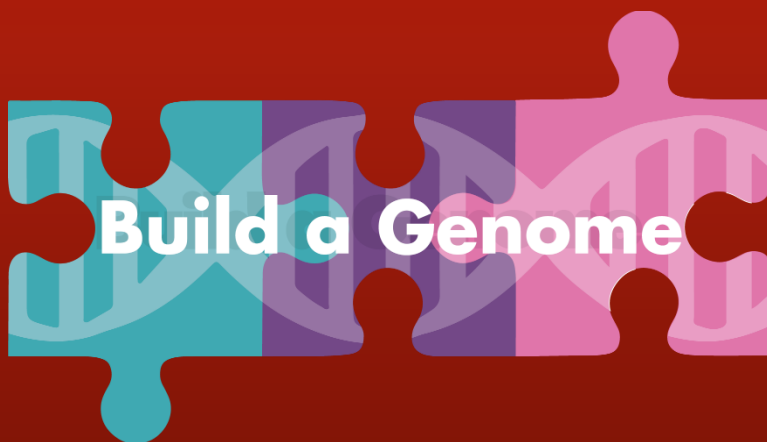


Build-a-Genome

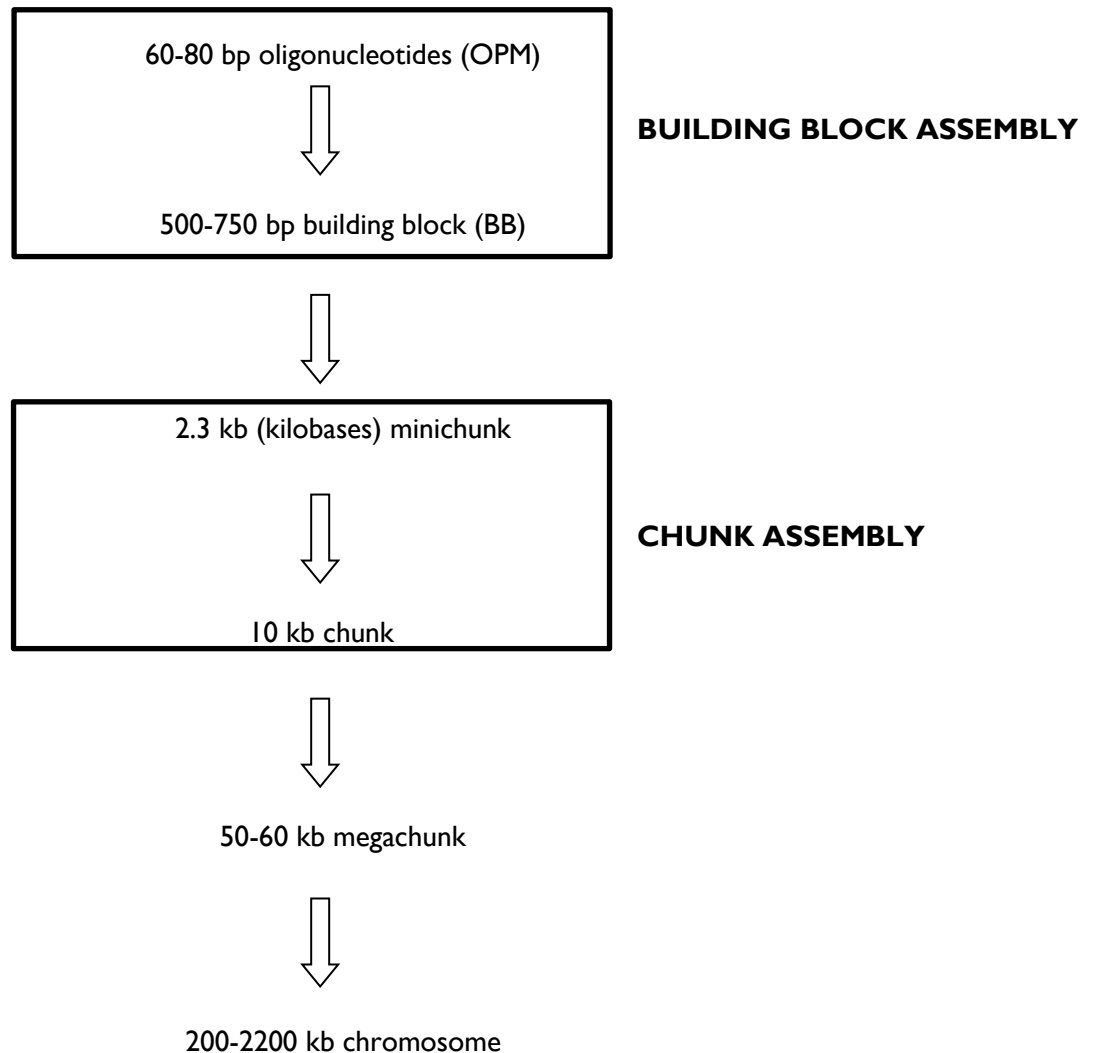
Synthesis of Red Fluorescent Protein
Lab Manual



Contents

OVERVIEW OF GENOME ASSEMBLY STEPS.....	2
OVERVIEW OF BUILDING BLOCK ASSEMBLY STEPS	3
Oligo Design Bioinformatics Assignment	4
Creation of Oligo Mixes for TPCR	10
Templateless PCR.....	11
Finish PCR.....	13
Gel Electrophoresis.....	15
Gibson Assembly and Bacterial Transformation	18
Gibson Assembly.....	20
Bacterial Transformation.....	20
DNA Sequence Analysis	24
Troubleshooting PCR	28
Design Information for the Instructors.....	29

OVERVIEW OF GENOME ASSEMBLY STEPS



Note to instructors: Due to the dropping cost of gene synthesis, it is no longer feasible to have students perform building block assembly as part of the Synthetic Yeast Project, since we can directly purchase minichunks at a lower cost.

However, we think it is still valuable to have students go through the process of building block assembly during the boot camp phase of the course. Students can synthesize either a portion of the synthetic yeast genome that has already been completed (for example a portion of chromosome III), or they can assemble a model gene such as GFP or RFP. The advantage of using a model gene is that students will be able to visualize that they are successful after transforming bacteria (this is especially true of RFP, where the bacterial colonies will be pink). This allows colony-screening PCR (to confirm synthesis of the gene) to be skipped during the boot camp phase if time is limited.

OVERVIEW OF BUILDING BLOCK ASSEMBLY STEPS

Step 1: Oligo design

The goal of building block synthesis is to assemble a “building block (BB)” of ~750 bp. We use software to divide the BB into single -stranded DNA pieces that are 60-80 bp long (oligonucleotides or oligos). Individual oligos are then ordered from commercial suppliers.

Step 2: Templateless PCR

Templateless PCR is performed to form a larger construct (the building block, or BB) from mixtures of oligonucleotides.

Step 3: Finish PCR and Gel electrophoresis

Finish PCR is performed to amplify the full-length building block that is produced in small quantities in the templateless PCR. The success of the finish PCR is gauged by running a portion of the finish PCR reaction on a gel.

Step 4: Ligation and Bacterial Transformation

Building blocks that are of the correct size are cloned into a genetic cloning vector and used to transform bacteria. Bacteria are transformed with the cloned construct to create a population of clones for sequencing and future assembly into larger synthetic yeast constructs.

Step 5: Colony screening PCR

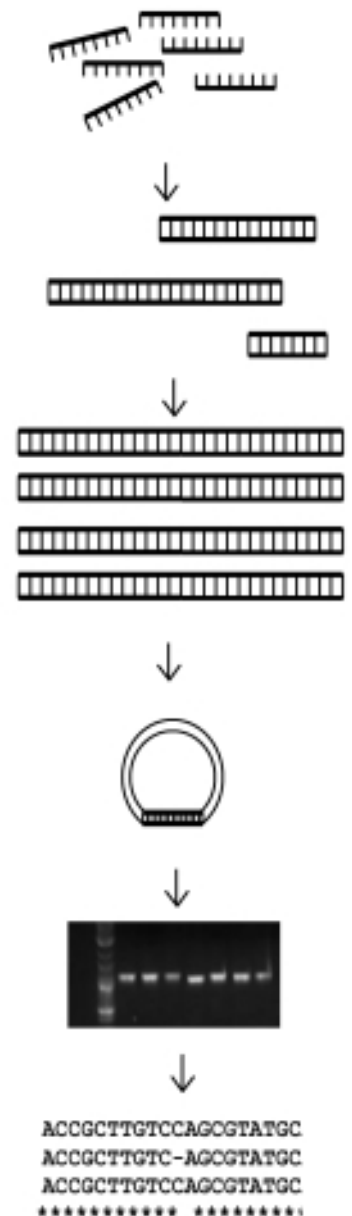
We verify that each bacterial clone has a DNA insert of the correct size by performing PCR and gel electrophoresis.

Step 6: DNA Sequence Analysis

Bacterial clones containing building blocks that are the correct size are sequenced to make sure that there are no sequence errors. Once the DNA has been sequenced, we compare the DNA sequence of each clone to the desired sequence to identify those without sequence errors (we call these “perfect clones”). This completes BB synthesis.

Step 7: Troubleshooting PCR

Since many BBs will not assemble correctly the first time, we change the templateless PCR and finish PCR reaction conditions to enable assembly of each building block.



Oligo Design Bioinformatics Assignment

Our goal is to build a section (a “building block”) of the synthetic yeast genome. We do not want to change the sequence of amino acids (we do not want to create dead yeast!), but we want to change which codons encode the amino acids. This is termed **recoding the genome**.

To recode the genome, we must obtain the DNA sequence of each building block. Our goal is to use the most preferred codons to recode the yeast genome, thereby optimizing protein synthesis. We can derive a DNA sequence from the protein sequence by “reverse translating” the protein to obtain a nucleotide sequence (note that for any protein sequence one can obtain multiple different DNA sequences depending upon which codons you choose to use).

The amino acid sequence of the building block that we will be using is:

MAKINGSYNTHETICGENESTHISISMYFAVRITEGENEWHATISYRS

(we are using the one letter abbreviation for amino acids as shown in the table below)

Abbreviation	Amino acid
A	alanine
C	cysteine
D	aspartic acid
E	glutamic acid
F	phenylalanine
G	glycine
H	histidine
I	isoleucine
K	lysine
L	leucine
M	methionine
N	asparagine
P	proline
Q	glutamine
R	arginine
S	serine
T	threonine
V	valine
W	tryptophan
Y	tyrosine
X	stop codon

To design our oligonucleotides, we will use a program called Gene Design, available at www.genedesign.org.

Step 1: Reverse Translate to obtain a nucleotide sequence

Click on “Design a Gene”. Enter the amino acid sequence of the building block from the previous page (using the single letter amino acid code). Click on “Reverse Translate”. You should get a nucleotide sequence as an output from the program.

Step 2: Recoding the genome using the preferred codons of *S. cerevisiae*

Synonymous codons are codons that encode the same amino acid. We can recode the genome by substituting synonymous codons for the original codons. Organisms tend to favor certain synonymous codons over others (these are called preferred and non-preferred codons), a phenomenon we term **codon bias**. Preferred codons are associated with more efficient translation while non-preferred codons are associated with less efficient translation. If we use all preferred codons when recoding the genome, we are “optimizing the genome”.

Click “Back” on your browser. Notice that at the bottom you can select the organism that you are working on. This table will then show you the most preferred codons for each of the 6 organisms (H. sapiens, *S. cerevisiae*, *E. coli*, *B. subtilis*, *D. melanogaster* and *C. elegans*).

1. For which amino acids is the same codon preferentially used in all 6 species? How many differences are there in preferred codon usage between *S. cerevisiae* and *C. elegans*?

Select *S. cerevisiae* as your organism and click on “reverse translate” to obtain a nucleotide sequence. The nucleotide sequence that you are given now uses the preferred codons of *S. cerevisiae* to encode your building block. This potentially alters the nucleotide sequence without altering the protein sequence.

2. When thinking about gene expression, what is one advantage to using the preferred codons to encode your synthetic gene(s)? By changing the DNA sequence, what are some potential problems that we may encounter at the level of the RNA? By using only the 21 preferred codons of *S. cerevisiae* instead of all 64 codons, what potential problems may we encounter?

Step 3: Other options for recoding the genome

At the bottom of the web page, click on “Codon Juggling”. This will give you several different versions of the sequence: an optimized version, a most different sequence, etc. Scroll to the bottom of the page and notice that all of the sequences encode the same amino acids.

The overall GC content of yeast is 38%. Notice that for each of the nucleotide sequences that you are given the %GC is very similar. The program is designed to maintain the GC content of the DNA when selecting which codon to use for each amino acid (in order to obtain the proper GC content, the most preferred codon is not always the one that is used at each position).

3. Why might we want to control the %GC in each building block? What properties of the DNA sequences would be affected by a change in the %GC?

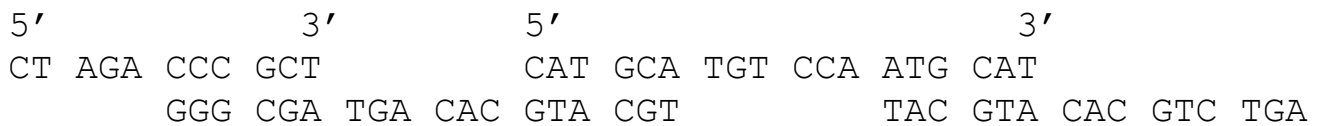
Step 4: Breaking the building block into overlapping oligonucleotides

Now that we have our nucleotide sequence, we need to have DNA chemically synthesized. However, only single stranded DNA can be chemically synthesized. Rather than ordering one piece of DNA that is the length of our building block (750 bp), we order our DNA as oligonucleotides of approximately 60 bp.

4. If we ordered one piece of DNA that was 750 bases long from the synthesis company, what percent of the DNA molecules would be perfect (without any synthesis errors) if the efficiency of synthesis were 99.5% for each step of oligo synthesis? If we ordered a piece of DNA that was 60 bases long, what percent of the DNA molecules would be perfect if the efficiency of synthesis were 99.5% for each step of oligo synthesis?

We order our oligonucleotides as single-stranded DNA and then use enzymes to synthesize the complementary strand, thereby producing double-stranded DNA. We design our oligos to partially overlap one another (we term these “gapped oligos”). This allows the end of one oligonucleotide to serve as a primer for DNA synthesis.

- Each oligonucleotide can only be extended in one direction. Why? On the figure below depicting overlapping oligos, fill in the missing DNA nucleotides and use an arrow to indicate the direction in which each section of DNA is synthesized.



Step 5: Optimizing oligonucleotide melting temperatures.

- A building block is typically comprised of 12-18 oligonucleotides. How does the use of oligos that overlap one another ensure that the oligos assemble in the correct order to form the building block?

One of the reasons that we need the Gene Design program is that the program will adjust the length of the oligos to design oligos that have the same T_m in their overlap regions.

- Why might it be important to have a uniform T_m when assembling oligos into a building block?

Click “Back” on your browser and then click “BB design (sequence overlap)”. Make sure the settings are at “Target oligo length” of 80 bp and “overlap melting temperature” of 56deg. Make sure that “generate gapped oligos” is selected. Click “Design oligos”. Notice that in the white box at the top you see four DNA strands: the top and bottom strands are the two complete strands of the building block and in between them you see the overlapping oligos that can be used to synthesize the building block. At the bottom of the page, you should see all of your oligos that are needed to make this building block listed (all written in the 5' to 3' direction).

8. What is the length of your building block? How many oligos are required to synthesize the sequence?

 9. Add up the lengths of all of your oligos. What is the total length of all the oligos together? Why does this differ from the overall length of the (building block (think about how the oligos assemble)?

 10. What is the average length of the oligos and what is the range of oligo lengths? Why aren't all of the oligos the same size (why have some been lengthened and others shortened)?
- Hit the “Back” button on your browser. Change the “Overlap melting temperature” to 64 degrees.
11. What about the oligos has changed? How would this affect the stability (T_m) of the double stranded DNA?

 12. What would be the advantage to using longer oligos for gene synthesis? What would be the advantage to using smaller oligos for gene synthesis? (Think about your answers about error rates above).

Step 6: Designing a sequence for the synthetic yeast project

Finally, go to the GeneDesign homepage and select “Building block design (restriction site overlap)”. Go to the database and find the sequences of the building blocks that you will be assembling. Copy one of the sequences into GeneDesign and click “Design oligos”.

13. What is the length of your building block?

How many oligos are required to synthesize the sequence?

What is the average length of the oligos and what is the range?

What is the average melting temperature of the oligos?

14. Click “Assembly oligos (FASTA format)” and submit the list of oligos for one building block along with your assignment. Repeat for your remaining building blocks.

Creation of Oligo Mixes for TPCR

Typically, 12-18 oligos are used to synthesize each building block (BB). These oligos are ordered individually and are delivered to us in 96 well plates. Each 96-well plate can contain oligos to make up to 4-5 complete building blocks.

For each building block that you are assigned, you will need to combine these 12-18 oligos together to create a templateless primer mix (TPM). This TPM will be used in a templateless PCR reaction (TPCR), which will anneal the oligos together and extend them to make longer DNA sequences, including only a small amount of the full-length building block sequence. Next week, you will use the outer primer mix (OPM) to amplify the full-length BB in a finish PCR reaction (FPCR) so that the predominant DNA following FPCR is your desired full-length BB.

Templateless primer mix (TPM)

The templateless primer mix (TPM) contains the 12-18 oligos that comprise the building block. All primers must be present at a concentration of 300nM in the TPM (a dilution of 1/20). These dilutions should be done in a 1.7ml microcentrifuge tube, and primer mixes should be stored at -20°C when not in use.

1. Label your tube with: the BB name, the type of oligo mix (TPM), and the primer concentration (300 nM)
2. To the tube, add 10 ul of each of the TPM oligos
3. Add H₂O to get a final volume of 200ul.
4. Mix thoroughly and keep the tube on ice.

In some cases, there may be more than 20 oligos that comprise a building block. Mix 10 ul of each oligo and add no water. Even though the concentration will be slightly more dilute than 300 nM this does not seem to have any effect on reaction efficiency.

Outer primer mix (OPM)

For each building block, the outer primer mix (OPM) contains **only** the first and last oligos. The outer primers must be present at a concentration of 3uM in the OPM (the oligos must both be diluted by 1/2). Again, these dilutions should be done in a 1.7ml microcentrifuge tube, and primer mixes should be stored at -20°C when not in use.

1. Label your tube with: the BB name, the type of oligo mix (OPM), and the primer concentration (3 uM)
2. To the tube, add 10 ul of each of the OPM oligos
3. Mix thoroughly
4. Put this primer mix into the box marked "OPM mixes" to be stored until next week

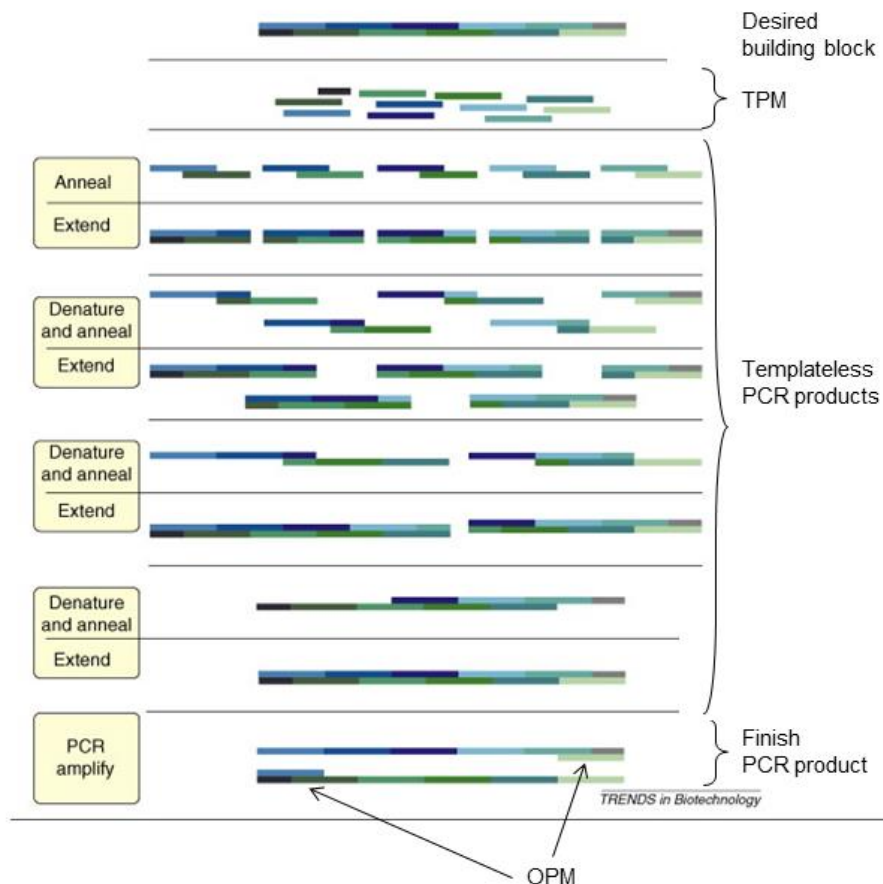
Templateless PCR

Notes to instructors:

- The protocol below is designed for each student to synthesize one building block. Students can set up templateless PCR reactions for multiple building blocks by creating master mixes for multiple reactions.
- The positive control given to students consists of a plasmid and two primers that are known to amplify well.

The goal of templateless PCR is to combine a mixture of short single-stranded DNA oligonucleotides (the TPM) and assemble them into a longer double-stranded DNA sequence, which we term a building block (BB). This building block DNA may have a sequence that is completely new and may not have ever previously existed in nature.

The oligos are annealed together during the templateless PCR reaction and are then extended through multiple rounds of PCR until finally a small amount of the full-length BB (approximately 500-750 bp) is synthesized.



Trends in Biotechnology
Volume 27, Issue 2, p63–72, February 2009

Reaction Setup

- Before you set up reactions, make sure you reserve a spot in the PCR machine. You will need to sign up for 3 spaces in the PCR machine-label each with your initials and either “BB”, “PC” or “NC”.

1. You will be performing 3 PCR reactions (1 building block (BB) plus one positive control reaction (PC) and one negative control reaction (NC).

2. Combine all reagents listed below into three different PCR tubes (these are the very small tubes).

Reagent	Vol per 25ul reaction
Water	10 ul
2X Phusion Master Mix	12.5 ul
Total	22.5 ul

3. To the first tube, add 2.5 ul of the TPM.

4. To the second tube, add 2.5 ul from the tube marked PC. This will be your positive control reaction.

5. To the third tube, add 2.5 ul of water. This will be your negative control reaction.

6. Cap your tubes and **make sure that they are sealed tightly** so that the liquid will not evaporate.

7. Place your tubes in the PCR machine in the position for which you signed up. Make sure that you have recorded which sample (BB, PC, or NC) is in each position in the PCR machine.

Reaction Conditions:

1 cycle:

- 94°C, 3 minutes
- 55°C, 30 seconds
- 72°C, 1 minute

5 cycles:

- 94°C, 30 seconds
- 69°C, 30 seconds
- 72°C, 1 minute

5 cycles:

- 94°C, 30 seconds
- 65°C, 30 seconds
- 72°C, 1 minute

20 cycles:

- 94°C, 30 seconds
- 61°C, 30 seconds
- 72°C, 1 minute

1 cycle:

- 72°C, 3 minutes

Finish PCR

Note to instructors:

- **The protocol below is designed for each student to synthesize one building block. Students can set up templateless PCR reactions for multiple building blocks by creating master mixes for multiple reactions.**
- **The positive control given to students consists of a plasmid and two primers that are known to amplify well.**

We just used templateless PCR to assemble a small amount of the full-length building block (BB) from oligonucleotides. However, templateless PCR produces very little of the full-length BB. Because templateless PCR is so inefficient at synthesizing the full-length BB, it also creates many smaller DNA fragments that are intermediates in building block synthesis. Therefore, the templateless PCR that we performed last week resulted in a mixture of DNA products: a small amount of the full-length BB is present among DNA products of varying sizes.

The goal of finish PCR is to exponentially amplify the building block sequence so that most of the DNA present in the mixture following finish PCR will be the full-length BB. This is accomplished by using the outer primer mix (OPM), which contains two oligos to amplify only full-length DNA. The OPM primers bind to either end of the building block (which now serves as a template), enabling only BBs that are full-length (and therefore contain both ends of the BB sequence) to be amplified by PCR (see figure on page 11).

Reaction Setup

- Before you set up reactions, make sure you reserve a spot in the PCR machine. You will need to sign up for 3 spaces in the PCR machine-label each with your initials and either “BB”, “PC” or “NC”.

1. Obtain your templateless PCR reactions from the PCR machine.
2. You will no longer need your PC and NC reactions today. To store them, transfer each to a separate tube by pipetting the liquid into a 1.7 ml tube. Label each tube with:
 - Your initials
 - ”TPCR”
 - Either “PC” or “NC”

Put these tubes away to be frozen for next week, when we will run it on a gel-if the PC reaction shows a band of DNA and the NC reaction does not, it will indicate that you properly set up your TPCR reactions.

2. For your BB reaction, add 175 ul of water, cap the tube and invert to mix. We will use this TPCR BB reaction in step 4 below.

You will be performing 3 PCR reactions (one building block (BB) plus one **NEW** positive control reaction (PC) and one **NEW** negative control reaction (NC).

3. Combine all reagents listed below into 3 different PCR tubes (the very small tubes).

Reagent	Vol/25ul reaction
Water	8 ul
2X Phusion Master Mix	12.5 ul
OPM	2 ul
Total	22.5 ul

4. To the first tube, add 2.5 ul of the TPCR BB reaction from step 2 above. Check with you instructor if you should keep or discard the remainder of the TPCR BB reaction.

5. To the second tube (PC), add 2.5 ul from the tube labeled PC.

6. To the third tube (NC), add 2.5 ul of water.

7. Cap your tubes and **make sure that they are sealed tightly** so that the liquid will not evaporate. Place your tubes in the PCR machine in the position for which you signed up. Make sure that you have recorded which sample (BB, PC or NC) is in each position in the PCR machine.

Reaction Conditions:

1 cycle:

94°C, 3 minutes

25 cycles:

94°C, 30 seconds

55°C, 30 seconds

72°C, 1 minute

1 cycle:

72°C, 3 minutes

Gel Electrophoresis

So far, we have performed two steps of synthetic gene assembly to create a building block: (1) templateless PCR to assemble oligos into a small amount of full-length building block, and (2) finish PCR to amplify the full-length BB so that it becomes the most abundant DNA following PCR.

Now, we need to check how well each of these steps worked by running our PCR products (and controls!) on an agarose gel to verify whether we have assembled a significant amount of the full-length BB. Agarose gel electrophoresis will separate DNAs based upon their size. Therefore, if assembly of our BBs has been successful, we expect to see one DNA band whose size should approximate that of our desired BB. You can determine the size of each of your DNAs by comparing the size of your band to the size of the DNA marker (the 2-Log DNA ladder), which you will run on the gel with your DNAs.

Remember, we also have 4 control reactions (the PC and NC controls from the templateless PCR and the PC and NC controls from the finish PCR).

Pouring a Gel:

1. Weigh out 0.5 g of agarose on a piece of weigh paper. Transfer to an Erlenmeyer flask. Add 50 ml of 1x TAE.

2. Place the flask on a hot plate and heat until the agarose is completely transparent and colorless. Using a rubber heat mitt, periodically swirl the flask to ensure that the agarose is melting evenly. If there is any cloudiness or particulates, then place the agarose back on the hot plate for additional time until clear (check frequently).

3. Remove the flask of clear agarose and allow it to cool. This will take about 10 min. If you can hold your hand on the flask for 20 sec, then the agarose is cool enough to pour. (Warning: if poured while too hot, the agarose will leak and warp the plastic gel tray.)

4. While the agarose is cooling, place the gel tray into the gel box and add the black blockers and a comb.

5. When the agarose is cool, bring it to the instructor to add 1 ul of ethidium bromide to the melted agarose (**CAUTION: ethidium bromide is a mutagen. Wear gloves at all times when handling gels and gel equipment!**).

6. Swirl the agarose to incorporate the ethidium bromide and pour the agarose into the gel tray – your gel should be about 1 cm thick.

7. Allow at least 20 minutes for the gel to solidify. Once solid, carefully remove the comb and the black blockers, and place the solidified gel (still on the tray) into the gel box so that the wells are oriented on the same side as the black electrode.

8. Add enough 1x TAE buffer to completely cover the gel by about 1 cm.

Preparing your samples:

1. Transfer each of your PCR products from the small PCR tubes to a new larger tube. Label each tube with either “BB”, “FPCR PC”, or “FPCR NC”.
2. Obtain your TPCR PC and TPCR NC tubes that were stored in the freezer from last week. You should have 5 total samples now.
3. On a piece of parafilm, spot out 2 ul of 5x DNA loading dye with a P20 pipet. You should have 5 spots (one for each of your PCR products). Wait until your gel has started to solidify before beginning this step since the spots will dry out after ~10 minutes.
4. Add 3 ul of water to each spot of dye.
5. Add 5 ul of each PCR product to a separate spot of dye:
 1. TPCR PC reaction
 2. TPCR NC reaction
 3. FPCR PC reaction
 4. FPCR NC reaction
 5. Building block (BB)

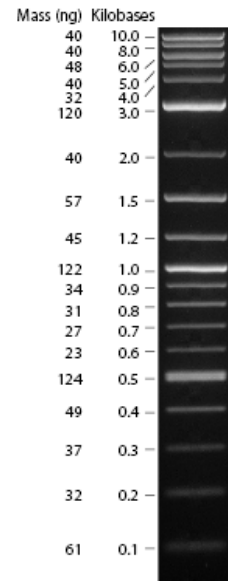
Running a Gel:

1. Into the first lane of the gel load 5 ul of the 2-Log DNA ladder (mixed with water and dye).
2. Into lanes 2-5, load 9 ul of each of your PCR products (mixed with water and dye).
3. Place gel lid with electrodes on gel box. REMEMBER DNA RUNS FROM NEGATIVE (black) TO POSITIVE (red). CHECK YOUR LEADS: BLACK SHOULD BE AT THE TOP, RED SHOULD BE AT THE BOTTOM!
5. Set voltage to 100V. You should see bubbles coming up from the wire electrode.
6. After a few minutes, check to make sure everything is going smoothly. Run gel approximately 30 minutes or until the dye is 2/3 of the way down the gel.

7. Take a picture of your gel.

8. Use the DNA ladder (a picture is on the right) to approximate the size of each of your building blocks. Make sure you check that your building blocks are the correct size - do not assume that because you have a band, it is the correct product!

9. You may discard all of your positive and negative control reactions once you have run your gel, but **be sure to keep your building block DNA in a tube that is clearly labeled with your initials and the name of the DNA building block.**



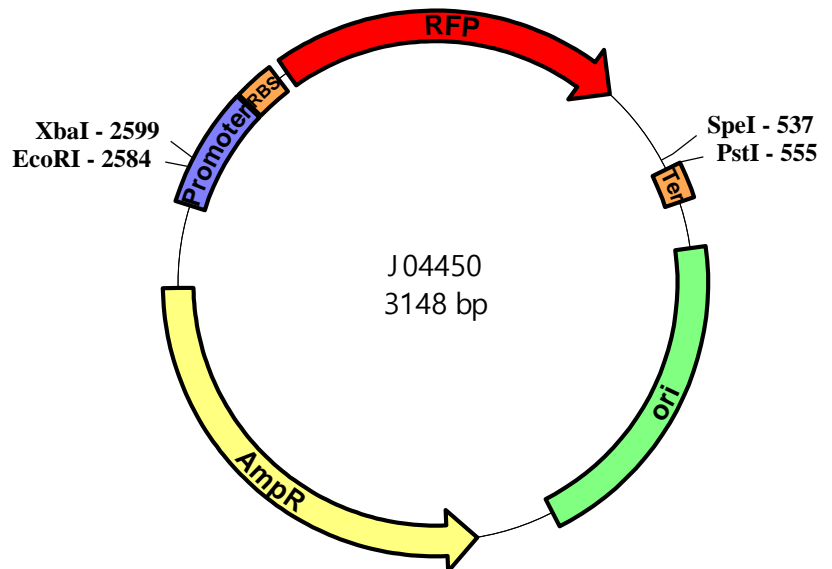
Gibson Assembly and Bacterial Transformation

Note to instructors: The RFP PCR product can be cloned into any blunt end cloning vector (we previously used the StrataClone Blunt vector). However, if you want the RFP gene to be expressed, you need to clone the gene into a vector containing a promoter. We make this simply by PCR amplifying the vector below with primers that overlap the RFP gene, thereby allowing Gibson assembly. The *DpnI* restriction enzyme is used to remove any template vector with RFP.

Following the production of our DNA building block (which codes for the Red Fluorescent Protein-RFP) by templateless and finish PCR and verification that most of this DNA is the correct size by gel electrophoresis, we now must clone this building block into a plasmid vector. This cloning step is important for two reasons. First, it allows us to separate the population of DNA molecules into individual DNA molecules, thereby enabling us to screen the individual molecules to identify those that have no errors in the DNA sequence. Second, cloning allows us to create a permanent frozen bacterial stock of each plasmid so that we may store each building block.

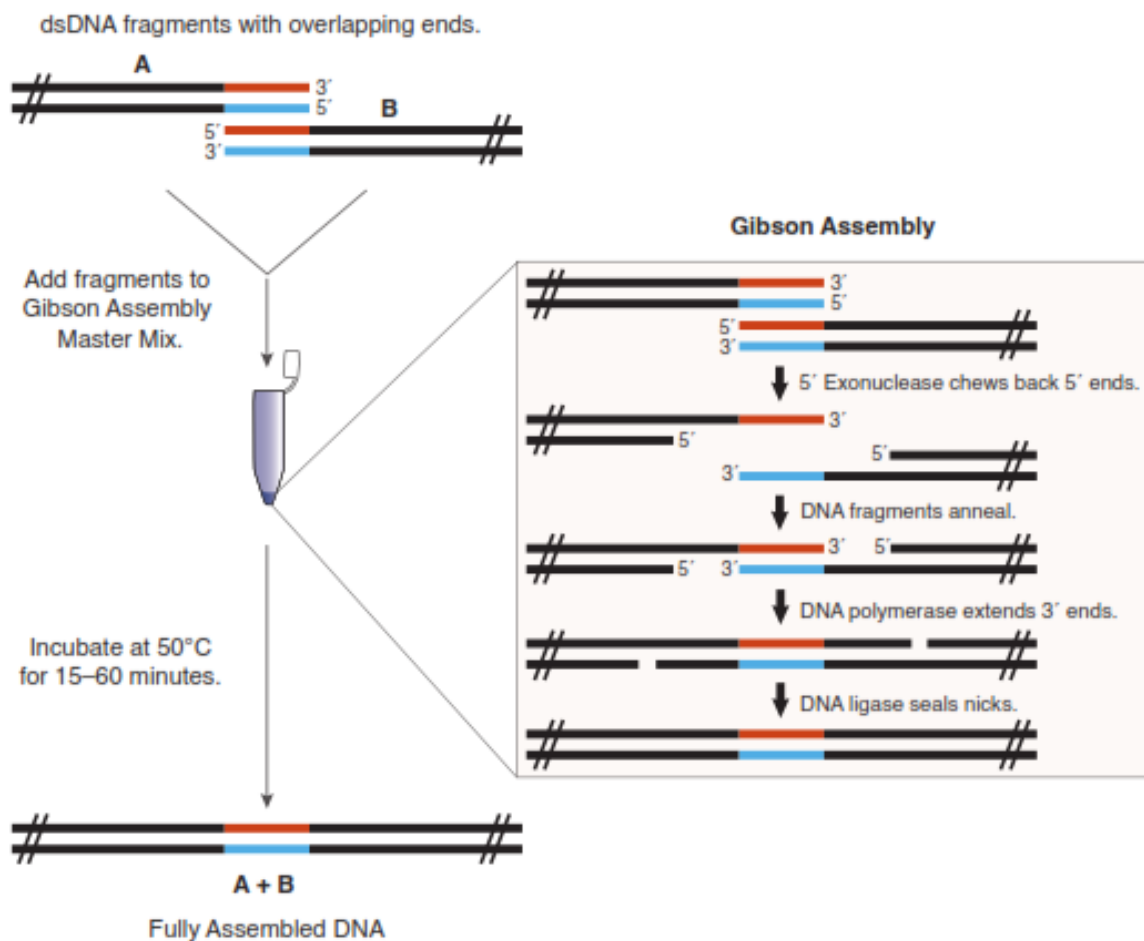
Our building block PCR product (the RFP gene) will be cloned into a plasmid called J04450 which has the following features:

- Bacterial origin of replication (ori)
- Multiple cloning site (MCS)
- Selectable marker (ampicillin resistance gene)
- Bacterial promoter
- Ribosome binding site (RBS)
- Transcription terminator (Ter)



Traditionally, we would clone the PCR product into the vector by using restriction enzymes to cut both the vector and DNA insert and then we would join them together with the enzyme DNA ligase. Instead of the traditional method, we will use a newer method called Gibson assembly. Gibson Assembly was developed by Dr. Daniel Gibson and his colleagues at the J. Craig Venter Institute. It allows for successful assembly of multiple DNA fragments, regardless of fragment length or end compatibility (the presence of compatible restriction sites). It has been rapidly adopted by the synthetic biology community due to its ease-of-use, flexibility and suitability for large DNA constructs.

Gibson assembly involves combining the vector (fragment A below), the DNA insert (fragment B below) and the Gibson Assembly Master Mix. The Gibson Assembly Master Mix



includes three different enzymes:

- The exonuclease creates single-stranded 3' overhangs in each of the DNA fragments. Note that the two fragments, A and B, have the same sequence at their ends (overlapping ends). The 3' overhangs therefore allow the fragments, which have complementary sequences, to anneal together.
- The DNA polymerase synthesizes DNA to fill in gaps within each annealed fragment.
- The DNA ligase seals each of the two DNA strands.

Gibson Assembly

Remember that the Gibson Assembly Master Mix contains enzymes and therefore **should be kept on ice at all times!**

1. Label a PCR tube (the tiny tubes) with your initials.
2. In that tube (**on ice**), combine:

Vector	5 ul
Building block insert	5 ul
Gibson Assembly Master Mix	10 ul

2. Place tube in the PCR machine and incubate at 50C for 15 minutes. (During this step, you can complete steps 1 and 2 under Bacterial Transformation below).
3. Remove tube from PCR machine and place on ice.

Bacterial Transformation

Once the vector and DNA insert are joined together with the Gibson Assembly Master Mix, the resulting recombinant DNA molecule is transformed into bacteria. Only a small number of bacterial cells in the population will become transformed. We can detect transformed cells because the plasmid DNA contains an antibiotic resistance gene; bacterial cells that become transformed will become resistant to the antibiotic (in this case, ampicillin). Ampicillin acts as a selective agent, allowing only those cells that contain a plasmid to grow.

E. coli that efficiently take up foreign DNA are said to be *competent*, and the method that they use to take up DNA is called *transformation*. Bacterial cells can be made competent by growing them at log phase, harvesting them from the growth media and then treating with calcium chloride. (It has been suggested that the Ca^{2+} neutralizes the negative charge of the phosphate backbone of the DNA. This neutralization prevents the DNA from being repulsed by the negatively charged phospholipids on the cell membrane, allowing the DNA to more efficiently enter the cell.) These competent bacteria can be stored at -80°C and subsequently used for transformation.

Competent bacteria take up DNA when the membrane is permeabilized, typically by heating quickly to 42°C . It is a tricky process because bacteria must be kept cold right until the heat shock step. **Taking the bacterial cells off of ice even briefly can lead to a dramatic decrease in transformation efficiency.**

It is important to use sterile technique throughout the transformation procedure. Keep the lids on your Petri dishes as much as possible and keep the foil top on the beaker of

microcentrifuge tubes. Remove the cap as briefly as possible when pipetting the media. **Please remember to wash your hands with soap before leaving the lab!**

This is the easiest step to "mess up" in the entire workflow. Cells must be kept cold at all times! As soon as they thaw, they start dying. The more time passes (up to the point where LB is added) the more they die and the fewer transformants you'll get. Chill labeled tubes, ligation mixes, etc., on ice prior to mixing of cells and DNA (ligation mix).

1. Get three LB/Amp plates and label them "BB", "PC", and "NC". Put the plates in the 37°C incubator to warm up.
2. Label three microcentrifuge tubes (the larger size tubes) "BB", "PC", and "NC". Close the caps and place these empty tubes on ice.
3. Add 2.0ul of the Gibson Assembly reaction to the "BB" tube. Place tube back on ice.
4. Add 2.0ul of the positive control DNA to the "PC" tube. Place tube back on ice.
5. Bring your tubes (**on ice**) up to the front to have 25ul of thawed competent cells transferred into each of your microcentrifuge tubes.
6. Incubate all tubes on ice for 30 minutes.
7. **Bring your ice bucket and a timer to the water bath**, and heat shock your tubes by placing at 42°C for 30 seconds. **Timing is critical here!**
8. Immediately (ie, **bring your ice bucket to the water bath**) place heat shocked cells back on ice for 2 minutes.
9. Add 350 ul room temperature SOC medium to each transformation reaction and then put in a microcentrifuge rack at room temperature.
10. Put the rack with tubes in the 37°C incubator for 1 hr.
11. Pipette 100 ul from each tube onto corresponding LB/Amp plates (only remove the tops from the plates long enough to add the cells). Spread bacteria by adding 5-6 sterile glass beads, shaking the plates with beads and then pouring beads into the waste.
12. Allow the liquid to soak into the plates. Once the plates appear dry, incubate the plates overnight at 37°C (When you are incubating or storing plates, they should be stored upside down to prevent condensation from dripping onto them).

Name:

Bacterial Transformation Assignment

1. Complete the table below:

	Number of white colonies	Number of pink colonies	Total number of colonies	Percent pink colonies
Building block				
Positive control				
Negative control				

2. When performing Gibson assembly, why is it crucial that the vector and insert contain the same sequences at each of their ends? What would happen if you tried to use Gibson assembly to join a vector and insert that did not have complementary DNA ends?

3. When cloning using restriction enzymes, we must very carefully pick which enzymes we use based on the sequence of our insert DNA. How is Gibson assembly more flexible, allowing you to clone an insert regardless of its sequence in contrast to traditional cloning with restriction enzymes?

4. Is plasmid J04450 a cloning vector or an expression vector? What sequences on the vector allow you to tell what type of vector it is?

5. How would you be able to tell if your **Gibson assembly reaction** didn't work (what differences would you expect in the number and/or color of bacterial colonies on your three plates)?

6. How would you be able to tell if your **bacterial transformation** didn't work (what differences would you expect in the number and/or color of bacterial colonies on your three plates)?

7. It would have been easy to take the building block PCR product and sequence it directly without first cloning it into a plasmid. Why is it important to perform cloning (putting the PCR product into a plasmid and transforming it into bacteria) before performing DNA sequencing to identify one DNA molecule with an error-free version of the RFP sequence?

8. When you pick bacterial colonies for DNA sequencing, it is crucial to pick only one colony and not to accidentally combine two colonies that are next to each other on the bacterial plate. Why would it be a problem if you did accidentally combine two colonies together (think about your answer to question 6 above)?

DNA Sequence Analysis

Now that we've screened our clones by colony-screening PCR to verify that they contain an insert of the correct size, we need to sequence the inserts to verify that they contain a building block without any sequence errors. Today we will learn to analyze DNA sequences to identify "perfect clones". What are the two most common sources of errors in the DNA when we are assembling DNA building blocks?

When sequencing data is sent to us, we receive not only a text file containing the sequence of the DNA insert, but we also receive the data from the sequencing machine in the form of a color-coded electropherogram. The electropherogram, or trace, represents the data obtained from sequencing detector, with the height of each peak representing the strength of the signal. We can therefore see the quality of the sequencing data that was obtained as well as investigate any ambiguities in the sequence. You will notice that the signal at the end of the electropherogram is not as strong as at the beginning; the peaks are much shorter and broader and become difficult to distinguish from one another. What is the reason that the strength of the signal tapers off?

Often, toward the end of a DNA sequence, instead of an A, T, C or G, the electropherogram will indicate N as the nucleotide. What does this N represent and why do they occur more frequently at the end of a sequencing read?

The data from a single reaction is termed a sequencing read. Our building blocks are 500-750 bp long. Looking at the electropherogram, would one sequencing reaction be sufficient to accurately determine the sequence of our clones? Why or why not?

In fact, we sequence each clone twice- once with a primer that anneals at the T7 sequence and once with a primer that anneals at the T3 sequence (these are in fact the same primers that we used for colony screening PCR). We call these "forward" and "reverse" sequencing reads. Based on the orientation of the T7 and T3 sequences relative to the insert in the vector, why does this ensure that we will get good sequencing data across the entire building block insert?

Now we need to determine if our clones contain a sequence that perfectly matches the building block or if they have DNA sequence errors. To accomplish this, we use a bioinformatics tool called Clustal W. Unlike BLAST, which performed local alignments (looking for short 11 bp “words” that were common between sequences), Clustal W uses global alignment.

Go to Moodle and download the zip file “Sequence Analysis Files” under the Lab section.

Go to the Clustal W homepage (www.ebi.ac.uk/Tools/clustalw2/index.html). Input the title of your sequence. Input the building block sequence. The line before the building block sequence must contain >Name.of.Sequence (no spaces). Skip a line and input the forward sequencing reaction, preceded by >Name.of.Sequence. Click Align.

Clustal W gives you a scores table indicating the pairwise alignment similarity score (out of 100). It also provides a DNA alignment. Residues that are identical in the two sequences marked with a *. Why does the alignment extend past the end of the building block sequence?

Click the back arrow and add the sequencing data from the reverse sequencing reaction, preceded by >Name.of.sequence. Why doesn't the alignment work well now?

Go to the Sequence Manipulation Suite (http://www.bioinformatics.org/sms2/rev_comp.html), make sure that you are on the Reverse Complement page and input your sequence. Click Submit. Cut and paste this sequence and replace the previous reverse sequencing read in Clustal W. The alignment should look much better now. For some sequenced clones, the reverse sequencing read will need to be reverse complemented, but for others the forward sequencing read will need to be reverse complemented. Why can we never predict whether the forward or reverse read will need to be reverse complemented?

Now, open the sequencing files for Analysis 1 in the Assignment folder. Input the data into Clustal W. Remember that you will need to reverse complement one of the sequencing reads. Then determine the nature and position of all mutations in the sequenced clone (ie deletion of A at nucleotide 140, transition of C to T at 34, etc.).

Remember, if your forward and reverse reads do not agree, one of the sequences is probably HIGHER quality than the other at every individual discrepant base (it's more likely the ends close to the sequencing primer are more reliable than the ends farther away from the sequencing primer).

Only record a mutation if the forward and reverse reads agree with each other and disagree with the building block sequence.

Click View Alignment File, print the sequence alignment, and **submit this with your assignment.**

	Name of building block	Mutations present in sequence
Analysis 1		
Analysis 2		
Analysis 3 (You only need to do 3-1 for this part)		

Error rates in building block synthesis

For analysis 3, there are 4 different clones of the same building block. Align the sequences for each of the 4 clones to the reference building block and identify all of the mutations in those 4 sequences. **Only record a “mutation” if the forward and reverse reads agree with each other and disagree with the building block sequence. Also, only record a mutation if it falls within the bounds of the building block sequence.**

Sequencing ID	Name of Building block	Length of BB (nt)	# insertions	# deletions	# transitions	# transversions	Total # mutations
Analysis 3-1							
Analysis 3-2							
Analysis 3-3							
Analysis 3-4							

Total # of mutations in all 4 sequences combined: _____

Total # nucleotides sequenced (ie size of your BB x 4): _____

We would like to know the overall error rate for creation of our building blocks (we call this value α) since this information will help us to determine the efficiency of our method and protocols. The error rate can be calculated as follows:

$\alpha = (\text{Total \# mutations found}) / (\text{Total \# nucleotides sequenced that are not vector sequence})$.

For example: if you found 13 mutations in 4 clones of a 750 bp BB, then $\alpha = 13 / (4 * 750) = 0.001$

From the information above for the building block you analyzed, $\alpha =$ _____

If the probability of a clone being perfect (p_c) is $p_c = e^{-\alpha L}$ (where L is the length of the BB), what is the probability that any one of your clones is perfect? _____

However, what we really need to know is how many clones we need to send for sequencing. We want to send as few clones in a batch as possible (sequencing is expensive!) while still having a high probability that the batch will contain one clone with a perfect sequence. If the probability of having one clone in the batch be perfect (p_b) is $p_b = 1 - (1 - p_c)^B$ (where B is the number of clones in a batch), then what is the chance that a batch of 18 sequences will have a perfect clone? _____

A batch of 12 sequences? _____

Troubleshooting PCR

BB assigned _____ Expected size of BB (bp) _____

BB assigned _____ Expected size of BB (bp) _____

BB assigned _____ Expected size of BB (bp) _____

What was the problem with your original finish PCR reaction for each building block (no DNA product, multiple distinct DNA bands, smeared DNA band)?

What is your hypothesis for why your PCR did not work?

What did you change to try to get your finish PCR to work (annealing temp changes, extension temp changes, additional rounds of PCR, etc.)? Remember, each of these BBs must be synthesized eventually, so as much information as you can give us about what has already been tried to troubleshoot each specific BB will be very helpful.

What was the result and how would you interpret it (attach gel images of your original PCR and you troubleshooting gel)? Is this PCR product ready to be ligated into the vector? If not, why not?

Design Information for the Instructors

The RFP gene that this project is based on is described here:
http://parts.igem.org/Part:BBa_E1010:Design

The sequence of the gene itself is:

```
atggcttctccgaagacgttatcaaagagttcatgctgttcaaagttcgtatggaaggttcggttaacgggtcacgagttcgaatcgaaggtgaaggtg
aaggtcgtccgtacgaaggtaccagaccgctaaactgaaagttacaaaggtgggtccgctgcttgggacatcctgtcccgcagttccagta
cggttccaaagcttacgtaaacaccggctgacatcccggactacctgaaactgtccttcccgaaggttcaaatgggaacgtgtatgaaactcgaa
gacgggtggtgtgttacggttaccaggactcctcctgcaagacgggtgagttcatctacaaagttaaactgctggttaccacttcccgtccgacggtc
cggttatgcagaaaaaacatgggttgggaagcttccaccgaacgtatgtaccgggaagacggtgctctgaaaggtgaaatcaaatgctgctgaaact
gaaagacgggtgctactacgacgtgaagttaaaaccactacatggctaaaaaccggttcagctgcccgggtgcttataaaaccgacatcaactggac
atcacctcccacaacgaagactacaccatcgttgaacagtacgaacgtgctgaaggtcgtcactccaccgggtgcttaataacgctgatagtctagtga
gatcgc
```

And the oligos used to synthesize the gene are:

```
RFP.o01 ATGGCTTCCTCCGAAGACGTTATCAAAGAGTTCATGCGTTTCAAAGTTCGTATGGAAGGTTCCG
RFP.o02 CTTACCTTCACCTTCGATTTTCAAACTCGTGACCGTTAACGGAACTTCCATACGAACTTTGAA
RFP.o03 TTCGAAATCGAAGGTGAAGGTGAAGGTCGTCCGTACGAAGGTACCCAGACCGCTAAACTG
RFP.o04 CCCAAGCGAACCGGCAGCGGACCCACCTTTGGTAACTTTAGCGGTCTGGGTACC
RFP.o05 CCGCTGCCGTTTCGCTTGGGACATCCTGTCCCGCAGTTCAGTACCGTTCCAAAGCTTACGTT
RFP.o06 GACAGTTTCAGGTAGTCCGGGATGTCAGCCGGGTGTTTAAACGTAAGCTTTGGAACCGTACTG
RFP.o07 ATCCCGGACTACCTGAAACTGTCCTTCCCGGAAGGTTTCAAATGGGAACGTGTTATGAACTTCAAG
RFP.o08 GGAGTCCTGGGTAAACGGTAAACAACACCACCGTCTTTCGAAGTTCATAACACGTTCCCATTT
RFP.o09 GTTACCGTTACCCAGGACTCCTCCCTGCAAGACGGTGAGTTCATCTACAAAGTTAAACTGCGTGGTAC
RFP.o10 GCATAACCGGACCGTCCGACCGGAAGTTGGTACCACGCAAGTTTAACTTTGTAGATGAA
RFP.o11 TCCGACGGTCCGGTTATGCAGAAAAAACCATGGGTTGGGAAGCTTCCACCGAACGTATGTAC
RFP.o12 CAGACGCATTTTATTTTACCTTTTACAGAGCACCCTTCCGGGTACATACGTTCCGGTGGAAAGCTTC
RFP.o13 CTGAAAGGTGAAATCAAAATGCGTCTGAAACTGAAAGACGGTGGTCACTACGACGCTG
RFP.o14 GGCAGCTGAACCGGTTTTTTAGCCATGTAGGTGGTTTTAACTTCAGCGTCGTAGTGACCACC
RFP.o15 GCTAAAAAACCGGTTTCAGCTGCCGGGTGCTTACAAAACCGACATCAAACCTGGACATCACCTCC
RFP.o16 ACGTTCGTAAGTGTCAACGATGGTGTAGTCTTCGTTGTGGGAGGTGATGTCCAGTTTGTATGC
RFP.o17 ACCATCGTTGAACAGTACGAACGTGCTGAAGGTCGTCCTCCACCGGTGCTTAATAACGCTG
RFP.o18 GCGATCTACTAGCACTATCAGCGTTATTAAGCACCGGTGG
```

The plasmid vector is J04450 and information on the vector is available at:

http://parts.igem.org/Part:BBa_J04450

The vectors used to amplify the plasmid vector are:

```
J04450.Fwd    tgctagttagatcgcCTACTAGAGCCAGGCATCAAATAAAACGAAAGGCTCAG
J04450.Rev    cttcggaggaagccatTCTAGTATTTCTCCTCTTTCTCTAG
```

(Yellow and green represent the overlap with the RFP gene for Gibson assembly while the upper case letters represent homology to the J04450 vector that is being amplified).