Chapter 7
Yeast Colony PCR

The S. cerevisiae strains that we are using this semester were constructed by the Saccharomyces Gene Deletion Project. In the project, yeast investigators systematically replaced every ORF in the yeast genome with a bacterial kanamycin resistance (KAN<sup>R</sup>) gene. In this lab, you will use the polymerase chain reaction (PCR) to identify the disrupted MET genes in your deletion strains. Thermostable DNA polymerases (above) play a key role in PCR.

Objectives
At the end of this lab, students should be able to:

- describe the reactions that are occurring at the different temperatures used in PCR cycles.

- design oligonucleotide primers to amplify sequences with PCR.

- explain how changes to the annealing temperature and extension time affect the production of PCR products.

- design and carry out a PCR strategy that distinguishes three met deletion strains.
Chapter 7

In a previous lab, you used different culture media to distinguish between different *S. cerevisiae met* mutants. Your results may have allowed you to tentatively identify your strains. In this lab, you will use the polymerase chain reaction (PCR) to more conclusively identify the mutant strains. This chapter begins with an overview of the PCR and the *Saccharomyces* Gene Deletion Project. You will use this knowledge to design and carry out a strategy for identifying *met* deletion strains by yeast colony PCR.

**Polymerase chain reaction overview**

The polymerase chain reaction (PCR) revolutionized molecular biology. With PCR, researchers had a tool for amplifying DNA sequences of interest from extremely small amounts of a DNA template. Indeed, billions of copies can be synthesized from a single DNA molecule in a typical PCR reaction. The development of PCR grew out of research on DNA polymerases and the discovery of thermostable DNA polymerases able to withstand extended heat treatments that denature most proteins (Sakai *et al.*, 1988). Today, PCR is a standard technique that is widely used to analyze DNA molecules and to construct novel recombinant molecules.

Thermostable DNA polymerases are central to PCR. The first description of PCR used a DNA polymerase from *E. coli*, which denatured and had to be replaced after each round of DNA synthesis (Sakai *et al.*, 1985). The procedure was much-improved by replacing the *E. coli* polymerase with a DNA polymerase from *Thermus aquaticus*, a bacterium that thrives in thermal springs at Yellowstone National Park. The *T. aquaticus* DNA polymerase, or *Taq* polymerase, functions best at temperatures of 70-75°C and can withstand prolonged (but not indefinite) incubation at temperatures above 90°C without denaturation. Within a few years, the *Taq* polymerase had been cloned and overexpressed in *E. coli*, greatly expanding its availability. Today, the selection of polymerases available for PCR has increased dramatically, as new DNA polymerases have been identified in other thermophilic organisms and genetic modifications have been introduced into *Taq* polymerase to improve its properties.

PCR involves multiple rounds of DNA synthesis from both ends of the DNA segment that is being amplified. Recall what happens during DNA synthesis: a single-stranded oligonucleotide primer binds to a complementary sequence in DNA. This double-stranded region provides an anchor for DNA polymerase, which extends the primer, *ALWAYS* traveling in the 5’ to 3’ direction. Investigators control the start sites for DNA replication by supplying oligonucleotides to serve as primers for the reaction (shown below for *Your favorite gene Yfg*). To design PCR primers, investigators need accurate sequence information for the primer binding sites in the

![Diagram of PCR primers](attachment:PCR_diagram.png)
target DNA. (Note: Sequence information is not needed for the entire sequence that will be amplified. PCR is often used to identify sequences that occur between two known primer binding sites.) Two primers are required for PCR. One primer binds to each strand of the DNA helix.

PCR begins with a denaturation period of several minutes, during which the reaction mixture is incubated at a temperature high enough to break the hydrogen bonds that hold the two strands of the DNA helix together. Effective denaturation is critical, because DNA polymerase requires single-stranded DNA as a template. The initial denaturation segment is longer than subsequent denaturation steps, because biological templates for PCR, such as genomic DNA, are often long, complex molecules held together by many hydrogen bonds. In subsequent PCR cycles, the (shorter) products of previous cycles become the predominant templates.

Following the initial denaturation, PCR involves a series of 30-35 cycles with three segments, performed at different temperatures. PCR reactions are incubated in thermocyclers that rapidly adjust the temperature of a metal reaction block. A typical cycle includes:

- a denaturation step - commonly 94°C
- a primer annealing step - commonly 55°C
- an extension step - commonly 72°C

PCR reactions include multiple cycles of denaturation, annealing and extension. Each cycle of PCR includes three different temperatures. In the denaturation step, the hydrogen bonds holding DNA helices together are broken. In the annealing step, oligonucleotide primers bind to single-stranded template molecules, providing starting points for processive DNA polymerases that extend the primer sequence. DNA polymerases become more active at the extension temperature, which is closer to their optimal temperature. Investigators adapt the temperatures and timing of the steps above to accommodate different primers, templates and DNA polymerases.
PCR is indeed a chain reaction, since the DNA sequence of interest roughly doubles with each cycle. In ten cycles, a sequence will be amplified ~1000 fold \((2^{10}=1024)\). In twenty cycles, a sequence will be amplified ~million fold. In thirty cycles, a sequence can be theoretically amplified ~billion fold. PCR reactions in the lab typically involve 30-35 cycles of denaturation, annealing and extension. To understand PCR, it's important to focus on the first few cycles. PCR products of the intended size first appear in the second cycle. Exponential amplification of the intended PCR product begins in the third cycle.

During the first cycle, the thermostable DNA polymerases synthesize DNA, extending the 3’ ends of the primers. DNA polymerases are processive enzymes that will continue to synthesize DNA until they literally fall off the DNA. Consequently, the complementary DNA molecules synthesized in the first cycle have a wide variety of lengths. Each of the products, however, has defined starting position, since it begins with the primer sequence. These “anchored” sequences will become templates for DNA synthesis in the next cycle, when PCR products of the intended length first appear. The starting template for PCR will continue to be copied in each subsequent cycle of PCR, yielding two new “anchored” products with each cycle. Because the lengths of the “anchored” products are quite variable, however, they will not be detectable in the final products of the PCR reaction.

DNA strands of the intended length first appear during the second cycle. Replication from the “anchored” fragments generates PCR products of the intended length. The number of these defined length fragments will double in each new cycle and quickly become the predominant product in the reaction.
Yeast Colony PCR

PCR products of the intended length first appear in the second cycle.
The “anchored” fragments generated during the first PCR cycle begin with either the primer 1 or primer 2 sequence. During the second cycle, replication begins at the other primer site, generating a PCR product is capped at both ends with primer sequences.

Most PCR protocols involve 30-35 cycles of amplification. In the last few cycles, the desired PCR products are no longer accumulating exponentially for several reasons. As in any enzymatic reaction, PCR substrates have become depleted and the repeated rounds of incubation at 94°C have begun to denature Taq polymerase.

Primer annealing is critical to specificity in PCR

Good primer design is critical to the success of PCR. PCR works best when the primers are highly specific for the target sequence in the template DNA. Mispriming occurs when primers bind to sequences that are only partially complementary, causing DNA polymerase to copy the wrong DNA sequences. Fortunately, investigators are usually able to adjust experimental parameters to maximize the probability that primers will hybridize with the correct targets.

PCR primers are typically synthetic oligonucleotides between 18 and 25 bases long. When designing a primer, researchers consider its $T_m$, the temperature at which half of the hybrids formed between the primer and the template will melt. In general, the thermal stability of a hybrid increases with the length of the primer and its GC content. (Recall that a GC-base pair is stabilized by three H-bonds, compared to two for an AT pair.) The following formula provides a rough estimate of the $T_m$ of oligonucleotide hybrids. In this formula, $n$ refers to the number of nucleotides, and the concentration of monovalent cations is expressed in molar (M) units.

$$T_m = 81.5°C + 16.6 (\log_{10}[K+ + Na+]) + 0.41 (%[G + C]) - (675/n)$$

When possible, researchers design primers that are similar in length and have a 40-60% GC composition. The sequences are designed so that the $T_m$'s of the primer-DNA hybrids are within a few degrees of the annealing temperature. Adjusting the $T_m$'s of the primers to be close to the annealing temperature favors specific hybrids over less specific hybrids that may contain a few mismatched bases. A hybrid formed between a primer and a non-target sequence with even one mismatched base has a $T_m$ that is lower than that of the fully hydrogen-bonded hybrid. If DNA polymerase extends the mismatched primer, incorrect PCR products will be generated. When mispriming appears to be a problem in a PCR reaction, investigators have several options to
increase the yield of the desired product. They can increase the length and/or GC content of the primers, alter the salt concentrations (results may be hard to predict) or increase the annealing temperatures.

When designing a PCR reaction, investigators also consider the nature of the template DNA. A variety of DNA templates can be used for PCR. Depending on the purpose of the experiment, investigators could choose to use genomic DNA, a plasmid or a cDNA (complementary DNA generated by a reverse transcriptase from mRNA). PCR can also be done with much cruder preparations of DNA, such as a bacterial or yeast colony. The more complex the template (its length in bp), the greater the probability that it will contain another sequence that is very similar to a primer sequence. For example, the haploid yeast genome is 12 Mbp long and contains only one copy of each MET gene. The probability that a non-target sequence in the yeast genome is similar enough to a 25-nucleotide MET primer to cause mispriming is reasonably good. Furthermore, these sequences with small mismatches may outnumber the target sequence. With complex templates such as genomic DNA, therefore, investigators can sometimes reduce the impact of mismatched hybrids by decreasing the amount of template DNA in the reaction. (Using too much template is the most common error in yeast colony PCR.)

The components of a PCR reaction are simple, consisting of the DNA template, primers, dNTPs, a buffer containing MgCl₂ (polymerases use dNTPs complexed with Mg²⁺), and the thermostable polymerase. For our experiments, we will be using a master mix that contains all of the components except the template DNA and the primers. The use of a master mix ensures that all reactions have identical reagents and it also reduces the number of transfers requiring micropipettes. The smaller number of transfers is particularly advantageous, because it reduces the opportunities for cross-contamination of reagents. PCR is an exquisitely sensitive procedure. Some researchers even use special barrier tips for their micropipettes, which contain filters that prevent samples from reaching the barrel of the micropipettes.

**Saccharomyces Genome Deletion Project**

The publication of the yeast genome sequence opened new opportunities for yeast geneticists. Knowing the DNA sequence of the yeast genome, geneticists could now take advantage of the high frequency with which yeast exchange genes by homologous recombination to generate mutants of their own design. Homologous recombination normally occurs during meiosis and during certain kinds of DNA repair. During homologous recombination, two closely related DNA sequences align with one another, breaks appear in the DNA molecules and strand exchange occurs when the breaks are repaired. Homologous recombination allows researchers to replace a chromosomal gene with a DNA construct of their own design. To use this strategy, investigators first construct a replacement cassette in which a marker gene is flanked by sequences that are identical to chromosomal DNA sequences at either side of the the yeast target gene. The sequence is then introduced into cells by chemical transformation (Chapter 12) or by
electroporation. Transformed cells that have incorporated the replacement cassettes into chromosomal DNA can be identified by selecting for the marker gene in the replacement cassette.

The *Saccharomyces* Genome Deletion Project (SGDP) used this approach to systematically replace each of the predicted ORFs in the *S. cerevisiae* genome with a kanamycin resistance (*KAN*R) gene. For each ORF, researchers used a series of PCR reactions (below) to construct cassettes in which the *KAN*R gene was flanked by short DNA sequences that occur upstream and downstream of the targeted ORF on the *S. cerevisiae* chromosome. The cassettes were then used to transform the BY4742 strain (Brachmann *et al.*, 1998). Strains that had incorporated the *KAN*R gene were selected on plates containing analogs of kanamycin (Winzeler *et al.*, 1999).

The *S. cerevisiae* strains that we are using in our experiments are part of the SGDP collection of deletion mutants. In this lab, you will use PCR to identify which MET genes have been disrupted in your strains. The PCR primers were designed and tested by the SGDP to verify strain identities. Available primers include two gene-specific primers (GSP) for each of the MET genes. One of the GSPs, GSP Primer A, is located 200-400 bp upstream of the initiation codon. The second GSP, GSP Primer B, is an antisense primer that binds within the ORF. We also have an antisense primer that binds 250 bp within the *KAN*R gene (KAN Primer B).
Chapter 7

Exercise 1 - Physical properties of PCR primers

The table below lists the primer pairs used by the SGDP to analyze the deletion mutants. In designing the primers, researchers aimed for primer pairs that had similar physical properties, i.e. length and $T_m$, and would generate PCR products that were several hundred base pairs long. Like most genomes, the *S. cerevisiae* genome has both AT-rich and GC-rich regions (Goffeau et al., 1996). Non-coding regions, in particular, tend to be enriched in AT base pairs, which are stabilized by fewer hydrogen bonds than GC pairs.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer A</th>
<th>Primer B</th>
</tr>
</thead>
<tbody>
<tr>
<td>MET1</td>
<td>TTCTATTTTCGTTATTTTTCG</td>
<td>AAATGAACCTGATCAGTAGCAAAAC</td>
</tr>
<tr>
<td>MET2</td>
<td>AAGTCATGTTAATCGTTTGAGTTTGT</td>
<td>GTCCAAGTATGTTGGGATCTGAGTAA</td>
</tr>
<tr>
<td>MET3</td>
<td>GTAATTTTGTAACTCACCGATCTTCTC</td>
<td>CATTCTTCTTTAACGACATCTCTAGC</td>
</tr>
<tr>
<td>MET5</td>
<td>TTCATACGTCCGATATATCTTCTTTC</td>
<td>GGTATTCAATGAGTCCTTGATGTCCT</td>
</tr>
<tr>
<td>MET6</td>
<td>ATGCCATGGATGCACTAATTTAAGG</td>
<td>AAAACCTTGGTCGTAAMAGGAGAGT</td>
</tr>
<tr>
<td>MET7</td>
<td>GTGGTTAAGAGAAGGCAACTA</td>
<td>TCAATCATTCCAAATGTCATAG</td>
</tr>
<tr>
<td>MET8</td>
<td>ATGCCATGGATGCACTAATTTAAGG</td>
<td>AAAACCTTGGTCGTAAMAGGAGAGT</td>
</tr>
<tr>
<td>MET10</td>
<td>AAAGAAGAACACTACATCAATCTTCCCA</td>
<td>AGTTTTAAGGACCCAACATCAAAAG</td>
</tr>
<tr>
<td>MET14</td>
<td>AAAGAATAGGGTCTTTTCATTTTCG</td>
<td>GATTGTACCTTTACCTGACCGCCTT</td>
</tr>
<tr>
<td>MET16</td>
<td>GCTGACAAAAGAATTGGATAAAAGA</td>
<td>ATATACTCTTTTACCTGACCGCCTT</td>
</tr>
<tr>
<td>MET25</td>
<td>CATCCCTAGAAACCATTTGTCACAT</td>
<td>GCAGAATGTGTACAAATATCAGCAC</td>
</tr>
<tr>
<td>CYS3</td>
<td>ACCCCATACACTCTCTCTTGTTAT</td>
<td>ATAGGTGTTAGCTGAGAAGATTGT</td>
</tr>
<tr>
<td>CYS4</td>
<td>ACAACTTCACCTCACCAAGAAGT</td>
<td>TCAAGTCTTCTAGCTGTTTGGAT</td>
</tr>
<tr>
<td>KAN$^R$</td>
<td></td>
<td>CTGCAGCGAGGAGCCGTAAT</td>
</tr>
</tbody>
</table>

Note: The primer sequences are written in Courier font, a nonproportional font that is often used for nucleic acids, because all letters have the same width. What is immediately apparent about the primer lengths?

In this lab, you will use the SGDP primers to identify your three deletion strains. Fill in the table below with information about the length, GC-content and $T_m$ for the primers designed for your strain. To find this information, use one of several online tools for primer analysis:

http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/
http://www.basic.northwestern.edu/biotools/oligocalc.html

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer A</th>
<th>Primer B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Length</td>
<td>% GC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KAN$^R$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

70
Yeast Colony PCR

**Exercise 2 - Predict the sizes of the PCR products**

In the next lab, you will analyze your PCR reaction products on an agarose gel that separates DNA molecules according to their sizes. To interpret those results, it will be important to have calculated the expected sizes of the PCR products from your reactions. To assist those calculations, prepare a simple map of the genomic region containing your gene with the primer sequences aligned against the genome sequence. All PCR products should contain a portion of the *MET* gene's 5'-flanking region because of primer A. A PCR product may or may not contain portions of the *MET* gene's CDS, depending on whether you are analyzing a strain with the native or disrupted *MET* gene. To construct your map, you can take advantage of special genome sequence records prepared by SGD curators. These records contain the CDS for your *MET* gene together with 1 kb of upstream and 1 kb of downstream sequence. SGD curators generated these records because researchers are often interested in studying regulatory elements that control transcription of a gene and the processing of gene transcripts. In *S. cerevisiae*, these regulatory elements are usually located within 1 kb of the CDS.

The figure on the right shows a map for the *SAM1* gene. The binding site for primer A is 357 nucleotides upstream of the *SAM1* initiation codon (nucleotide 1001). Primer B-anchored products add 280 bp of CDS to the PCR product. The expected size of the PCR product is 357 + 280 bp, or 637 bp. If the deletion strain had been used for PCR, the *SAM1* primers A and B would not generate a PCR product. Instead, *SAM1* primer A and *KAN^R* primer B would generate a 607 bp (357 + 250) product, because the *KAN^R* primer B binds to nucleotides 231-250 of the *KAN^R* CDS.

You will need two browser windows for this exercise. Each member of the group should work with a single gene.

**Find the genomic sequence for your gene.**
- Navigate to your gene's summary page in the SGD (yeastgenome.org)
- Click the Sequence tab at the top of the summary page.
- Cursor down to the gene sequence for S288C. Select "Genomic sequence +/- 1kb" from the dropdown box.
- Note below the starting and ending coordinates for the sequence and calculate the length of the sequence. (You should see the ATG start codon at nucleotides 1001-1003.)

Length of sequence (bp) ___________ Length of the coding sequence ___________
Align the primer sequences with the genomic sequence.

To find the position on the primers in the genomic sequence, we will use NCBI's BLAST tool. BLAST stands for Basic Local Alignment Search Tool and can be used to align either protein or nucleic acid sequences. You will learn more about the BLAST algorithms in Chapter 9.

- Direct your browser to the NCBI site and select BLAST from the list of resources on the right.
- Select Nucleotide BLAST from the list of Basic BLAST programs.
- Click the box “Align two or more sequences.” Copy the “genomic sequence +/- 1kb” from SGD and paste the sequence into the lower Subject Sequence box.
- Type the Primer A sequence for your gene in query box.
- Adjust the BLAST algorithm for a short sequence. The primer sequences that we are using are 25 nucleotides long. This is shorter than the default value of 28 for “words” in BLASTN (the algorithm for comparing nucleotide sequences). BLAST will not align two sequences if the match is smaller than 28 nucleotides. Expand the “algorithm parameters” at the bottom of the page. Select a “word size” less than 28.
- Click BLAST. The BLAST results bring up a table that shows each match between your primer and the genome sequence. The top result should be a perfect match between your primer and the genome sequence. (Check your typing if it isn’t a perfect match!) Record the starting and ending nucleotides in the genomic DNA sequence where it matches the primer sequence.

- Repeat the BLAST alignment for primer B. Click “Edit and Resubmit” at the top of the BLAST results page. Clear the query box and type in the sequence of primer B. Click BLAST and record the alignment results. In the results, note that the primer nucleotide numbers are ascending, while the genomic DNA nucleotide numbers are in descending order. This is because Primer B sequence is the reverse complement of the gene sequence.

Draw a map of your gene and primer binding sites in the space below. Include the start codon and distances in bp.

Calculate the sizes of the PCR products that would be generated with:
Primer A and Primer B
Primer A and KAN<sup>78</sup> primer B
**Exercise 3 - Design yeast colony PCR**

In this lab, each team will be able to perform six PCR reactions to identify your three *S. cerevisiae* strains. You will have the option of using any of the MET primers listed above, as well as the KANₐ primer B. It will not be possible to test every strain with both the GSPA-GSPB and GSPA-KANₐ B combinations. Use your results from the selective plating experiment as you work with your team to devise a strategy that will allow you to positively identify your *met* strains.

List the primer pairs that you will use for the reactions, together with the predicted sizes of the PCR products in your notebook.

**Exercise 4 - Yeast colony PCR**

To prepare the reactions, you will first mix the primer pairs with a very small number of yeast cells that you transfer from a colony to the tube with the tip of a P-20 or P-200 micropipette. The colony and primers will then be heated at 98˚C for 15 minutes to disrupt the yeast cells. At that point, you will add an equal volume of a PCR master mix, containing nucleotides and the Taq polymerase, to each tube. The tubes will then be returned to the thermocycler for a typical PCR reaction.

1. **Label the PCR tubes.** The tubes are very small, so develop a code that you can use to identify the tubes. (Don't forget to include the code in your notebook. The code should indicate which primers and strains are mixed in each tube.)

2. **Prepare the primer mixtures.** The final volume of the PCR reactions will be 20 µL. The primer mixture accounts for half the final volume, or 10 µL. The primers stock concentrations are 2.0 µM each. Pipette 5.0 µL each of the two primers that you would like to use into each PCR tube. What will the final concentration of each primer be in the actual PCR reaction?

   NOTE: Because of the extraordinary sensitivity of PCR reactions, it is very important not to cross-contaminate tubes with unintended primers. Change tips between every primer transfer.

3. **Transfer a small quantity of yeast cells to each PCR tube.** Lightly touch the tip of a P20 or P200 micropipette to a yeast colony. Twirl the micropipette tip in the tube containing your primer mix to release the cells. The most common error is transferring too many yeast cells, which will interfere with the PCR reaction. The right amount of yeast will fit on the tip of a pin.

4. **Lyse the yeast cells.** Place the tubes in the thermocycler for 15 min at 98˚C.

5. **Set up the PCR reactions.** Remove the tubes from the thermocycler and add 10 µL of PCR master mix to each tube.

6. **Amplify the target gene sequences.** Return the tubes to the thermocycler and start the PCR program.
Chapter 7

Our thermocyclers are set for the following program:

- One cycle of denaturation: 95°C for 2 minutes
- 35 cycles of denaturation, annealing and extension:
  - 95°C for 30 sec.
  - 55°C for 30 sec.
  - 72°C for 1 minute
- One cycle of extension: 72°C for 10 minutes

Test yourself

1. A researcher plans to amplify the CDS of the SAM1 gene. The partial sequence below shows the coding sequences for the N-terminus and C-terminus of Sam1p. Design two 18-nucleotide long primers to amplify the sequence the SAM1 CDS.

   5'-ATGGCCGGGTACATTTTTATTC...........(CDS).......TCCAAGACTTTGAAGTTCTAA - 3'

2. Most of the primers that we are using for PCR have Tₘs that are slightly lower than the annealing temperature used in the PCR reactions. Thus, less than half of the target sites are expected to anneal with the primers in each cycle. Why would investigators choose to design primers that will not be fully annealed with the DNA template?

References


