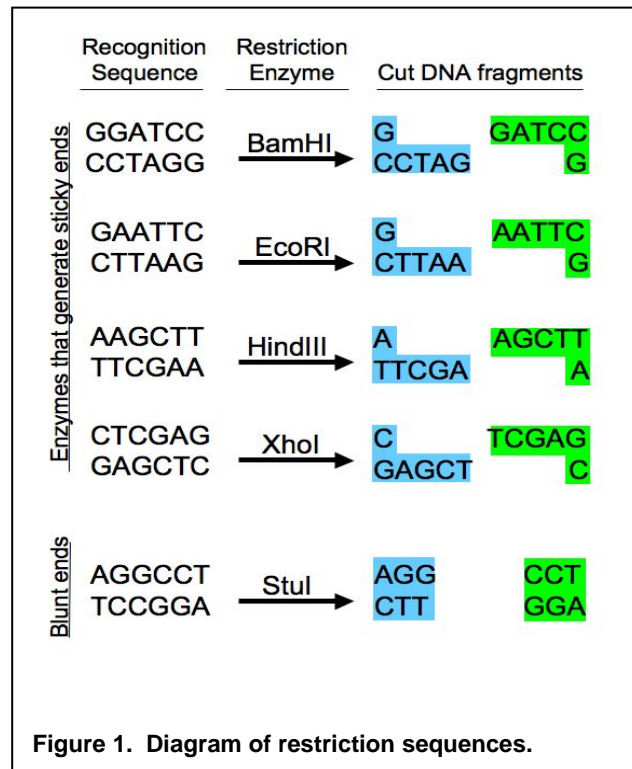


Exploring Genetic Variation in a Caffeine Metabolism Gene Part Two: RFLP and SNP Analysis

Background

Restriction endonucleases (or **restriction enzymes**) are some of the most powerful tools in modern molecular biology. These enzymes, produced by bacteria, interact with **recognition sequences**, highly specific nucleotide sequences in DNA molecules (Figure 1). They then cut both strands of the double-stranded DNA, creating DNA restriction fragments. Because the recognition sequence is very specific and different DNAs will have different numbers of these sequences, digestion of different DNAs with the same enzyme may yield differently sized restriction fragments. This restriction analysis can be used as a fingerprint to distinguish DNA molecules from each other. To date, more than 2500 restriction enzymes have been isolated, and many are commercially available, offering a wide variety of options for cutting DNA molecules at different sites.



In this lab, we will digest your PCR product of the amplified CYP1A2 gene sequence and examine it by using certain restriction enzymes. This will generate DNA fragments of a certain length, called **restriction fragment length polymorphisms (RFLPs)**. We can analyze this data to determine the genotype indicated and therefore make predictions about the phenotype. You will then be able to determine whether your model data set represents a homozygote or a heterozygote and whether this indicates that they will metabolize caffeine quickly or slowly.

Safety Note

Ethidium bromide is a carcinogen that we will use to stain the DNA in the gel. When ethidium bromide is visualized under UV light, it fluoresces orange. Make sure to wear gloves when working with ethidium bromide.

Protocol

1. In the interest of time, we will assume that your PCR worked and proceed with the next step of setting up a restriction digest. Fill in the columns on the right in the chart below to set up your restriction digests for a total volume of 30µL each.

Component	Stock Concentration	Final Concentration	Sacl Digest	Apal Digest	No enzyme
PCR			15µL	15µL	15µL
CutSmart Buffer	10X	1X			
Water					
Enzyme: Sacl	10 units/µL	5 units			
Enzyme: Apal	10 units/µL	5 units			

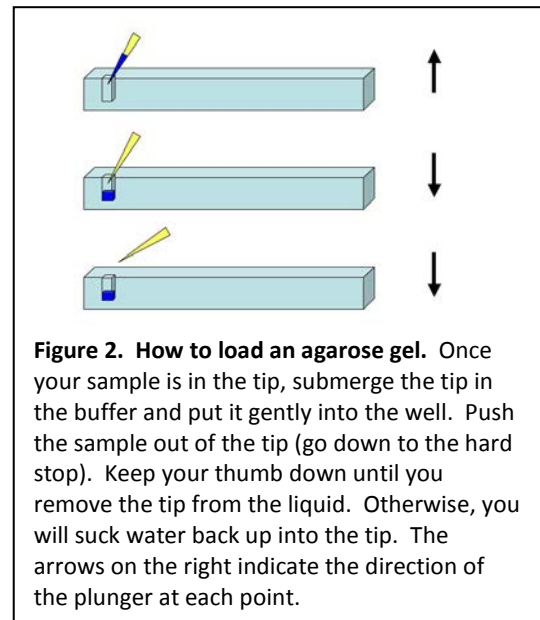
2. Set up your digests (or negative control) in 0.65 mL microfuge tubes. Label them clearly with your initials and indicate them as either Sacl or Apal or no enzyme. Gently flick the tube to mix, centrifuge briefly, and incubate the Sacl digest at 37°C for 1 hr and the Apal and no enzyme tubes at room temperature for 1hr.

3. While the digest is incubating, prepare a 1.2% agarose gel (100mL in 1X TAE) with 1:20000 ethidium bromide and 1L of 1X TAE. Make sure to wear gloves when using ethidium bromide.

4. When the digest is complete, add 6µL of DNA loading dye to each tube and pipet up and down to mix. Load 15µL of each sample into a different well of the gel in the following order: no enzyme; Sacl; Apal. Make sure to load 5µL 100bp ladder

(<https://www.neb.com/products/n3231-100-bp-dna-ladder>) as well. See the figure 1 if you need help.

5. Run the gel at 150 volts until the blue dye is about $\frac{3}{4}$ down the gel. Image the gel on the UV transilluminator.



Concept Check:

Below is the CYP1A2 sequence. Recall that there are two CYP1A2 alleles that differ in sequence as the result of a Single Nucleotide Polymorphism (SNP). The underlined and bolded base at position 494 indicates the SNP in this allele.

Allele 1

```
GGA GAGCGATGGG GAGGGCTTCC CCCTTAGCAC AAGGGCAGCC CTGGCCCTGG      53
CTGAAGCCCA ACCCCAACCT CCAAGACTGT GAGAGGATGG GGACTCATCC CTGGAGGAGG 113
TGCCCTCCT GGTATTGATA AAGAATGCC TGGGGAGGGG GCATCACAGG CTATTTGAAC 173
CAGCCCTGGG ACCTTGCCA CCTCAGTGTC ACTGGGTAGG GGGAACCTCT GGTCCCTTGG 233
GTATATGGAA GGTATCAGCA GAAAGCCAGC ACTGGCAGGG ACTCTTTGGT ACAATACCCA 293
GCATGCATGC TGTGCCAGGG GCTGACAAGG GTGCTGTCCT TGGCTTCCCC ATTTTGGAGT 353
GGTCACTTGC CTCTACTCCA GCCCCAGAAG TGGAAACTGA GATGATGTGT GGAGGAGAGA 413
GCCAGCGTTC ATGTTGGGAA TCTTGAGGCT CCTTCCAGC TCTCAGATTC TGTGATGCTC 473
AAAGGGTGAG CTCTGTGGGC A 494
CAGGACGCAT GGTAGATGGA GCTTAGTCTT TCTGGTATCC AGCTGGGAGC CAAGCACAGA 554
ACACGCATCA GTGTTTATCA AATGACTGAG GAAATGAATG AATGAATGTC TCCATCTCAA 614
CCCTCAGCCT GGTCCCTCCT TTTTCCCTG CAGTTGGTAC AGATGGCATT GTCCAGTCT 674
GTTCCCTTCT CGGCCACAGA GCTTCTCCTG GCCTCTGCCA TCTTCTGCCT GGTATTCTGG 734
GTGCTCAAG 743
```

The SNP site is variable, of course, and the alternate allele of CYP1A2 has a C nucleotide in place of an A. This alternate allele sequence is shown below.

Allele 2

```
GGA GAGCGATGGG GAGGGCTTCC CCCTTAGCAC AAGGGCAGCC CTGGCCCTGG      53
CTGAAGCCCA ACCCCAACCT CCAAGACTGT GAGAGGATGG GGACTCATCC CTGGAGGAGG 113
TGCCCTCCT GGTATTGATA AAGAATGCC TGGGGAGGGG GCATCACAGG CTATTTGAAC 173
CAGCCCTGGG ACCTTGCCA CCTCAGTGTC ACTGGGTAGG GGGAACCTCT GGTCCCTTGG 233
GTATATGGAA GGTATCAGCA GAAAGCCAGC ACTGGCAGGG ACTCTTTGGT ACAATACCCA 293
GCATGCATGC TGTGCCAGGG GCTGACAAGG GTGCTGTCCT TGGCTTCCCC ATTTTGGAGT 353
GGTCACTTGC CTCTACTCCA GCCCCAGAAG TGGAAACTGA GATGATGTGT GGAGGAGAGA 413
GCCAGCGTTC ATGTTGGGAA TCTTGAGGCT CCTTCCAGC TCTCAGATTC TGTGATGCTC 473
AAAGGGTGAG CTCTGTGGGC C 494
CAGGACGCAT GGTAGATGGA GCTTAGTCTT TCTGGTATCC AGCTGGGAGC CAAGCACAGA 554
ACACGCATCA GTGTTTATCA AATGACTGAG GAAATGAATG AATGAATGTC TCCATCTCAA 614
CCCTCAGCCT GGTCCCTCCT TTTTCCCTG CAGTTGGTAC AGATGGCATT GTCCAGTCT 674
GTTCCCTTCT CGGCCACAGA GCTTCTCCTG GCCTCTGCCA TCTTCTGCCT GGTATTCTGG 734
GTGCTCAAG 743
```

1. The restriction enzyme *SacI* has the recognition sequence GAGCTC. Find this site in both allele sequences above and highlight it. (Hint: this is not far from the SNP.)

2. SacI cuts between the T and the last C.

A. If SacI cuts at this site, how many pieces of DNA do you get? _____

B. How big is each piece? _____ bp & _____ bp

3. The restriction enzyme Apal cuts at the sequence GGGCCC. Look for this sequence at the SNP site in Allele 2 and highlight it with a different color. Notice that this site does NOT occur in Allele 1 due to the SNP variation.

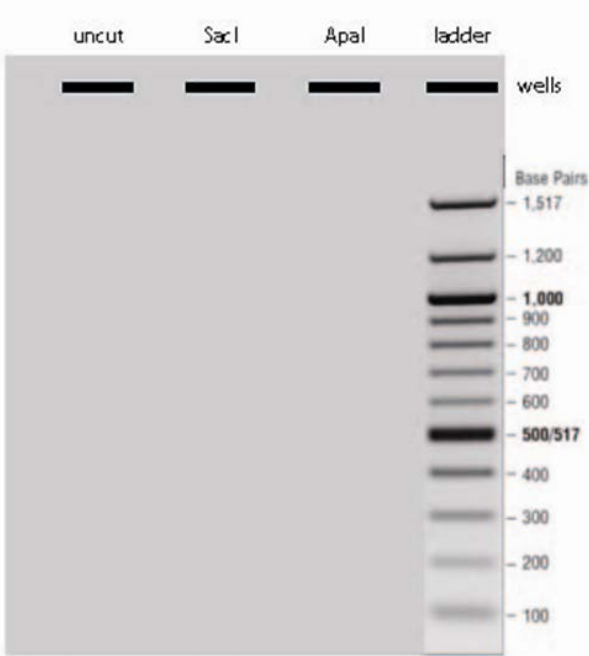
A. If Apal ***only*** is used to digest allele 2, how many pieces of DNA do you get? _____

B. Apal cuts between the penultimate and final C of its restriction site. How big will each piece be?

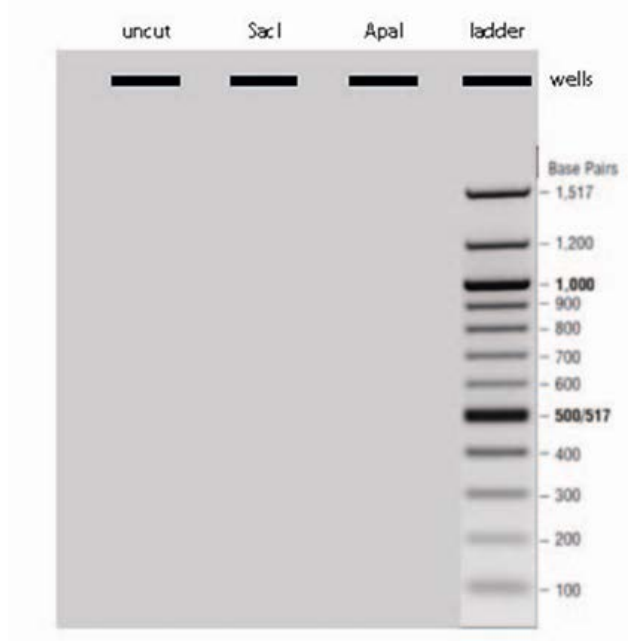
_____ bp & _____ bp

4. Everyone has two CYP1A2 alleles (a maternal copy and a paternal copy). Thus, individuals may be homozygous for either allele 1 or 2, or heterozygous. Your PCR samples should contain an equal mix of amplified CYP1A2 sequences from both your paternal and maternal alleles. Below are three agarose gel diagrams. Sketch the positions (sizes) of CYP1A2 restriction fragments that you would expect to observe based on the possible genotypes indicated.

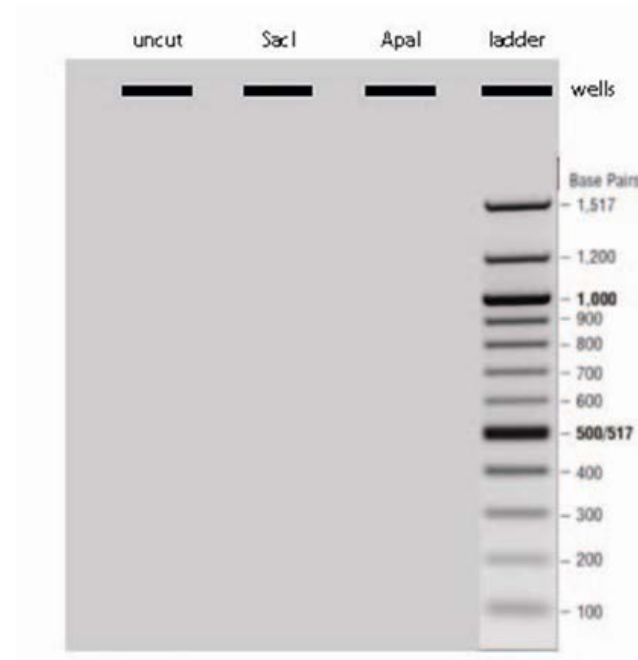
Homozygous Allele 1



Homozygous Allele 2



Heterozygous



Analysis

Finally, it is time to determine whether you are a fast or slow caffeine metabolizer. Those who have are homozygous for Allele 1 are considered fast metabolizers, and the amount of caffeine consumed should have no effect on their cardiac health. Those who have a C in either allele are considered slow metabolizers, and consumption of high amounts of caffeine (greater than two cups a day) increases their risk for a non-fatal heart attack (Cornelis MC et al. (2006) Coffee, CYP1A2 genotype, and risk of myocardial infarction. *JAMA* 295(10):1135-41).

- A. Based on your personal experience with caffeine, what do you predict your phenotype will be, fast or slow? _____
- B. Does your data set indicate that you are a fast or a slow metabolizer of caffeine?

- C. Are you a homozygote or heterozygote? _____
- D. Based on the information presented, which phenotype would you predict is dominant, fast or slow? _____
- E. Which is the dominant allele (C or A)? _____
- F. Which is the recessive allele (C or A)? _____

Thought Questions

In this laboratory, we used the restriction enzyme *Apal* to distinguish the alleles and identify the SNP. How would you distinguish SNPs that do not result in a restriction enzyme recognition sequence?

In this laboratory, we amplified an intron within the *CYP1A2* gene. The A allele has been correlated to an increase in protein expression of *CYP1A2* (Sachse C et al. (1999) Functional significance of a C→A polymorphism in intron 1 of the cytochrome P450 *CYP1A2* gene tested with caffeine. *Br J Clin Pharmacol.* 47(4):445-9). Hypothesize ways that non-coding DNA might influence gene expression.

We analyzed one gene in this laboratory, but it is now relatively simple and affordable to sequence whole human genomes. What do you think will be the impact of having this kind of genetic information readily available? What types of traits should test for, and when should we do it?