



Laboratory Exercise: Exploring Genetic Variation in a Caffeine Metabolism Gene

Yvelande Zephyr and Susan Walsh¹
¹Rollins College
1000 Holt Ave - 2743
Winter Park, FL 32789

Corresponding Author: sjwalsh@rollins.edu

Accepted for publication March 9, 2015

Citation:

Zephyr, Yvelande and Walsh, Susan (2015). Exploring Genetic Variation in a Caffeine Metabolism Gene. *Genetics Society of America Peer-Reviewed Education Portal (GSA PREP)*: 2015.001; [doi:10.1534/gsaprep.2015.001](https://doi.org/10.1534/gsaprep.2015.001)

Synopsis:

This two-week lab project for majors demonstrates the core concept of genetic variation by having students analyze a single nucleotide polymorphism (SNP) from their own DNA. Briefly, DNA is isolated from cheek cells, and a specific region of the *CYP1A2* intron is amplified via PCR. Restriction fragment length polymorphism (RFLP) analysis is then employed using the enzyme *Apal* to detect the SNP, and *SacI* serves as a control for the digestion. Using agarose gel electrophoresis, each genotype can be distinguished by the size of the RFLPs. To verify the technique, unknown data are compared against individuals reported to be homozygous A, homozygous C, or heterozygous by the direct-to-consumer (DTC) genetic testing company 23andme. Students can generate a hypothesis based on their own behavior, test it, and interpret the results. RFLP is a more cost-effective technique than sequencing-based laboratories. Importantly, this exercise can be scaled up or down to meet the needs of majors and nonmajors alike. Overall, this series of laboratories exposes students to a variety of basic molecular biology techniques, correlates to an observable phenotype without causing serious health concerns, and caters to students' innate desire for self-discovery.

Introduction:

Students are intrinsically interested in learning more about themselves, and this particular series of laboratories addresses precisely that, without the complications of testing for a trait that could cause immediate concerns about health. Approximately 90% of adults regularly consume caffeine, making it the most popular stimulant in the world (Frary *et al.*, 2005; Fredholm *et al.*, 1999; Gonzalez de Mejia and Ramirez-Mares, 2014). As caffeine is a routine part of many undergraduates' lives, they can immediately make predictions about its effects on their behavior.

The physiological effects of caffeine are partially mediated by the cytochrome P450 (CYP) genes (Butler *et al.*, 1989). Specifically, the main enzyme associated with the demethylation of caffeine, the first major step in metabolizing caffeine in humans, is the liver enzyme cytochrome P450 1A2 (*CYP1A2*) (Butler *et al.*, 1989). The *CYP1A2* gene is located on chromosome 15q24.1 (Sachse *et al.*, 1999). Within intron 1 of *CYP1A2*, there is a single nucleotide polymorphism (SNP), rs762551, that is either an A or a C (Chida *et al.*, 1999; Sachse *et al.*, 1999). Fast metabolizers are homozygous A, while those who are homozygous C or heterozygous are slow metabolizers (Sachse *et al.*, 1999). In this lab, students can easily distinguish the two alleles using restriction fragment length polymorphisms (RFLPs) by cutting with the enzyme *Apal*. *Apal* cuts only if a C is present in the allele, yielding a distinct profile for each genotype.

Students first isolate their own buccal cells using phosphate buffered saline as a mouthwash, and then proceed to isolate genomic DNA through detergent cell lysis, Proteinase K digestion, phenol extraction, and alcohol precipitation (Koppikar and Mulherkar, 2006). Depending on the level of the students, some of these steps are best done by a teaching assistant or instructor, but upper level genetics students have handled the entire process themselves during a single

four-hour lab period. As an alternative, genomic DNA has been successfully isolated in a shorter time span using DNAzol or a genomic DNA kit (see alternate procedures). The genotyping is then determined by using PCR followed by restriction digestion (Figure 1). Conveniently, the SNP in question contains an Apal sequence to distinguish the two different alleles (only those with a C will be cleaved) and a SacI site near the SNP that will be cleaved regardless. In this way, the students have a positive control for their restriction digestion. The advantage of RFLPs over other genotyping experiments is that this experiment will not require the additional cost and preparation required for sequencing. In order to best prepare students for the laboratory, they are provided with a video (<http://www.hhmi.org/biointeractive/polymerase-chain-reaction-pcr>) and worksheet to depict PCR. To explain the RFLP experiment, students are given the sequences of both alleles and asked to identify the restriction sites for both Apal and SacI, describe the sizes that will occur when the enzymes do or do not cut, and then draw out the gels for homozygotes and a heterozygote. In this way, students can form an educated guess about the variety of results they may achieve and make predictions based on their own personal experiences after consuming caffeine.

Approach/Method (Instructor Guidelines):

Introduction of Lab:

The method described below has been performed in two four-hour labs for majors. However, it has also been completed over multiple shorter labs for nonmajors. This laboratory sequence reinforces molecular biology techniques (DNA isolation, PCR, and restriction digestion) and genetic variation (homozygote versus heterozygote and SNPs). There are periods of down time during the laboratory where worksheets modeling the experiment on paper are utilized.

Suggestions and specific instructions for each lab protocol follow:

Lab 1: Genomic DNA Purification from Buccal Cells and PCR:

DNA Isolation Procedure 1:

Although some protocols suggest using kits to isolate genomic DNA (see alternate procedures), we have consistently isolated genomic DNA successfully using a longer, more traditional and affordable method (Koppikar and Mulherkar, 2006). In brief, this procedure is as follows. Before eating a meal, participants were asked to vigorously swish in approximately 15 mL sterile PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, pH 7.4) in their mouths for at least 60 seconds while chewing on their cheeks to isolate as many cells as possible. The cells were centrifuged at 4000xg for 5 minutes. Afterward, the cells were washed with 20 mL fresh PBS and centrifuged again. They were then resuspended in 1 mL PBS, transferred to a microfuge tube, and spun at 4000xg for 5 minutes at room temperature. The cells were resuspended in 0.5 mL lysis buffer (20 mM Tris-HCl pH 8, 50 mM NaCl, 10 mM EDTA, 0.5% SDS, and 100 µg/mL Proteinase K) and incubated at 50°C for 1 hour. For students in

nonmajors courses, the rest of the DNA isolation was done for them. The cell lysate was extracted with an equal volume of phenol, then extracted with phenol:chloroform, and finally chloroform. Sodium acetate (3 M, pH 5.5) was added at 1/10 the volume of the cell lysate, and an equal volume of isopropanol was added. The samples were centrifuged at 14000xg for 20 minutes at 4°C. The DNA pellets were washed with cold 70% ethanol, and the pellets allowed to air dry for 5 minutes before resuspending in 30 μ L 10 mM Tris-HCl pH 8. DNA was quantified by absorbance at 260 nm using a nanodrop and diluted to 10 ng/ μ L in 10 mM Tris-HCl pH 8. Rate-limiting steps in this procedure in terms of numbers of students are the number of microfuges and the nanodrop. Yield typically ranges from 100 ng/ μ L to 2 μ g/ μ L.

Alternate DNA Isolation Procedures:

2. DNAzol (Invitrogen Cat #10503-027) can be used according to the manufacturer's instructions after pelleting the cells in a 2 mL microfuge tube. In brief, 1 mL of DNAzol was added to the cell pellet, which was resuspended by pipetting gently with a wide-bore pipet tip before adding 0.5 mL 100% ethanol. After inverting, the DNA is visible in the supernatant and can be spooled and collected into a new tube. The DNA is washed with 75% ethanol twice, briefly dried, and resuspended in 0.2 mL 8 mM NaOH. The sample is then neutralized by adding 6.4 μ L of 1 M HEPES. Yield is approximately 10 ng/ μ L. This procedure can be accomplished in approximately 30 minutes.

3. The Wizard Genomic DNA Purification Kit (Promega Cat #A1120) can be used according to the manufacturer's instructions after pelleting the cells in a microfuge tube. The only exception was to resuspend the DNA pellet in 10 mM Tris pH 8 in the rehydration step. Yield is approximately 25 ng/ μ L. This procedure can be accomplished in approximately 90 minutes, if the rehydration step is shortened.

PCR:

Primers (forward primer: 5'- GAGAGCGATGGGGAGGGC; reverse primer: 5'- CCCTTGAGCACCCAGAATACC) were designed against SNP rs762551. PCRs contained 50 ng of template DNA, 1.5 mM MgCl₂, 0.2 mM dNTPs, 1.25 units of Taq DNA polymerase (New England Biolabs) in a total volume of 50 μ L. Students set up a cocktail for the class of all components except their DNA. Nonmajors typically assembled this PCR during an additional lab period, starting with diluting their own DNA to 10 ng/ μ L, and the reagent cocktail was made for them. PCR parameters were 94°C, 5 min; 94°C 30 sec, 58°C 30 sec, 68°C, 1 min for 35 cycles; 68°C 5 min. Samples were stored in the freezer until the following lab period.

Optional Analysis:

Students are asked to analyze the ingredients of the cell lysis buffer and describe the purpose of each. In addition, they are asked to calculate the melting temperature (T_m) of the primers and explore the SNP website.

Lab 2: RFLP of PCR:

After the PCR, the 15 μ L of the product was divided into three tubes, and each was digested using either SacI (New England Biolabs), Apal (New England Biolabs), or without enzyme in a total volume of 30 μ L. The SacI digest was incubated at 37°C for 1 hour, and the Apal digest was incubated at 25°C for 1 hour. Loading dye was added, and 20 μ L of the reactions were electrophoresed on a 1.2% TAE-agarose gel containing ethidium bromide and compared to a 100 bp DNA ladder (New England Biolabs) using a UV transilluminator. The amount of time for digestion can be reduced if necessary, provided that there is adequate enzyme to complete the reaction quickly. By resolving only part of the sample, students have additional material to run if a mistake occurs. Depending on the length of the lab period and the skill level of the students, the restriction digestion and gel electrophoresis can be divided into two lab periods.

Optional Analysis:

Students can analyze the primer sequences using a nucleotide BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) against the human genomic + transcript to make predictions about other amplified products that appear in their samples. They can then design primers *in silico* that they predict will be more specific to this region. Similarly, students can perform the PCR along a temperature gradient from 50°C to 60°C to demonstrate the specificity that is generated from higher annealing temperatures.

Students should hypothesize how a mutation in an intron could lead to a phenotypic change. In that sense, this exercise dissolves the antiquated theory that introns are merely “junk DNA” and forces students to think about eukaryotic gene regulation and protein function beyond the coding region. Specifically, the A allele is correlated to increased expression under certain conditions in certain haplotypes (Sachse *et al.*, 1999; Akillo *et al.*, 2003).

The SNP used in this experiment was correlated to phenotype in those of European or Costa Rican ancestry (Cornelis *et al.*, 2006; Sachse *et al.*, 1999). Students can debate whether this SNP will be predictive across those of different ethnic backgrounds and design experiments to test their hypothesis. They can also explore primary literature by reading about additional SNPs that have been implicated in caffeine consumption (rs2472297 (Sulem *et al.*, 2011) and rs4410790 (Cornelis *et al.*, 2011)), reinforcing the concept that complex traits are often mediated by small effects of multiple genes.

Since this RFLP can only technically distinguish between a C or any other nucleotide at that position (yielding two alleles), class data could be used to calculate basic Hardy-Weinberg. Multiple class data sets could be used to compare changes in Hardy-Weinberg across multiple populations. Alternatively, class data could be compared to the appropriate published population diversity data on the SNP website (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=rs762551).

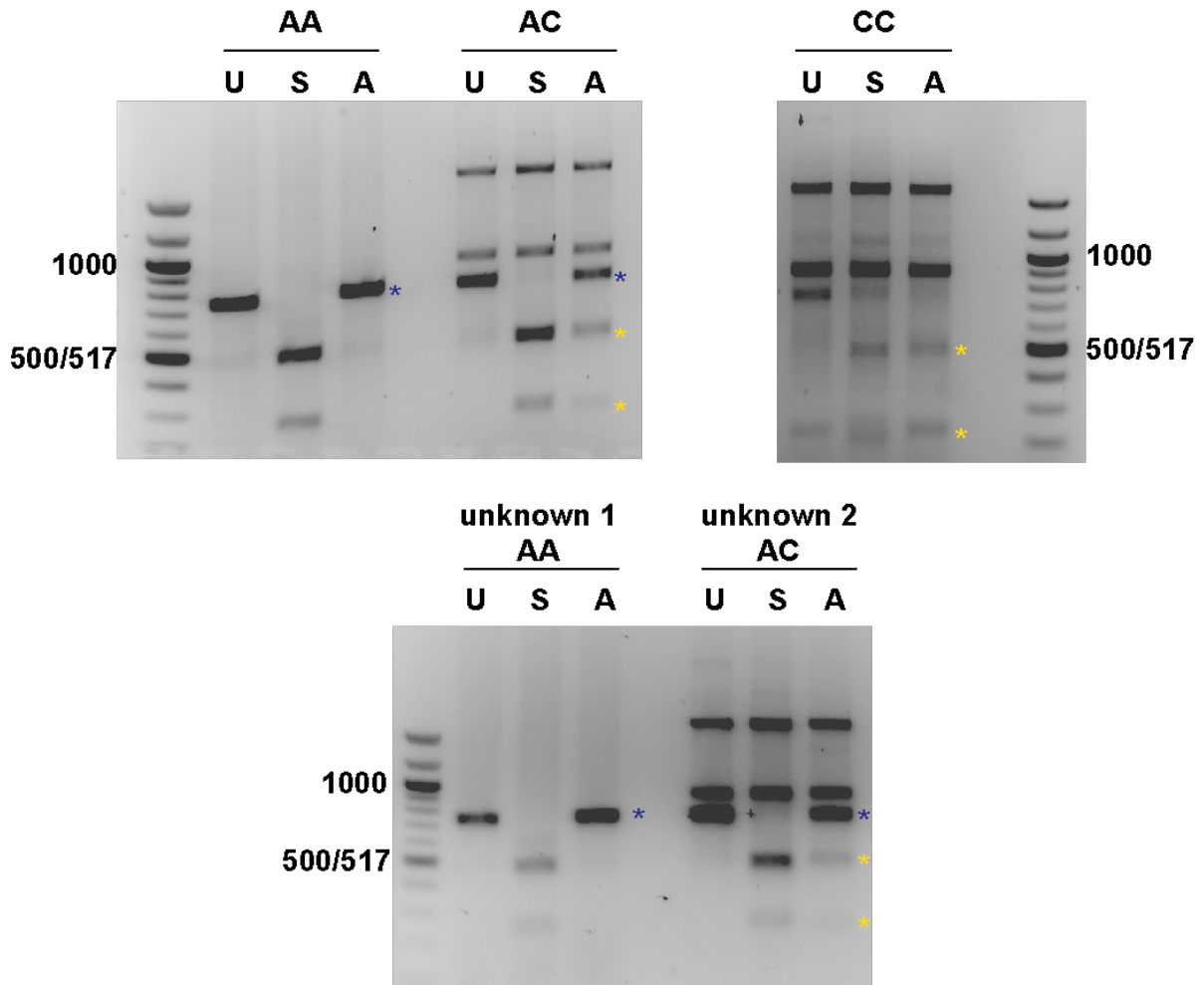


Figure 1. PCR-RFLP analysis of liver enzyme cytochrome (CYP1A2) SNP rs762551 using the restriction enzymes Apal and Sacl. DNA was isolated from human cheek cells and separated by gel electrophoresis on a 1.2% TAE-agarose gel. Apal (A) did not digest alleles containing an A, resulting in 743 bp (blue asterisks), but Apal did digest alleles containing a C (494 and 249 bp; yellow asterisks). Sacl (S) was used as a control restriction digest, and uncut PCR (U) was also run. Comparison is to a 100bp ladder. The top shows individuals sequenced by 23andme and their corresponding genotypes, and the bottom shows two individuals with unknown genotypes.

Table 1. Genotype dataset from IRB-approved subjects.

AA	AC	CC	Total
5	2	1	8

Lab Safety:

These labs require instruction on how to safely work with Proteinase K, phenol, chloroform, ethidium bromide and UV light. Students should wear appropriate protective eyewear, gloves, and lab coats during these procedures. In addition, phenol and chloroform should be used with adequate ventilation, preferably in a laboratory fume hood. Chemical waste (phenol, chloroform, and ethidium bromide) should be disposed of properly. Tubes containing human cells should also be treated as a biohazard.

Student Evaluation:

Depending on the level of students and the quality of the data, the students were asked to either write a full laboratory paper including primary literature sources or to simply fill out a list of questions identifying themselves as a heterozygote or homozygote and as a slow or fast metabolizer of caffeine. Majors were also required to keep their procedure, data, and conclusions in their lab notebooks, which are collected throughout the semester to measure completeness.

Justification:

Students are engaged in this particular laboratory because of their own desire to know more about themselves and explain their own behavior. In the process, they master molecular biology techniques and witness the core concept of genetic variation within a human population. They identify their own genotype for one particular trait using RFLP and see that a single nucleotide change can have an effect on their phenotype. Additionally, the procedure highlights the concept of the nature of genetic material and leads to hypotheses about how genes are inherited.

The core competencies covered in this lab directly pertains to the application of the process of science: observational strategies, hypothesis testing, experimental design, and evaluation of experimental evidence. Students formulate hypotheses based on their own behavior and confirm their predictions through analysis of their genotype. Students also reinforce communication skills formally through scientific writing. Informally, we have found that students like to talk about their own genotype with their friends and families, allowing for communication at different levels to various audiences. In particular, students may ask their parents about their caffeine metabolism, making inferences about their parents' genotypes and transmission genetics. The ease of this experiment allows for a larger discussion on genetic testing, in terms of what should be tested for, when genetic testing should occur, and how tightly genotype matches phenotype. These questions all feed into examining the relationship between science and society.

Acknowledgements:

Data presented in this paper were contributed from willing participants via a protocol approved by the Institutional Review Board at Rollins College. We would like to thank the American Society of Cell Biology and COMPASS for providing a grant to purchase the reagents necessary to do this experiment in collaboration with the Upward Bound program.

References:

- Aklillu E, Carrillo JA, Makonnen E, Hellman K, Pitarque M, Bertilsson L, Ingelman-Sundberg M. Genetic Polymorphism of CYP1A2 in Ethiopians Affecting Induction and Expression: Characterization of Novel Haplotypes with Single-Nucleotide Polymorphisms in Intron 1. *Molecular Pharmacology* 2003;64(3):659–69.
- Butler MA, Iwasaki M, Guengerich FP, Kadlubar FF. Human cytochrome P-450PA (P-450IA2), the phenacetin O-deethylase, is primarily responsible for the hepatic 3-demethylation of caffeine and N-oxidation of carcinogenic arylamines. *Proc Natl Acad Sci U S A.* **1989;86:7696-7700.**
- Chida M, Yokoi T, Fukui T, Kinoshita M, Yokota J, Kamataki T. Detection of three genetic polymorphisms in the 5'flanking region and intron 1 of human CYP1A2 in Japanese population. *Jpn J Cancer Res*1999;90:899–902.
- Cornelis MC, El-Sohehy A, Kabagambe EK, Campos H. Coffee, CYP1A2 genotype, and risk of myocardial infarction. *JAMA* 2006;295:1135–41.
- Cornelis, M.C. et al. 2011. Genome-Wide Meta-Analysis Identifies Regions on 7p21 (AHR) and 15q24 (CYP1A2) As Determinants of Habitual Caffeine Consumption. *PLoS Genet.* 7(4): e1002033.
- Frary CD, Johnson RK, Wang MQ. Food sources and intakes of caffeine in the diets of persons in the United States. *J Am Diet Assoc.* 2005;105(1):110-113.
- Fredholm BB, Bättig K, Holmén J, Nehlig A, Zvartau EE. Actions of caffeine in the brain with special reference to factors that contribute to its widespread use. *Pharmacol Rev.* 1999;51:83-133.
- Gonzalez de Mejia E, Ramirez-Mares MV. Impact of caffeine and coffee on our health. *Trends Endocrinol Metab.* 2014;25(10):489-492.
- Koppikar, P. and Mulherkar, R. 2006. A simple method for extraction of high molecular weight genomic DNA from buccal cells in mouthwash. *Indian Journal of Biotechnology* 5: 477-481.
- Sachse C, Brockmöller J, Bauer S, Roots I. Functional significance of a C-->A polymorphism in intron 1 of the cytochrome P450 CYP1A2 gene tested with caffeine. *Br J Clin Pharmacol.* 1999;47(4):445-9.
- Sulem, P. et al. 2011. Sequence variants at CYP1A1–CYP1A2 and AHR associate with coffee consumption. *Hum Mol Genet.*20(10): 2071–2077.