Polymorphisms

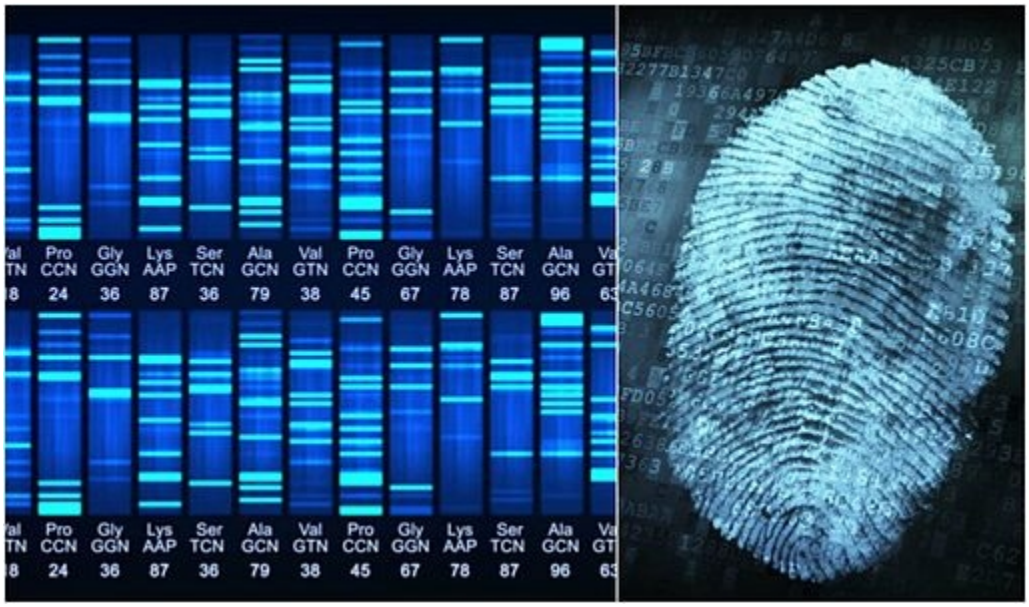
The difference in nucleotide sequences between humans lies between 0.1-0.4%. That means that people are greater than 99% similar. But when you look around the room at your classmates, you can see that that small difference amounts to quite a bit of variation within our species. The bulk of these differences aren't even within the coding sequences of genes, but lie outside in regulatory regions that change the expression of those genes. Imagine if there were mutations to the coding sequences, this could be very deleterious to the well-being of the organism. We say that the coding sequences of genes that ultimately lead to proteins has a **selective pressure** to remain the same. The areas outside of the coding sequences have a reduced and sometimes non-existent selection pressure. These areas are allowed to mutate in sequence and even expand or contract. Areas of changes or differences are called **polymorphic** (many forms). If you were to read a repetitive set of sequences and count the repetition, you'd make mistakes and lose count. Likewise, DNA polymerase will make errors or stutter in areas of repetitiveness and produce polymorphic regions.

Tandem Repeats

A type of polymorphism occurs due to these repeats expanding and contracting in non-coding regions. These regions are called variable number tandem repeats (**VNTRs**) or sometimes short tandem repeats (**STRs**). Any region or location on a chromosome is referred to as **locus** (loci for plural). Scientists use polymorphic loci that are known to contain VNTRs/STRs in order to differentiate people based on their DNA. This is often used in forensic science or in maternity/paternity cases. Any variation of a locus is referred to as an **allele**. In standard genetics, we often think of an allele as a variation of gene that would result in a difference in a physical manifestation of that gene. In the case of STRs, these alleles are simply a difference in number of repeats. That means the length of DNA within this locus is either longer or shorter and gives rise to many different alleles. VNTRs are referred to as **minisatellites** while STRs are called **microsatellites**.

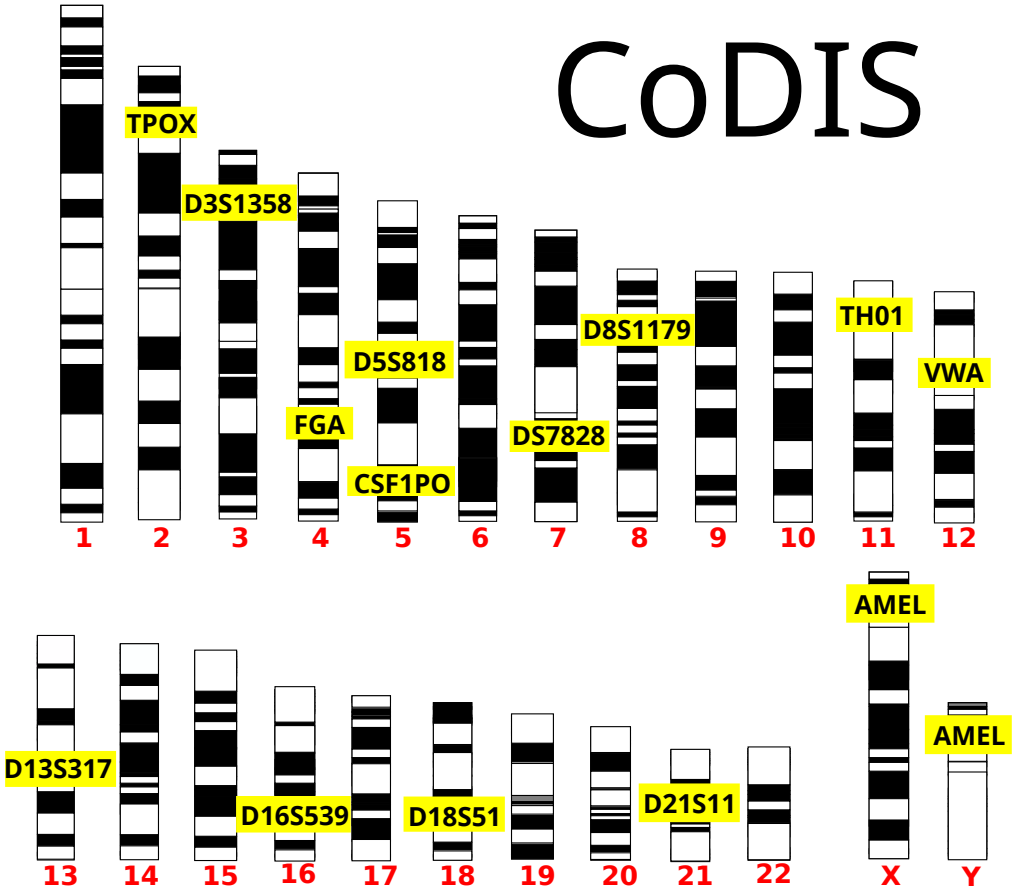
CoDIS

The FBI and local law enforcement agencies have developed a database called the Combined DNA Index System (**CoDIS**) that gathers data on a number of STRs. By establishing the number of repeats of a given locus, law enforcement officials can differentiate individuals based on the repeat length of these alleles. CoDIS uses a set of 13 loci that are tested together. As you



Credit: Shruthi.n.christ [CC BY-SA 4.0]

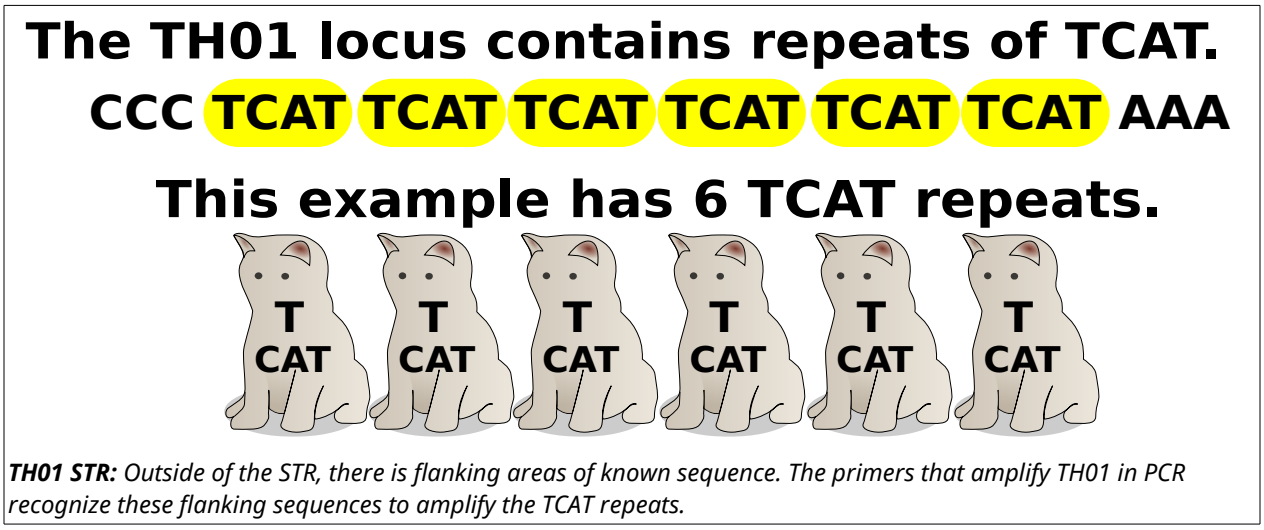
would imagine, people are bound to have the same alleles of certain loci, especially if they were related. The use of 13 different loci makes it statistically improbable that 2 different people could be confused for each other. Think about this in terms of physical traits. As you increase the number of physical traits used to describe someone, you are less likely to confuse that person with someone else based on those combinations of traits. Using the CoDIS loci increases the stringency since there are many alleles for each locus. The thirteenth locus in CoDIS (called AMEL) discriminates between male and female.



CoDIS STRs: *The FBI utilizes 13 different loci to discriminate between people. AMEL discriminates by gender and is located on the X & Y.*

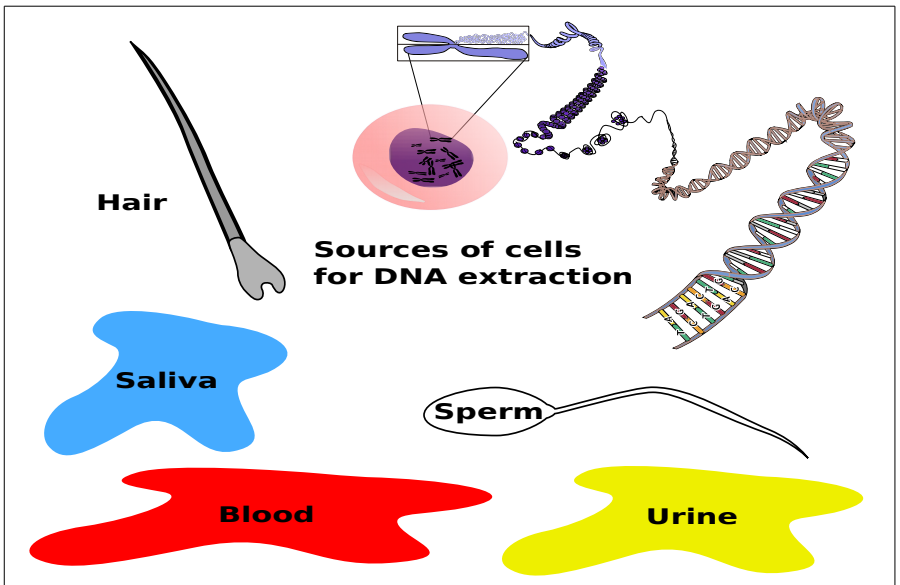
Crime Scene Investigation

This lab uses a CoDIS locus called TH01. TH01 is a locus on chromosome 11 that has a repeating sequence of TCAT. There are reported to be between 3-14 repeats in this locus. With the exception of X and Y in a male, all chromosomes have a homologous partner. Therefore, each individual will have 2 alleles for each CoDIS locus.



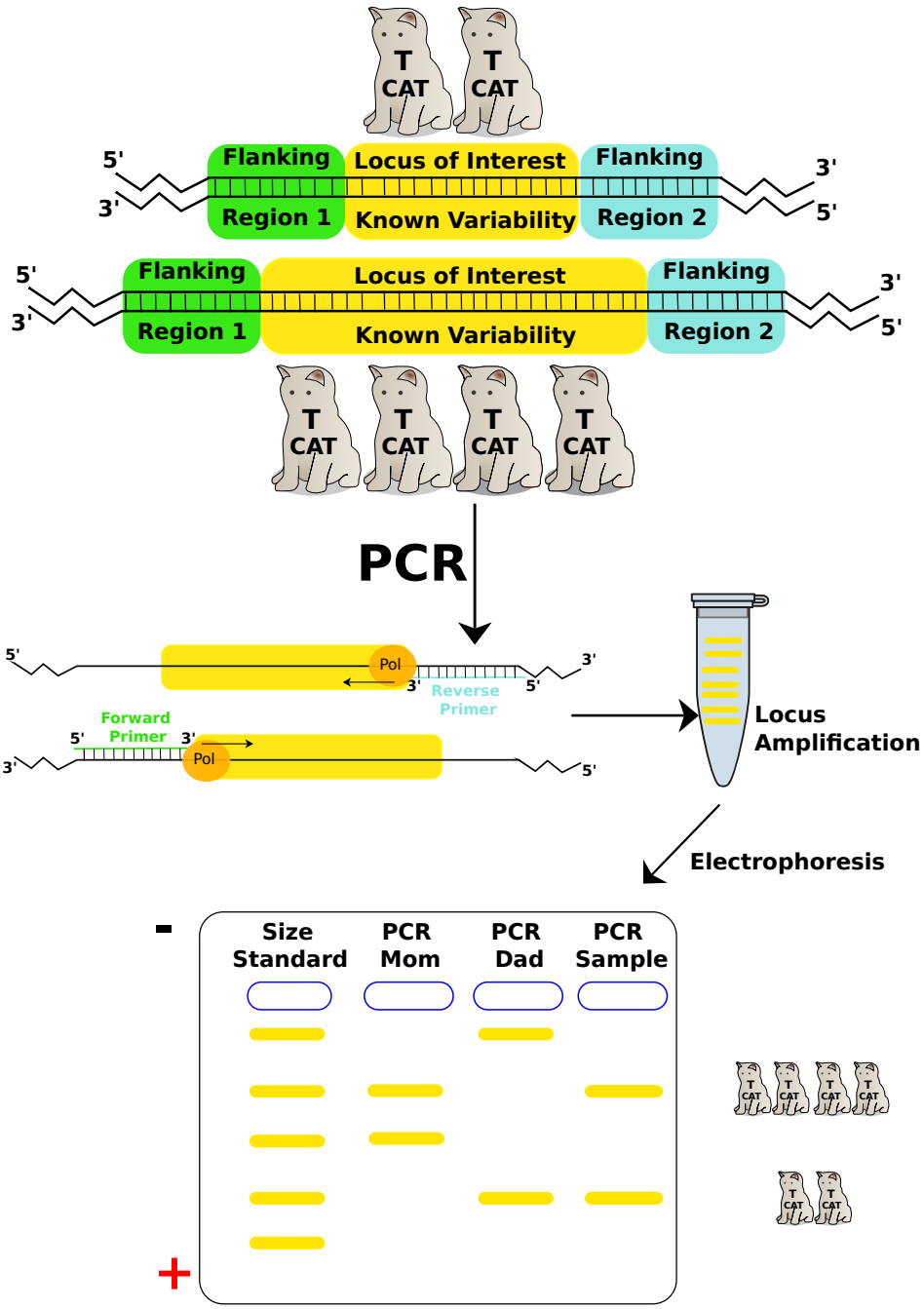
TH01 STR: *Outside of the STR, there is flanking areas of known sequence. The primers that amplify TH01 in PCR recognize these flanking sequences to amplify the TCAT repeats.*

At a crime scene, criminals don't often leave massive amounts of tissue behind. Scant evidence in the form of a few cells found within bodily fluids or stray hairs can be enough to use as DNA evidence. DNA is extracted from these few cells and amplified by PCR using the specific primers that flank the STRs used in CoDIS.



DNA evidence from a crime scene: *DNA can be extracted from cells found from various sources at a crime scene. PCR can amplify this small amount of DNA.*

Amplified DNA will be separated by gel electrophoresis and analyzed. Size reference standards and samples from the crime scene and the putative suspects would be analyzed together. In a paternity test, samples from the mother, the child and the suspected father would be analyzed in the same manner. A simple cheek swab will supply enough cells for this test.



TH01 locus used in a Paternity/Maternity test: *Individual PCR reactions are run for each sample (mom, dad, child). The TH01 primer pair specifically amplifies the locus. Each amplified sample is run on the same gel to resolve the different alleles of TH01 from each individual. From this test the sample could be the offspring from these 2 parents but use of more STRs would make it more definitive. Count the TCATs.*

External Resources

- Flash animation walking through what a STR is http://www.dnalc.org/view/15981-DNA-variations.html

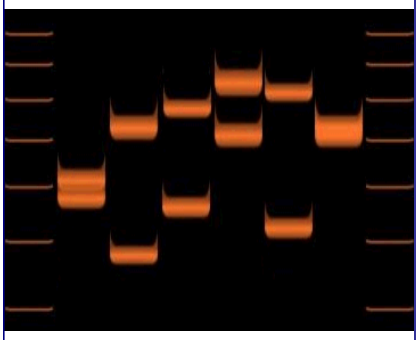
Minisatellite Activity

The minisatellite marker D1S80 is located at 1p35-p36. This VNTR is 16 bases long. With a variation of alleles between 3-24 repeats, the locus displays enough diversity to aid in distinguishing between people. Although this is not a CoDIS marker, use of multiple loci are required to definitively identify samples. The large repeat (16bp) permits the use of standard agarose gel electrophoresis to explore the diversity of this locus in our lab. PCR products range between 430 to 814bp long.

- D1S80-for: 5'-GAAACTGGCCTCCAAACACTGCCCGCCG-3'
- D1S80-rev: 5'-GTCTTGTTGGAGATGCACGTGCCCCTTGC-3'

1. PCR the DNA samples extracted from cheek cells using the PCR Beads
2. Pour 2% agarose into casting apparatus in refrigerator
    - 2 gels per class need to be made → 100ml of TBE with 2g agarose
    - add 5µl SYBR safe solution into the molten agarose before casting
    - place 2 sets of combs into the gel → at one end and in the middle
3. load gel with DNA ladder
    - Sample is from the PCR
4. Run gel at 120V for 30 minutes
5. Visualize on UV or Blue-light transilluminator

Example Results



By: PaleWhaleGail (CC BY-SA 3.0)

1. How many alleles are visible in each lane?
2. Are the genotypes distinguishable between individuals?
3. Are any of the alleles common between individual samples?

Jeremy Seto jseto@citytech.cuny.edu

Genetics leaves a bad taste in my mouth... or not

Some of our personal preferences arise from the way we were brought up. Culture plays a role in our likes and dislikes. Likewise, our experiences play a role in how we respond to certain stimuli. Another major factor that plays a role into our preferences comes wired in our genome. The DNA in our cells is the instruction manual for who we are. We are programmed to seek out things of a nutritive values in order to acquire raw materials like carbohydrates, proteins and lipids. In our search for nutritive compounds we have learned to avoid things that don't taste good. Bitter things have a tendency to be associated with toxic compounds in nature. When eating a food item for the first time, molecules hit our tongue and stimulate multiple sensations: sweet, sour, salty, savory and bitter. Attributed to these multiple taste types are a diverse family of receptors that bind to the molecules that result in our perception of these sensations. Something bitter might make us learn to avoid this food item in the future.

One type of bitter receptor senses the presence of a chemical called phenylthiocarbamide (**PTC**). This chemical chemically resembles toxic compounds found in plants but is non-toxic. The ability to taste PTC is comes from the gene called *TAS2R38*. This gene encodes a protein that on our tongues that communicates the bitterness of this chemical. There are two common alleles of this gene with at least five more uncommon variants. Within the two common forms, a **single nucleotide polymorphism** (**SNP**) is responsible for changing one amino acid in the receptor. It's this difference of one amino acid that results in the ability of the receptor to either respond or not respond to PTC. We inherit one copy of the gene from our father and one copy from our mother. Based on how our parents gametes were formed and what alleles we received during the fertilization event determines how we respond to this chemical. Because we each have 2 copies of this gene, we can utilize simple Mendelian genetics to understand which allele is dominant or recessive.

1. Place a piece of “Control” paper on the tongue and indicate if there is a taste
2. Place a piece of “PTC” paper on the tongue and indicate if there is a taste and the taste severity
3. Fill out the table for the class to identify how many non-tasters, tasters or super-tasters there are.
4. Indicate if you believe the trait is dominant or recessive (ability to taste or not taste)
5. Assign a descriptor allele for the dominant (a capital letter) or the recessive (a lowercase letter) and draw a Punnet square for the F_2 generation of 2 Heterozygous parents.
6. Compare the class tally of tasters and non-tasters in the class and discuss with your instructor if there is a clear dominance of this trait.

Table: PTC Tasting Tally

Phenotypes	Number	% Total
PTC Tasters (Dominant or Recessive)		
PTC Non-Tasters (Dominant or Recessive)		
Total		

Jeremy Seto jseto@citytech.cuny.edu

Questions:

1. How do you explain the presence of those who can't taste PTC, those who can taste it and those who really can't stand the taste of it?
2. This chemical is non-toxic and doesn't exist in nature. Do you think there is a **selective pressure** that confers an advantage to those who do taste it?

Exercise: Coding Bitterness

Prior to this exercise, review the **Central Dogma**.
The full coding sequence of ***TAS2R38*** is 1,002 bases (334 amino acids) long. A segment of the gene is shown below where the SNP (in red) occurs. Variant 1 is the version of the gene that encodes for the ability to taste PTC. Variant 2 is the version of the gene that is unable to bind to PTC. This SNP mutation is called a **missense** mutation because it changes the amino acid. Some mutations cause the insertion of a premature stop codon. This **nonsense mutation** results in a truncated protein and can be disastrous to the function. We already know that the simple substitution of one nucleotide translates to a change in one amino acid and determines the ability to taste PTC. Imagine if a large group of amino acids from the protein was missing.
With template strand (“Complement”) information:

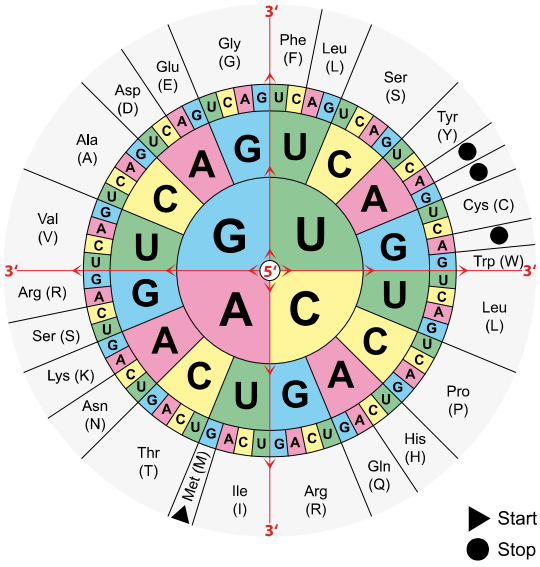
1. Write the sequence of the coding strand.
2. Write the sequence of the mRNA
3. Use the Genetic Code Chart to translate the amino acid sequence

Variant 1

Coding Strand: 5'-
Complement : 3'-TTC TCC GTC CGT GAC TCG-5'
mRNA : 5'-
Amino Acid :

Variant 2

Coding Strand: 5'-
Complement : 3'-TTC TCC GTC GGT GAC TCG-5'
mRNA : 5'-
Amino Acid :



Jeremy Seto jseto@citytech.cuny.edu

PCR Genotyping the TAS2R38 PTC receptor

- 5’-CCTTCGTTTTCTTGGTGAATTTTTGGGATGTAGTGAAGAGGCGG-3’ (Forward Primer)
- 5'-AGGTTGGCTTGGTTTGCAATCATC-3' (Reverse Primer)

1. PCR the DNA samples extracted from cheek cells using the PCR Beads
2. Pour 2% agarose into casting apparatus in refrigerator
    - 2 gels per class need to be made → 100ml of TBE with 2g agarose
    - add 5µl SYBR safe solution into the molten agarose before casting
    - place 2 sets of combs into the gel → at one end and in the middle
3. Digest PCR product with *Hae*III
    - remove 10µl of PCR product into a fresh tube
    - add 1µl of *Hae*III enzyme into tube
    - incubate for 10 minutes at 37°C
4. load gel with DNA ladder, Digested and Undigested
    - Undigested sample is from the original PCR
5. Run gel at 120V for 20 minutes
6. Visualize on UV transilluminator

SNP detection

The longer primer ends with the sequence "GG". Both alleles at this locus will amplify equally well with this primer set, however, one allele will have the sequence "GGGC" and another "GGCC". "GGCC" is the restriction site for the enzyme *Hae*III. The digestion of this amplified DNA will be digestible for one allele and yield a DNA fragment the size of the large primer (44 bp) as well as the remainder of the amplicon. Because of this difference in digestion profile of the amplicon, we can identify the 2 alleles at this locus.

Analysis questions

1. What is the size of the PCR product?
    - Perform an *in silico* PCR on the *Tas2R38* gene and identify the size of the amplicon.
2. The long primer is 44bp. If the amplicon of the allele digests, what are the sizes of fragments expected following *Hae*III?
3. Which allele is the one that can be identified through *Hae*III digestion?
    - use the results of the PTC paper test
4. Some lanes contain 3 bands instead of 1 or 2. Can you explain this?

Jeremy Seto jseto@citytech.cuny.edu