

Introduction—A series of 20th Century discoveries opened up a new world of molecular genetics that could not have been imagined by 19th Century geneticist like Gregor Mendel. The standard molecular genetic techniques that you will explore in this module are made possible by our modern understanding of the identity and composition of the genetic material (DNA!), how it is replicated and expressed as phenotypes. We will begin with a very brief overview of several important discoveries, and then take advantage of the knowledge gained over the past 100 years to conduct a virtual inquiry into the genetic basis of a phenotypic polymorphism.

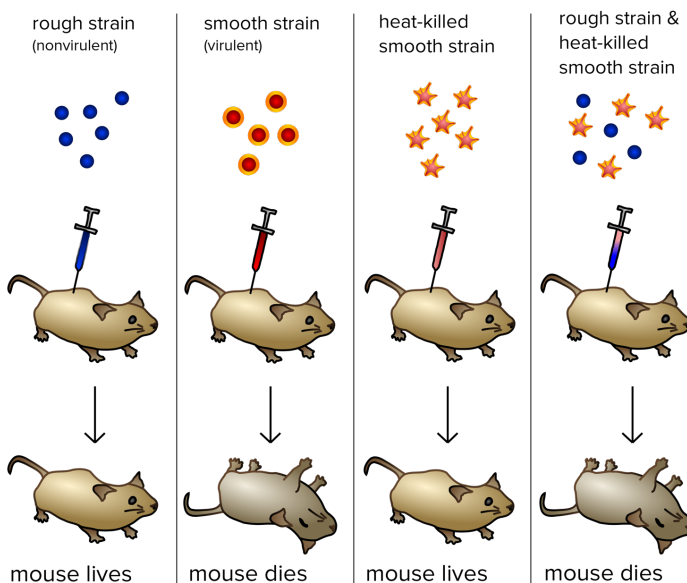


Figure 1. Summary of the Griffiths (1928) experiment demonstrating the “transforming principle”. Rough strain cells and heat-killed smooth strain cells do not kill mice, but smooth strain cells and rough strain cells that have been “transformed” by heat killed smooth strain cells do.

In **1928**, an English scientist named Frederick Griffith was investigating the different types of *Streptococcus pneumoniae* that could be cultured from the lungs of patients suffering from bacterial pneumonia. He and others noted that two different strains of the bacteria, named R (rough) and S (smooth) for the appearance of their surfaces, could spontaneously convert to the other form, presumably by mutations that changed the cells' surface. The cell shape affected the virulence of the bacteria; mice injected with S-type cells died, while mice injected with R-type cells were unaffected. Further, Griffith reported that mice unexpectedly died after he injected them with a mixture of

live R cells and heat-killed S cells, neither of which killed the mice when injected alone (Fig. 1). From these results, Griffith hypothesized that a substance was transferred from nonliving S cells to living R cells, transforming the benign recipient R cells into deadly S cells. Griffith's discovery of the “transforming principle” set the stage for new insights into the identity of the genetic material.

In **1944**, a Canadian-American scientist named Oswald Avery published results of his experiments regarding the transforming principle described by Griffith fifteen years prior. Avery and his colleagues developed conditions under which transforming activity could be measured in the absence of particular types of compounds (e.g., nucleic acids, proteins, lipids, carbohydrates). The elemental composition of the purified transforming compound was close to the theoretical values for DNA; most notably a high phosphorous content, which is characteristic of DNA, but not of proteins. The most definitive proof that the transforming principle was DNA was its sensitivity to specific enzymes, called DNases, that specifically degrade DNA. Avery and his colleagues were

able to show that transforming agent was not destroyed by enzymes that degrade proteins or RNA. Before the experiments of Avery and Griffith, the dogma of the time was that protein was the genetic material, as it was present in the nucleus in nearly equal amounts as DNA, and was structurally more diverse.

Contrary to popular belief, American scientist James Watson and British scientist Francis Crick did not discover DNA. In fact, nucleic acids had been discovered nearly a century earlier by Swiss scientist Friedrich Miescher. And in the decade subsequent to Avery's assertion that DNA was the hereditary material, the molecule had been much studied, notably by Austrian biochemist Erwin Chargaff who discovered the relationship between A/T and C/G nucleotides, and English physicist Rosalind Franklin who produced crucially important X-ray crystallography images of the molecule. In **1953** Watson and Crick proposed molecular structure for DNA (Fig. 2). The model's they proposed contains two principle features:

1. DNA is a double-stranded helix, with the two strands connected by hydrogen bonds. 'A' nucleotides are always paired with 'T's, and 'C's are always paired with 'G's, which is consistent with and accounts for Chargaff's rule.
2. The DNA double helix is anti-parallel, which means that the 5' end of one strand is paired with the 3' end of its complementary strand (and vice versa). As shown in Fig. 2, nucleotides are linked to each other by their phosphate groups, which bind the 3' end of one sugar to the 5' end of the next sugar.

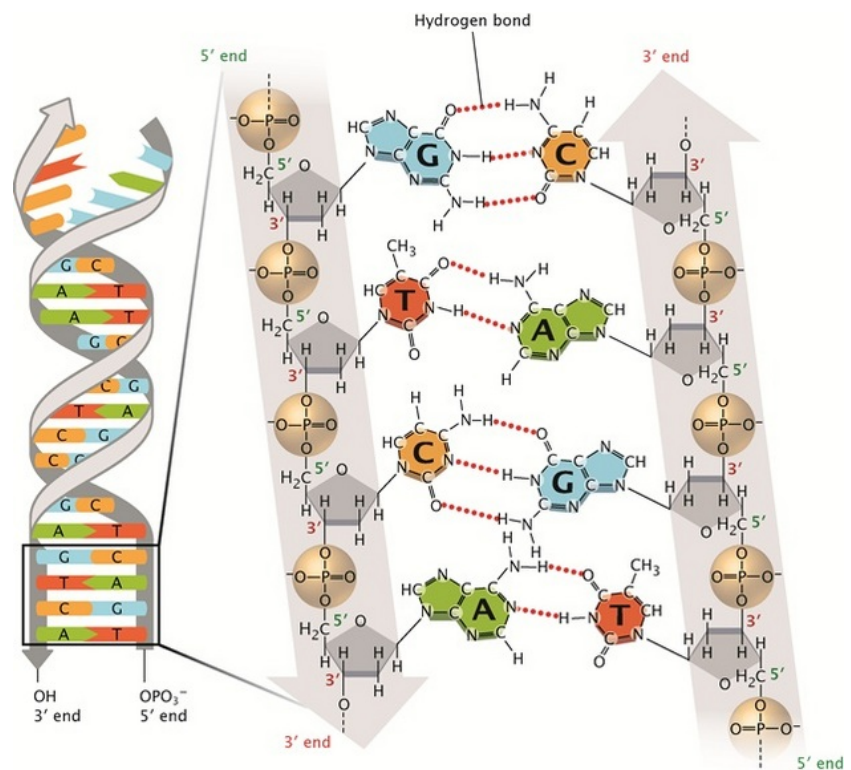


Figure 2. Depiction of the Watson-Crick model of DNA structure. Two-antiparallel strands of phosphate and deoxyribose sugars connected by hydrogen bonds between complimentary nucleotide bases.

An understanding of the molecular structure of DNA provided key insights into the mechanisms of replication and transcription. Insights that would eventually lead to our ability to rapidly synthesize incredible quantities of specific DNA sequences in the laboratory via the Polymerase Chain Reaction (PCR). The invention of PCR is attributed to American chemist Kary Mullis in **1985**, and the basic principle uses intrinsic cellular DNA replication mechanisms to target specific stretches of the DNA molecule for “amplification” via a chain reaction. The procedure requires placing genomic DNA into a tube with a batch of loose nucleotides (A, G, C, T) and a pair of synthesized short DNA segments that match segments on each side of the desired gene. These “primers” serve as starting points for DNA copying by a polymerase enzyme derived from heat tolerant

bacteria *Thermus aquaticus* (*Taq*). When the test tube is heated, the DNA separates into two strands. As the temperature is lowered the primers bind to their complementary portions of the DNA sample and the enzyme links the complementary loose nucleotides to the exposed DNA template. The heating and cooling is repeated, doubling the number of DNA copies every cycle. After thirty cycles a single copy of a piece of DNA can be amplified a million-fold.

Background and Objectives—Your objective for this experiment is to (virtually) isolate genomic DNA from a population of students who have been screened for the ability to taste bitter compounds and determine whether the population is in equilibrium (more on this later). In humans, certain chemicals in food stimulate particular taste cells on our tongue, which then send signals to the brain. The interpretation of these signals determines the flavor of the food. Taste cells on the tongue contain specific chemoreceptor proteins, of which there are five categories in humans: sour, salty, sweet, umami (savory), and bitter. There is only one type of receptor for sweet, sour, and umami, but there are possibly 30 types of receptors for bitter! In this lab we will explore variation in the ability to taste bitter compounds.

The ability to taste certain bitter substances is determined by genetic differences in the *TAS2R38* gene. This gene encodes a protein found in your taste buds that binds the specific molecules in foods that give rise to a bitter taste. The ability to taste or not is the phenotype, which is determined by the combination of alleles inherited at the *TAS2R38* locus. The allele (PAV) that allows tasting is dominant to the non-tasting allele (AVI). Thus, tasters may be homozygous dominant (PAV/PAV) or heterozygous (PAV/AVI) while non-tasters must be homozygous recessive (AVI/AVI). However, the *TAS2R38* gene does not explain all of the variation in bitter-compound tasting variability, so each individual's genotype at *TAS2R38* will not absolutely determine tasting ability.

Based on previous estimates of population-level taste phenotypes our naïve expectation is that 70% of students will be classified as tasters (HINT: this is your EXPECTED proportion of tasters for our population of students with unknown taste phenotypes!!). Bitter foods such as broccoli, cauliflower, and other members of the mustard family (Brassicaceae) contain thiocyanate compounds that stimulate the taste receptor proteins produced by *TAS2R38*. Food preference studies show that “tasters” tend to avoid these vegetables, and “nontasters” tend to eat a more varied diet that includes leafy vegetables. Before we begin our molecular genetics investigation of the *TAS2R38* gene we will learn more about using a series of bioinformatic tools.

Section 1: Bioinformatics

Major advances in the field of molecular genetics, coupled with advances in genomic technologies, have led to enormous growth in the genetic information generated by the scientific community. This flood of genomic information has led to a critical need for computer databases to store and organize data and for web-based tools to view and analyze data. Biological databases are large, persistent repositories of data, associated with software designed to query and retrieve the data stored within the system.

Bioinformatics is an interdisciplinary field of science that merges biology and computer science. One primary field of modern bioinformatics research involves DNA sequence analysis and genome annotation. The DNA sequences of thousands of organisms have been decoded and stored in databases, and full genomic sequences are now available

for numerous species. Genomic maps serve as a scaffold for orienting new sequence information. Until recently, a researcher interested in localizing a gene, or nucleotide sequence, was forced to manually map the genomic region of interest, a time-consuming and often painstaking process. Today, an abundance of existing sequence data, and high-quality, genome-wide maps enable rapid alignment of new gene sequence information with annotated data for the scientific community to use in their research. Bioinformatics has transformed the discipline of biology from a purely lab-based science to an information science as well. You will use the National Center for Biotechnology Information (NCBI) databases to learn more about the existing data for the *TAS2R38* gene.

LEARNING OBJECTIVES:

1. Learn to perform a BLAST search of the NCBI nucleotide database and interpret the search results
2. Understand how to locate genes and browse chromosomes in genomic databases

Procedures Part 1 (Use NCBI database to find *TAS2R38*)

1. Go to <http://www.ncbi.nlm.nih.gov/> to access the National Center for Biotechnology Information (NCBI) website. Click the 'Gene' link from the 'Popular Resources' menu and enter "*Homo sapiens TAS2R38*" in the search bar at the top of the page.

The screenshot shows the NCBI homepage. On the left is a sidebar with a 'Resource List (A-Z)' menu. The main content area is titled 'Welcome to NCBI' and features six large icons: 'Submit' (deposit data), 'Download' (transfer data), 'Learn' (find help), 'Develop' (use APIs), 'Analyze' (identify tools), and 'Research' (explore research). On the right is a 'Popular Resources' menu with links to PubMed, Bookshelf, PubMed Central, BLAST, Nucleotide, Genome, SNP, Gene (highlighted with a red arrow), Protein, and PubChem. Below this is an 'NCBI News & Blog' section with several news items dated October 2020.

2. Find and click the link for *TAS2R38* in the search results. Read through the information shown in the 'Summary' pane.
3. Scroll down to the 'Genomic context' pane. On what chromosome is *TAS2R38* located? What genes are located near *TAS2R38*? Click on the nearby genes; what are their functions? Record your responses on the worksheet.

4. Scroll down to 'Genomic regions, transcripts, and products' and click on 'FASTA' to view the nucleotide sequence of *TAS2R38*. Copy the FASTA sequence (beginning with the greater than sign) and paste into a new text file (this text file will be attached to your worksheet and turning in).



5. Copy and paste these primer sequences into the new text file below the *TAS2R38* sequence:
Forward 5'-AAAGGAGGCACAGACCACAG-3'
Reverse: 5'-GATGATTGCAAACCAAGCCAACCT-3'
6. Back in your web browser, click "Pick Primers" from the 'Analyze this sequence' pane on the right-hand side to open the 'Primer-BLAST' page.

```

TTGGTGCTTTTATGACAGCCTCAGTTCACAGTCACAAGTGTGCTATTCATGAATAACAATAAGGCTC
AACTGGCAGATTAAGATCTCAATTTATTTTATTCCTTTCTCTCTGCTATCTGTGGTCTGTGCCCTCTT
TCCTATTGTTCTGGTTCTCTGGGATGCTGACTGTCTCCCTGGGAAGGCACATGAGGACAATGAAGGT
CTATACCAAGAACTCTCGTGACCCAGCCTGGAGGCCACATTAAAGCCCTCAAGTCTCTTTGCTCCTTT
TTCTGCTTCTTTGTGATATCATCTGTGCTGCTTCTCATCTGTGCCCTACTGATTCTGTGGCGCGACA
AAATAGGGGTGATGTTTGTGGGATAATGGCAGCTTGTCCCTCTGGGCATGCAGCCATCCTGATCTC
AGGCAATGCCAAGTTGAGGAGAGCTGTGATGACCATTTCTGCTCTGGGCTCAGAGCAGCTGAAGGTAAGA
GCCGACCACAAGGCAGATTCGCCGACACTGTGCTGAGAATGGACATGAAATGAGCTCTTCATTAATACGC
CTGTGAGTCTTCATAAATATGCC

```

Analyze this sequence

Run BLAST

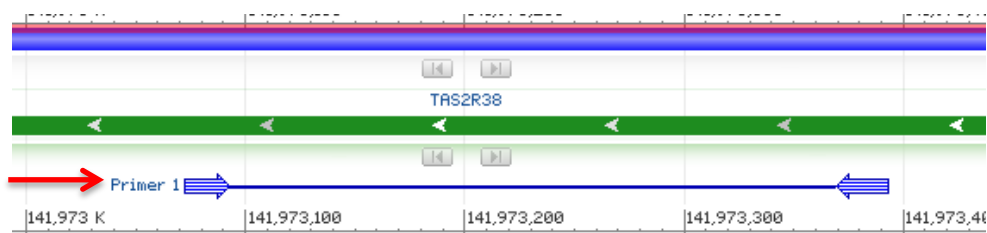
Pick Primers

Highlight Sequence Features

Find in this Sequence

Procedures Part 2 (Use Primer-BLAST to determine PCR amplicon size)

7. The sequence accession number is automatically placed in the 'PCR Template' search box. You must paste the forward and reverse primer sequences in the corresponding boxes in the "Primer Parameters" section and click "Get Primers".
8. Scroll down to the "Detailed primer reports" pane. What is the expected PCR amplicon size resulting from these primers? Record your response on the worksheet.
9. Scroll back up to the "Graphical view of primer pairs" containing the "Primer 1" position within the *TAS2R38* gene. Hover your cursor over "primer 1" to open a dialog box with information about the primers. Click the link next to "FASTA record:" in the dialog box. Copy and paste the amplicon FASTA sequence in your text document under the primer sequences.



Procedures Part 3 (Use BLAST to find PAV and AVI alleles)

10. Go back to <http://www.ncbi.nlm.nih.gov/> and click the 'BLAST' feature from the 'Popular Resources' menu and then click on the 'nucleotide blast' icon.
11. In the 'Enter Query' section, copy and paste the amplicon sequence from Part 2 (nucleotides only) into the search box. In the 'Choose Search Set' section, for

'Database', select "Nucleotide collection (nr/nt)" from the menu. In the 'Program Selection' section select "Optimize for Somewhat similar sequences (blastn)".

Standard Nucleotide BLAST

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)

[Reset page](#) [Bookmark](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

Query subrange [From](#) [To](#)

Or, upload file [Choose File](#) no file selected [+](#)

Job Title [Enter a descriptive title for your BLAST search](#)

☐ Align two or more sequences [+](#)

Choose Search Set

Database ☒ Standard databases (nr etc.): ☐ rRNA/ITS databases ☐ Genomic + transcript databases ☐ Betacoronavirus

Nucleotide collection (nr/nt) [+](#) [-](#) [+](#)

Organism [Optional](#) [Enter organism name or id--completions will be suggested](#) ☐ exclude [+](#)

[Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown](#)

Exclude [Optional](#) ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Limit to [Optional](#) ☐ Sequences from type material

Entrez Query [Optional](#) [You Tube](#) [Create custom database](#) [Enter an Entrez query to limit search](#)

Program Selection

Optimize for ☒ Highly similar sequences (megablast) ☐ More dissimilar sequences (discontiguous megablast) ☐ Somewhat similar sequences (blastn) [Choose a BLAST algorithm](#)

BLAST Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)

☐ Show results in a new window

12. Click on BLAST—Results may take a few moments to load

13. Enter "*Homo sapiens* (taxid:9606)" in the 'Organism' box under the 'Filter Results' header and click "Filter".

BLAST [®] » blastn suite » results for RID-SU5XPJE1016 [Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

[< Edit Search](#) [Save Search](#) [Search Summary](#) [?](#) How to read this report? [BLAST Help Videos](#) [Back to Traditional Results Page](#)

? Your search parameters were adjusted to search for a short input sequence.

Job Title	Nucleotide Sequence
RID	SU5XPJE1016 Search expires on 10-20 12:36 pm Download All
Program	BLASTN ? Citation
Database	nt See details
Query ID	lcl Query_44139
Description	None
Molecule type	nucleic acid
Query Length	22
Other reports	Distance tree of results MSA viewer ?

Filter Results

Organism [only top 20 will appear](#) ☐ exclude

[Type common name, binomial, taxid or group name](#)

[+ Add organism](#)

Percent Identity to E value to Query Coverage to

[Filter](#) [Reset](#)

14. Observe the list in the 'Descriptions' tab. Have you found records for the human *TAS2R38* gene? Record your observations on your worksheet.

15. In the list of significant alignments, click on the accession link for the "hit" containing the taster allele (AY258597.1)

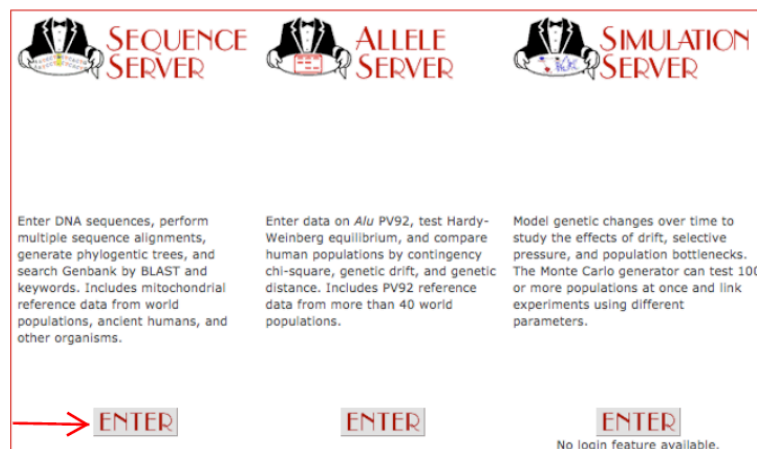
16. The bottom section lists the entire nucleotide sequence of the gene or DNA sequence that contains the PCR amplicon from Part 2. Highlight all the nucleotides between the beginning of the forward primer (base 297) and end of reverse primer (base

617). Paste this sequence into a text document. Then, delete all non-nucleotide characters and spaces and label the sequence as "Taster—PAV". Alternatively, you could repeat the Primer-BLAST steps from Part 2.

17. Repeat the steps above to copy the human non-taster allele (AY258598.1), and label as "Nontaster—AVI".
18. You now have compiled the 'generic' *TAS2R38* sequence (Part 1), the two PCR primers, the expected PCR amplicon (Part 2), and the amplicon sequences for the PAV and AVI alleles (Part 3). You will need to retain this information for Part 4 below, and Section 2 of the experiment (next week).

Procedures Part 4 (Use Multiple Sequence Alignment to Explore *TAS2R38*)

19. Return to the original BLAST results by clicking "reset" in the 'Filter Results' pane (or repeat the first few steps of Part 3 without filtering by *Homo sapiens*).
20. Find sequences of the *TAS2R38* gene from human tasters, human nontasters, chimpanzee (*Pan troglodytes*), bonobo (*Pan paniscus*), and gorilla (*Gorilla gorilla*).
21. Use only entries listed as "complete cds" (coding sequence). For each, open its Accession link, copy its complete nucleotide sequence by clicking on the FASTA link, and paste the sequences into your text document. Retain the original FASTA sequence labels (the text after the greater than sign).
22. Open the BioServers Internet site at www.bioservers.org. Enter 'Sequence Server' using the button in the left-hand column.



23. Create PTC gene sequences for comparison:
 - a. Click on 'Create Sequence' at the top of the page.



- b. Copy one of the *TAS2R38* sequences (from step above), and paste it into the Sequence window. Enter a name for the sequence, and click OK. Your new sequence will appear in the workspace at the bottom half of the page.

- c. Repeat for each of the human and primate sequences from step above. Also create a sequence for the forward primer used in your PCR amplification and for the amplicon.
24. Compare each of the following sets of sequences:
 - a. Human PTC taster vs. human PTC non-taster vs. predicted amplicons (from Part 3).
 - b. Human PTC taster vs. human PTC non-taster vs. chimpanzee vs. bonobo vs. gorilla.
25. Click on the Check Box in the left-hand column to compare two or more sequences.
26. Click on Compare in the grey bar. The results will appear in a new window. The sequences are displayed in rows of 25 nucleotides. Yellow highlighting denotes mismatches between sequences or regions where only one sequence begins or ends before another. Some of these are single nucleotide polymorphisms (SNPs) of which we will learn more about in Section 2 of the experiment.



27. Compare the Human taster sequence with the human non-taster sequence, and the predicted amplicons. Record your observations on the worksheet.
28. Compare the Human PTC taster sequence with the human PTC non-taster sequence and the other primates in the database (chimpanzee, bonobo, and gorilla). Record your observations on the worksheet.