CHAPTER 4:
How Is DNA Sequenced, and What Can We Learn?

INTRODUCTION

In this chapter, you will use bioinformatics software to analyze the nucleotide and amino acid sequences of alleles of TAS2R38 to determine how they differ. You will also explore how these genotypes relate to PTC taster phenotypes and discuss how small the differences are between the DNA sequences. Next, you will learn how to "read" the files created when DNA is sequenced. Lastly, you will learn about restriction enzymes in preparation for the activities in Chapter 5.

ACTIVITY: Finding TAS2R38 Differences

You are going to compare the DNA sequences of alleles of TAS2R38 to see how they differ and how those differences might relate to their phenotypes. Specifically, you will use DNA alignment software to compare the sequences of strong tasters and nontasters. How do the gene sequences of people with these PTC tasting phenotypes differ from one another?

NOTE: When you amplified your DNA in Chapter 3, you created many copies of a small fragment of the TAS2R38 gene. In this activity, however, you will be exploring the complete 1002-bp TAS2R38 gene.

MATERIALS

- Laptop or desktop computer with internet access
- 1 copy of **Finding TAS2R38 Differences (RM 4.1)**
- (Optional) Amino acid translation table (see **Table A.1** in the Appendix of this guide)

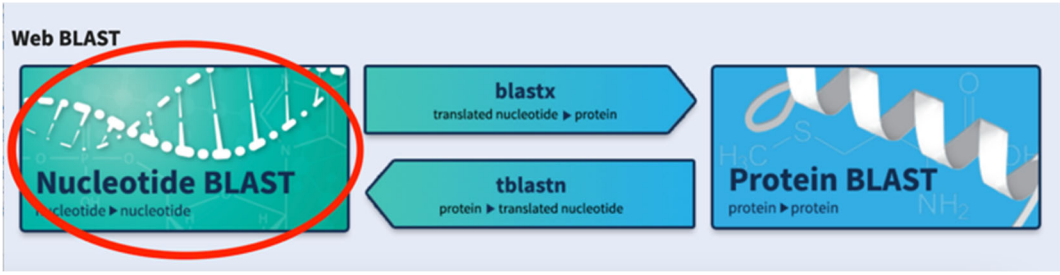
PROCEDURE

Before you begin, familiarize yourself with NCBI BLAST by watching the tutorial (4:03). Alternatively, read the Alignments Tab of the Web BLAST Quick Start Guide on the NCBI website.

Create an alignment between two nucleotide sequences: PTC taster and nontaster

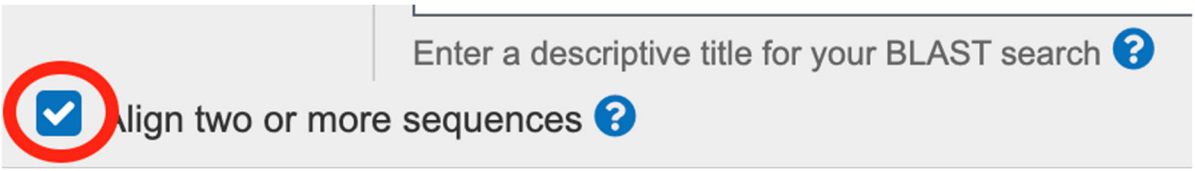
1. Navigate to the NCBI BLAST homepage.
2. Under the header "Web BLAST," click on the "Nucleotide BLAST" box (see **Figure 4.1**).

Figure 4.1: Nucleotide BLAST



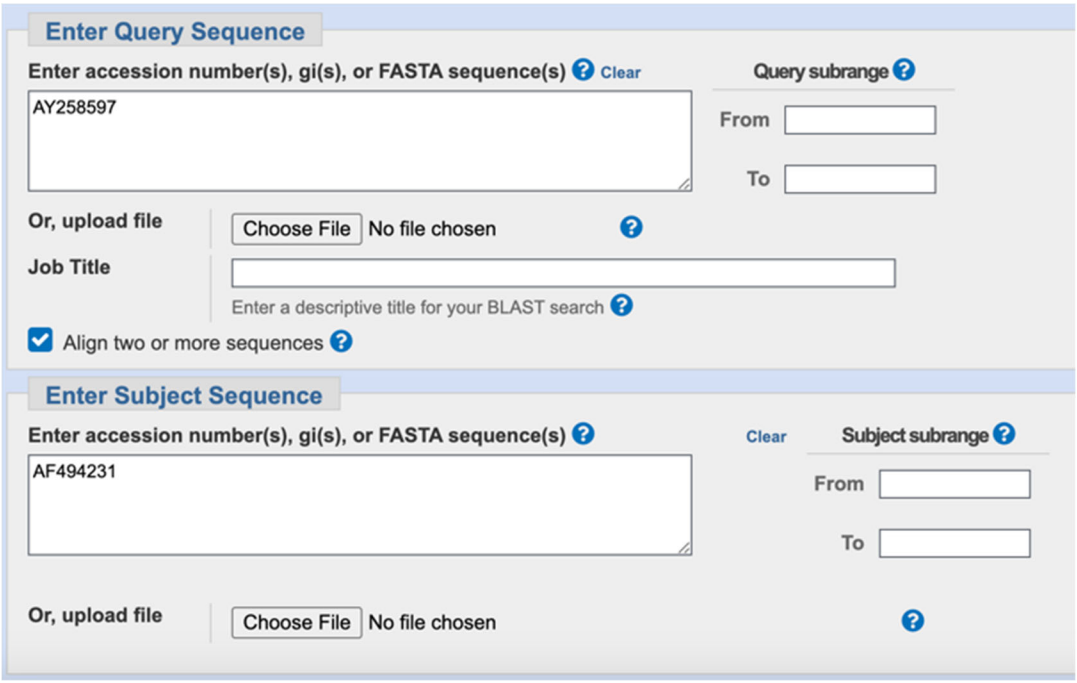
3. GenBank is a repository of nucleic acid and amino acid sequence files. You will be creating a DNA sequence alignment between two GenBank records: the nucleotide sequences of a PTC taster allele (AY258597) and a PTC nontaster allele (AF494231) of the TAS2R38 gene. Under "Enter Query Sequence," check the "Align 2 or more sequences" box (see **Figure 4.2**).

Figure 4.2: Align two or more sequences



4. Now, you should see two boxes, one called "Enter Query Sequence" and one called "Enter Subject Sequence" (see **Figure 4.3**). Enter the taster sequence accession number (AY258597) into the Enter Query Sequence box and the nontaster sequence accession number (AF494231) into the Enter Subject Sequence box.

Figure 4.3: Enter accession numbers for the query and subject DNA sequence

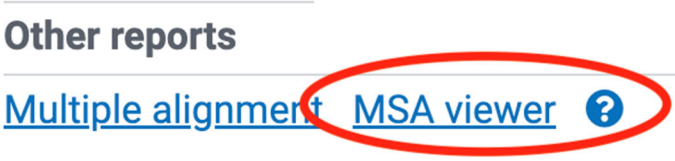


5. At the bottom of the page, select the "Show results in a new window" box and then click on the "BLAST" button (see **Figure 4.4**).

Figure 4.4: Show results and BLAST

6. An alignment analysis will open on a new page. Click "MSA viewer" next to "Other reports" (see **Figure 4.5**). A graphical representation of your nucleotide sequence alignment will open in a new window.

Figure 4.5: MSA viewer

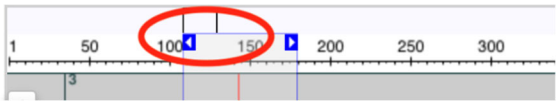


Analyze nucleotide differences between phenotypes

7. You should see three rows, labeled "consensus," "AY258597" (the taster sequence), and "AF494231" (the nontaster sequence). Above them you will notice a horizontal gray bar with several vertical red lines. These lines are ***SNVs*** (***single nucleotide variants)***, or differences in a gene's DNA sequence found within a population). Locate and count the red lines. How many SNVs do you see?

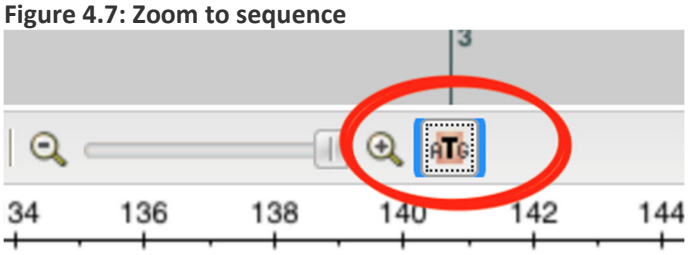
8. Move the uppermost rectangular slider until it is centered on the first single nucleotide variant (see **Figure 4.6**).

Figure 4.6: Move slider



9. Click the "Zoom to sequence" button (see **Figure 4.7**) to visualize both the consensus nucleotide sequence and the sequence of the taster and nontaster alleles at each SNV**.** You will notice that most of the taster and nontaster sequences are represented by dots, indicating that they are identical to the consensus sequence. Only the SNVs differ.

Figure 4.7: Zoom to sequence



10. What is the nucleotide position of the first SNV you find? Record the number as well as the difference in nucleotide sequences between the taster and nontaster alleles in the **Nucleotide Sequence Differences** table on **Finding TAS2R38 Differences (RM 4.1)**.

11. Move the slider to the next SNV and record its position and DNA sequence in **RM 4.1**. Continue with all the SNVs.

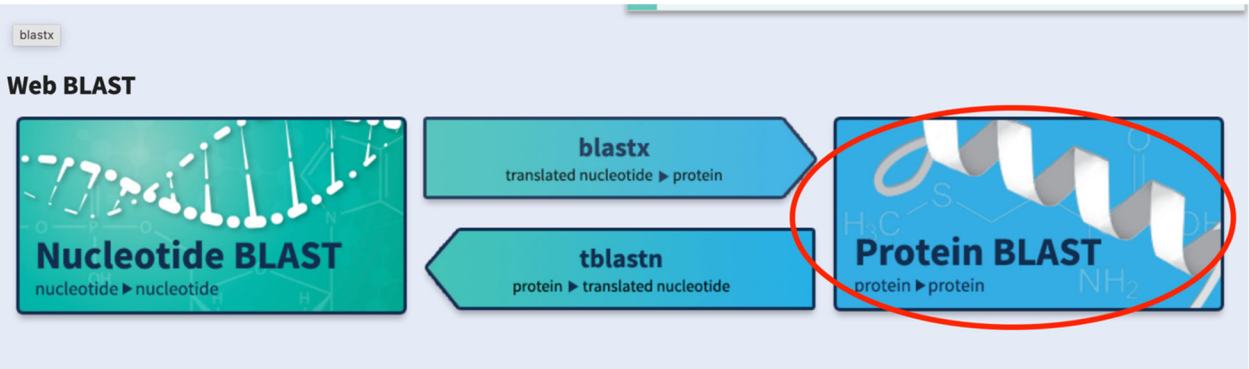
Analyze amino acid differences between phenotypes

You have identified SNVs in the taster and nontaster alleles. Now you will explore whether any of these nucleotide changes result in differences in the amino acid sequence of either allele. To do this, you will create another alignment in BLAST—this time aligning the amino acid translation of each of the same alleles.

12. Navigate to the NCBI BLAST homepage once more.

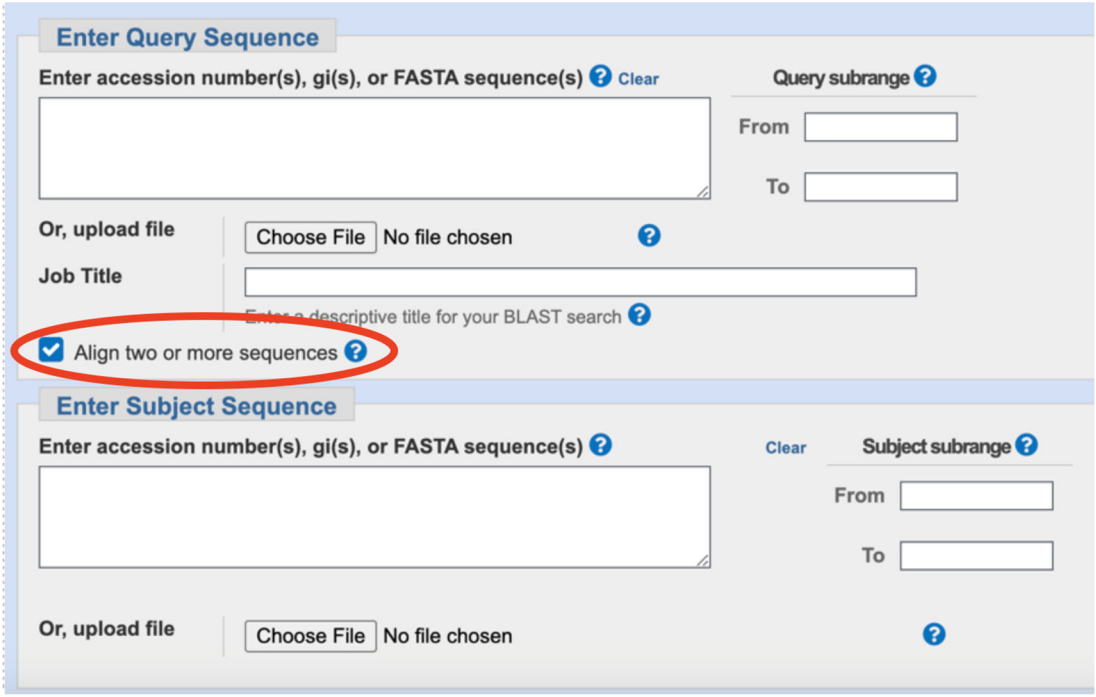
13. Under the header “Web BLAST,” click on the “Protein BLAST” box (**Figure 4.8**).

Figure 4.8: Protein BLAST



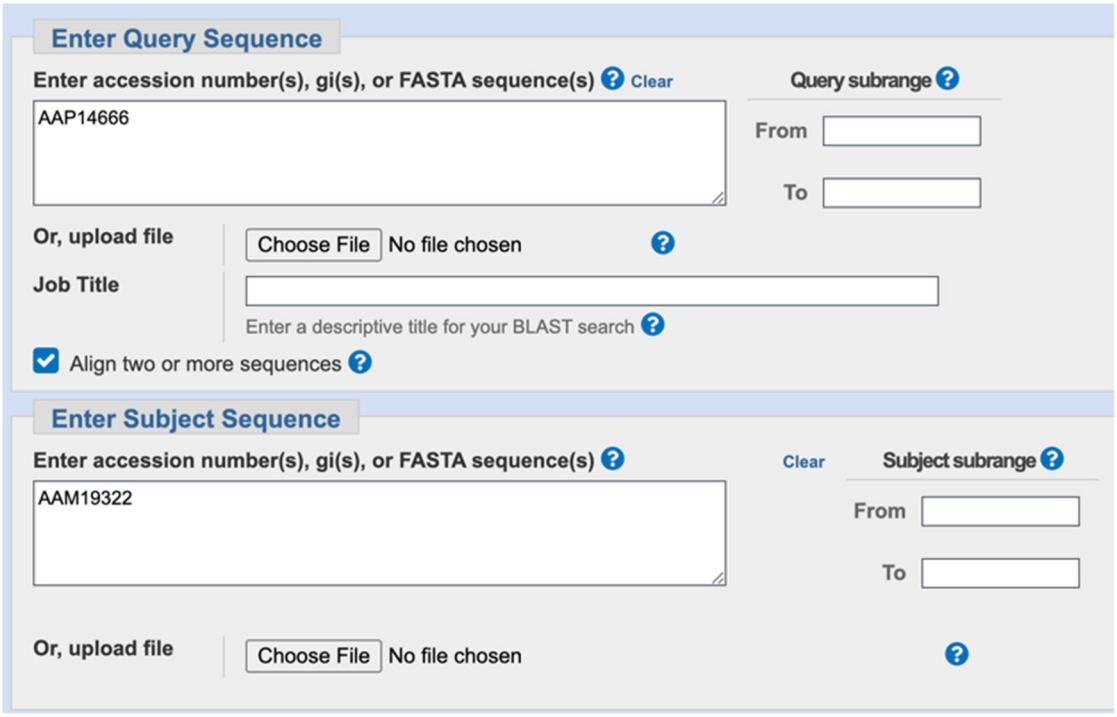
14. You will be creating a sequence alignment between two protein files from GenBank: the amino acid sequences of a PTC taster allele (AAP14666) and a PTC nontaster allele (AAM19322) of the TAS2R38 gene. Under “Enter Query Sequence,” check the “Align two or more sequences” box (see **Figure 4.9**).

Figure 4.9: Align two or more sequences



15. Now, you should see two boxes, one called “Enter Query Sequence,” and one called “Enter Subject Sequence” (see **Figure 4.10**). Enter the taster amino acid sequence accession number (AAP14666) into the Enter Query Sequence box and the nontaster amino acid sequence accession number (AAM19322) into the Enter Subject Sequence box.

Figure 4.10: Enter accession numbers for the query and subject amino acid sequences

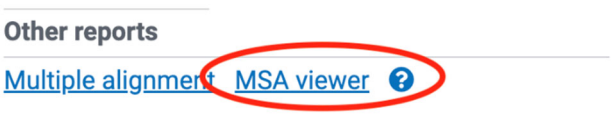


16. Select the "Show results in a new window" box and then click on the "BLAST" button at the bottom of the page (**Figure 4.11**).

Figure 4.11: Show results in a new window and BLAST

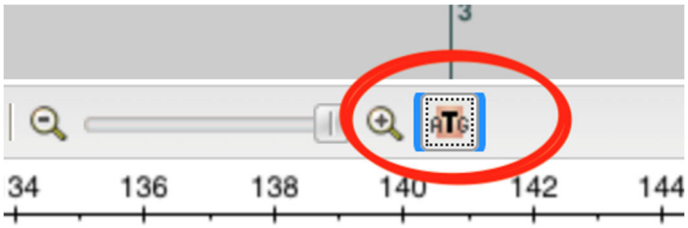
17. An alignment analysis will open on a new page. Click "MSA viewer" next to "Other reports" (see **Figure 4.12**). A graphical representation of your amino acid sequence alignment will open in a new window.

Figure 4.12: MSA viewer



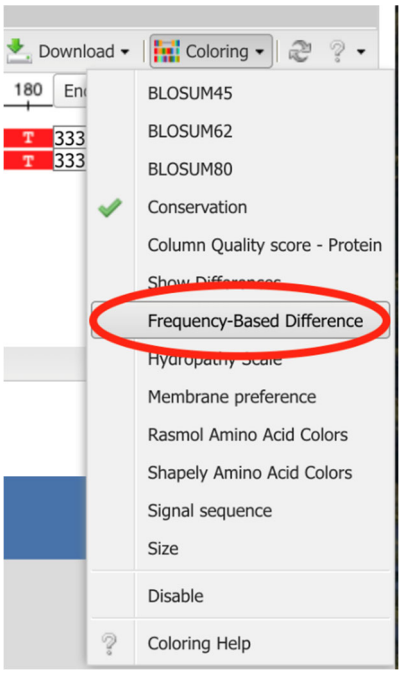
18. Click the "Zoom to sequence" button (see **Figure 4.13**) to visualize the amino acid sequences of the two TAS2R38 alleles in your alignment. You are going to scan the alignment in search of amino acid substitutions. As a reminder, AAP14666 is the taster sequence and AAM19322 is the nontaster sequence**.**

Figure 4.13: Zoom to sequence



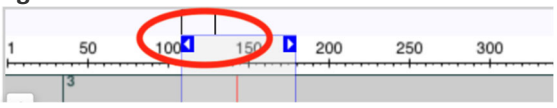
19. Under the "Coloring" pull-down menu, select "Frequency-Based Difference" (see **Figure 4.14**). Amino acids that differ between the two sequences will now be highlighted in red.

Figure 4.14: Frequency-based difference



20. Move the uppermost rectangular slider along the length of the sequence alignment to identify amino acid differences between the two sequences (see **Figure 4.15**). Record the amino acid positions and sequences in the **Amino Acid Sequence Differences** table on **Finding TAS2R38 Differences (RM 4.1)**.

Figure 4.15: Move slider



21. Before your class discussion, read the passage that follows to learn about the distinction between SNVs and SNPs.

READING: SNVs and SNPs

SINGLE NUCLEOTIDE VARIANTS VS. SINGLE NUCLEOTIDE POLYMORPHISMS

Between any two individuals, there are different nucleotides in their DNA on average once every 1,300 bases. These are single nucleotide variants (SNVs) like the ones you identified in alleles of the TAS2R38 gene. Some SNVs occur frequently in a population—at a rate of 1% or higher. These variants are known as ***single nucleotide polymorphisms***, or ***SNPs*** (pronounced "snips"). In both SNVs and SNPs, any nucleotide can be substituted—in other words, an A might become a C, G, or T. The substitutions are caused by errors during DNA replication or by ***mutagenesis*** (the process of permanent changes to DNA due to external agents). Single nucleotide substitutions can also occur in dividing somatic cells (for example, in cancer), but SNVs and SNPs must occur in ***germline DNA*** (the genome of reproductive cells) to be inherited.

While most SNVs occur in the non-coding portions of our DNA, some do produce phenotypic differences—like the ability to taste bitterness. Scientists use large human genomic datasets to map our genes and identify the location of these variants, just like you did with the TAS2R38 gene using BLAST.

How, then, do researchers determine that a SNV occurs with sufficient frequency to be called a SNP? They sequence DNA from databases of populations around the world, often examining thousands of individual sequences to determine the frequency of particular combinations of genetic variants.

ARE THERE ANY SNPS IN TAS2R38?

Three single nucleotide variants in TAS2R38 occur at sufficiently high frequencies to be classified as SNPs. These are located at nucleotide positions 145, 785, and 886. In Chapter 5, you will focus on the SNP at 145 in greater detail. If you wish to learn more about the biomedical research behind these SNPs, visit SNPedia and search for TAS2R38.

ACTIVITY: Reading DNA Chromatograms

MATERIALS

- Laptop or desktop computer with internet access
- 1 copy of **Exploring DNA Sequences (RM 4.2)**

NOTE FOR COLORBLIND STUDENTS: Being able to tell apart different colors is crucial to this activity. If you have red–green colorblindness, both macOS and Windows operating systems offer accessibility options that help users distinguish between colors that others experience as red or green. Make sure to ask your teacher for access to digital versions of the DNA chromatograms featured below.

BACKGROUND

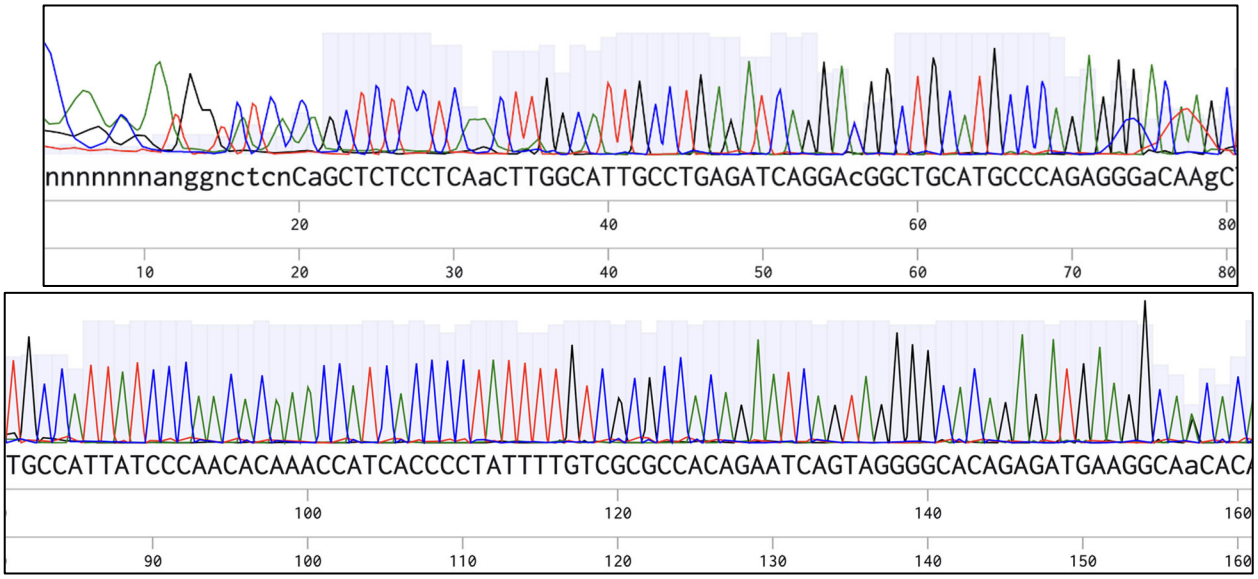
As you learned in Chapter 3, a DNA chromatogram is a visual representation of the nucleotide sequence of a DNA sample produced by an automated sequencing machine. In the DNA chromatogram, each color represents a specific nucleotide (see **Table 4.1**). The height of each color peak in the DNA chromatogram represents the relative fluorescent intensity of a nucleotide whose identity is determined by base-calling software. Should the software be unable to assign a nucleotide base with confidence, an "N" (for nucleotide) will appear as a placeholder.

Table 4.1: Key to the DNA chromatogram colors

Key to the DNA chromatogram Colors	
Black = Guanine (G)	**Red = Thymine (T)**
Blue = Cytosine (C)	**Green = Adenine (A)**

A DNA chromatogram looks like the image below—four different colored strands ("traces") representing the four nucleotides (see **Figure 4.16)**. The gray blocks behind the traces represent the ***confidence interval***—a measure of the probability that the software correctly identified the nucleotide in the DNA sequence. Basically, the taller the gray area, the more confident the program is in assigning the nucleotide.

Figure 4.16: Example DNA chromatogram



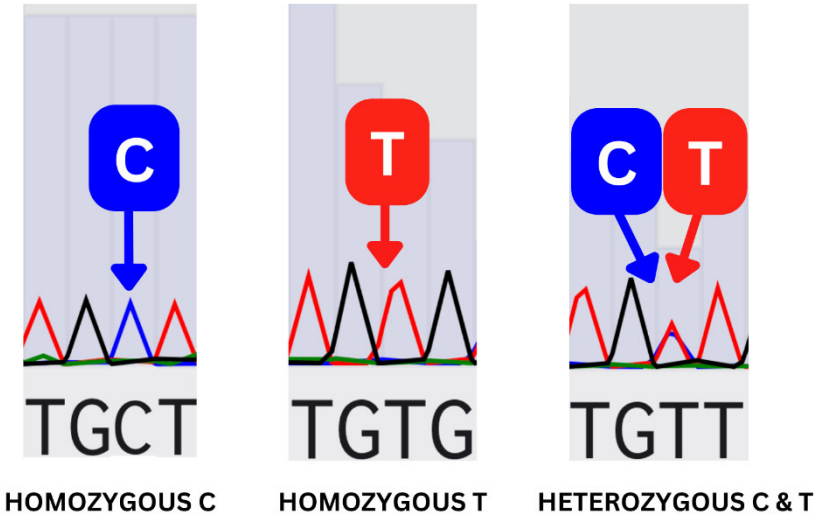
In a "good" DNA chromatogram, the peaks are spaced evenly (to the right of nucleotide position 20 in **Figure 4.16** above, for the most part). Peak heights may vary, which is normal. Often, the first 20–40 base pairs of a sequence are unclear (as shown in the lower confidence intervals above), and you can see many "N" (for "nucleotide") placeholders at positions where the computer cannot assign a nucleotide base with confidence.

Occasionally, broad peaks are visible at locations where the sequence repeats the same nucleotide multiple times. However, the trace will still show defined peaks at each nucleotide position; the software reads these as distinct signals and assigns the correct nucleotide bases (see **Figure 4.17)**.

Figure 4.17: Broad peaks in a DNA chromatogram

If a person is heterozygous for a particular base pair, you should see two peaks at that location. The images below represent sequences that are homozygous C, homozygous T, and heterozygous C and T (see **Figure 4.18)**. Note that in the heterozygous individual shown, the height of the peaks for C and T, respectively, are approximately half the height of the peaks for that position in the homozygous examples.

Figure 4.18: Examples of heterozygous vs. homozygous nucleotide peaks in DNA chromatograms

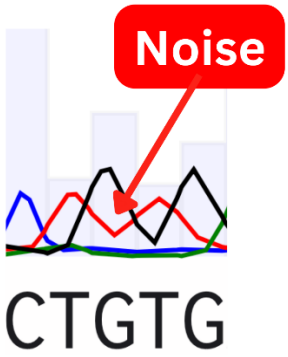


Sometimes, portions of a DNA chromatogram are illegible. There are numerous potential reasons for "noisy" traces, including poor DNA quality, low concentration of template DNA, suboptimal PCR conditions, and degradation of sequencing reagents.

How can you distinguish between a clear-cut, heterozygous position and a "noisy" one? In any DNA chromatogram, there is always some background "noise" from the other traces, but the base-calling software can usually tell the difference between "noise" and a heterozygous base pair at a position.

For this DNA chromatogram analysis, peaks that are less than half the height of the main peak at a nucleotide position can be considered “noise” (see **Figure 4.18**). Sometimes there are single broad peaks that span several nucleotide positions. These can also be considered “noise.”

Figure 4.18: “Noise” in a DNA chromatogram



PROCEDURE

Now try it yourself. Explore the DNA chromatogram on **Exploring DNA Sequences (RM 4.2)** and then answer the questions.

ACTIVITY: Explore the DNA Chromatogram of a Weak PTC Taster

MATERIALS

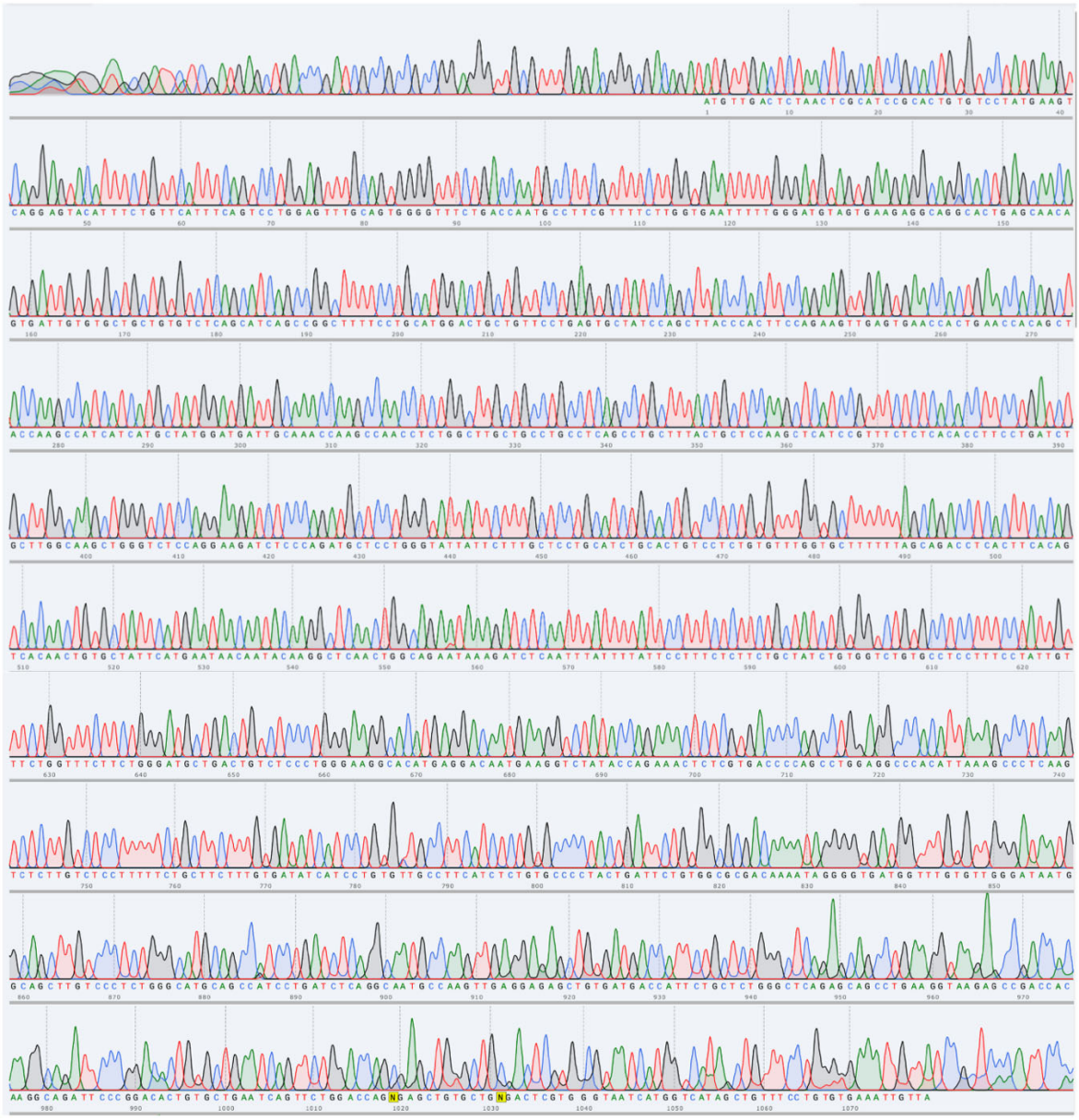
- 1 copy of **Finding TAS2R38 Differences (RM 4.1)**
- 1 copy of **Interpreting a Heterozygous DNA Sequence (RM 4.3)**

PROCEDURE

1. Now you have learned how to read a DNA chromatogram and you have identified SNPs which differentiate the taster from the nontaster allele. Review your answers on **Finding TAS2R38 Differences (RM 4.1)**. What do you predict you will find at those same nucleotide positions in the DNA sequence of a weak taster? Keep in mind they would have inherited one taster allele from one parent, and one nontaster allele from the other parent. Write your predictions in the table on **Interpreting a Heterozygous DNA Sequence (RM 4.3)**.

2. Next, examine the DNA chromatogram below (see **Figure 4.19),** which shows the DNA sequence of a heterozygous individual who must have one copy of each allele. Please note that this DNA chromatogram has been trimmed to reduce “noise” and remove "N" positions, and the confidence interval is not displayed.

Figure 4.19: DNA chromatogram of TAS2R38 from a heterozygous, weak PTC taster



3. Keeping the diagnostic SNP positions from **RM 4.1** in mind, examine those same nucleotide locations in the weak taster DNA chromatogram in search of heterozygous peaks. As a reminder, a heterozygous position will have two overlapping peaks with different colors. Note any heterozygous positions you find in the table on **RM 4.3**.

READING: Exploring the Differences in Our DNA

As you read the passage below, which will prepare you for the lab activities of Chapter 5, write down definitions in your science notebook.

WHY LOOK FOR SNPS?

SNPs are molecular signposts that help researchers study molecular evolution. These single nucleotide substitutions represent a historical record of minor, heritable changes accumulating in the human genome and being passed down from one generation to the next. Scientists can use mathematical formulas to determine how much time has passed since a particular nucleotide variant has emerged in a DNA sequence. With collections of DNA sequences from geographical populations worldwide, they can also map the movement of specific haplotypes from one place to another. This information provides insight into selective forces that have acted on human populations over time.

Some SNPs, such as synonymous nucleotide substitutions within ***exons*** (portions of coding sequence that are spliced together to make mRNA), do not affect the structure or expression of proteins. The same is true for most nucleotide substitutions in noncoding DNA, unless mutations occur in important upstream regulatory regions, such as promoter sites. (A ***promoter*** is a region of DNA upstream from a gene where proteins bind to initiate transcription of the gene). Nonsynonymous SNPs, such as those that cause amino acid substitutions, are noteworthy from both an evolutionary biology and biomedical standpoint. Mutations that cause physiological changes are valuable targets for drug development and genetic engineering.

SNPs are often detected when researchers compare the DNA sequences of many individuals at once. This process typically requires large-scale lab equipment and multiple personnel.

DETERMINING DIFFERENCES IN OUR DNA WITHOUT SEQUENCING

Large DNA sequencing projects can be expensive and time-consuming. However, sometimes simply knowing which nucleotide is present at a specific position in a gene's coding sequence can reveal which allele you have. In the activity **Finding TAS2R38 Differences**, you found places in which the sequence of nucleotides in the TAS2R38 gene was different between people who are PTC tasters and nontasters.

Knowing the locations of diagnostic SNPs like these allows you to use a less expensive and more widely available technology to determine your particular genotype. To do this, you can use restriction enzyme digestion followed by gel electrophoresis.

WHAT ARE RESTRICTION ENZYMES?

Restriction enzymes are specialized bacterial proteins that cut DNA into fragments at or near specific sequences of bases. When these proteins are used in the lab to cut double-stranded DNA at particular sites for diagnostic or gene cloning purposes, the technique is known as a ***restriction enzyme digest***.

In the early 1950s, scientists observed that certain strains of *E. coli*, a common bacterium found in the human gut, were resistant to infection by ***bacteriophages***—viruses that infect bacteria by injecting their DNA into the cell and commandeering the host cell's molecular processes to make more bacteriophages. Investigation of this primitive bacterial "immune system" led to the discovery of restriction enzymes, which restrict bacteriophage growth by recognizing and destroying phage DNA without damaging the host DNA. Subsequent studies demonstrated that restriction enzymes from different species of bacteria cut DNA at specific sequences, which are called ***restriction sites***. There are currently at least 3,000 different restriction enzymes commercially available!

One class of restriction enzymes recognizes specific 4–8-bp-long DNA sequences and typically cuts the strands at a particular position within or outside of the recognition site (See **Figure 4.20**). During a restriction enzyme digest *in vitro*, cuts at each of these recognition sites generate DNA fragments of different sizes. Then these DNA fragments can be separated by size using gel electrophoresis. You will use restriction enzyme digestion and gel electrophoresis to learn more about your own taster genotype in the next chapter.

Figure 4.20: An illustration of the restriction enzyme *HaeIII* and its recognition site