INTRODUCTION

A look around the world shows that people's skin comes in many different shades from the lightest pink to darkest brown. All of this variation can't be explained by differences in a single gene. Skin comes in such an array of colors because it is a **polygenic trait**. Polygenic traits are determined by a combined effect of more than one gene located at different **loci** (locations) throughout the genome and each with two or more **alleles**.

Studies into the genetics of human skin color have concluded that at least 34 genes (Sturm and Duffy 2012) have a detectable influence on skin color, but there are likely many more.

PART 2: Searching for skin color genes

How do we know which genes are involved in skin color? Scientists use genome-wide association studies to determine how differences in DNA among individuals or even populations relate to different phenotypes. These studies involve comparing DNA sequences across the genomes of a large number of individuals in search of variations that are consistently associated with particular phenotypes. These variations are called **single-nucleotide polymorphisms**, or **SNPs** (pronounced "snips"). A SNP is a variation in a single nucleotide at a particular position, or **locus**, in the genome. Not all single-nucleotide changes are SNPs. To be classified as a SNP, the change must be found in more than 1% of the population. SNPs are also called markers.

Once a SNP locus has been identified, scientists can use the differences in populations to determine the gene or genes that generate the phenotypes they see. Using this approach, many genes have been identified as having a role in determining skin pigmentation. The genes identified as having the strongest effect on skin color are *TYR*, *TYRP1*, *OCA2*, *SLC45A2*, *SLC24A5*, and *MC1R*. Among these, the *melanocortin 1 receptor* gene (*MC1R*) is the major contributor to normal pigment variation. To date, scientists have identified more than 50 SNPs within the *MC1R* gene. Each SNP represents a different allele, so there are more than 50 known alleles for the *MC1R* gene.

SNPs can also be used to gain insight into an individual's ancestry. Scientists have compared SNPs among populations of **indigenous** people, or people native to a particular place, and quantified how often a given allele (SNP) of a specific gene is present within that population. This measure of the relative frequency of a given allele, which is called **allele frequency**, is often expressed as a percentage. For example, the skin color gene *SLC24A5* has a SNP locus known as rs1426654. There are two alleles at this locus; an individual can have an adenine (A) or a guanine (G). Studies of indigenous populations have revealed that the allele frequency of the A allele is 100% among Europeans (making the frequency of the G allele 0%) and 2.5% among Africans (making the frequency of the G allele 97.5%). If you were to consider just these two groups of people, you would consider a person with two A alleles to most likely be of European ancestry, whereas a person with two G alleles most likely would be of African ancestry. By looking at the frequencies of SNPs throughout their genome, a researcher can make an informed hypothesis about an individual's ancestry based on their DNA, as you will see in Part 3.

5. To search for genes involved in determining skin color, scientists look for SNPs associated with different skin color phenotypes. SNPs are variations at a single nucleotide within the genome. How can a change in a single nucleotide be responsible for differences in skin color or the function of a gene in general?

Most phenotypic traits vary among the individuals in a population. For many traits, an individual's phenotype is determined by environmental influences as well as their genetic makeup. A measure of the degree to which differences in a trait are the result of genetics is called **heritability**. Heritability values range from 0 to 1.

- A heritability value of 1.0 means that 100% of the differences in a trait found among individuals in a population are caused by genetic differences.
- A heritability value of 0.0 means that 100% of the differences in a trait found among individuals in a population are caused by environmental differences.

For any given trait, heritability usually falls somewhere between these two extremes. As long as a trait has heritability greater than 0, it can be acted on by natural selection. Note that a measure of heritability is only useful for a given population at a given time. It is not an absolute and unchanging number.

Biologists estimate heritability of a trait in a population by comparing a trait among close relatives, for example, among parents and children or between twins. The more similar the expression of a trait is between close relatives, the more likely it is to have a high heritability.

6. Why are identical twins a good source of data for studies into the heritability of a trait?

In one particular study (Clark *et al.* 1981), scientists measured skin color using a **reflectometer** (a device that measures reflectance of a surface) in 134 pairs of twins. Some of the twins were identical, others were fraternal. Study participants lived in the same geographic area and had similar sun exposure. The study concluded that skin color has a heritability of 0.83.

7. Support this claim using evidence from the information provided: Differences in human skin color are caused primarily by differences in genetics.

8. We know that multiple genes (some with many different alleles) contribute to skin color, but genes alone do not account for the diversity of pigmentation we see among humans; environmental factors play a role. Propose an explanation for how one of these factors could alter the expression of skin color genes.

PART 3: Using skin color-related allele frequencies to infer ancestry

Have you ever wondered how companies are able to trace your ancestry by analyzing a saliva sample? By understanding genetics and mathematical modeling, and having the ability to sequence DNA, scientists can use computers to compare an individual's DNA against a database of known allele frequencies tied to different geographic regions. First, DNA is extracted from the saliva sample and analyzed using a SNP chip, a small piece of silicon glass with short pieces of DNA attached to it. These pieces of DNA are complementary to certain SNPs.

When an individual's DNA is exposed to the SNP chip, only the DNA molecule whose nucleotide is complementary "sticks," or anneals, to the chip. This identifies the nucleotide at a designated SNP locus. A single SNP chip can look at several hundred thousand SNPs in a sample. An individual's SNPs are then compared to the data bank to make a prediction about their ancestry. This process is accomplished using a mathematical model.

In this activity, you will practice analyzing a person's set of SNPs to make an estimate of their region of ancestry. Table 2 provides the allele frequencies for various SNPs of five different genes involved in human skin pigmentation. Frequency data is provided for four different indigenous populations: European, Chinese, Japanese, and African.

The analysis you are conducting involves two simplifications. First, you will look at profiles with only 13 SNPs and all the SNPs are related to skin pigmentation. In reality, estimating someone's ancestry requires looking at many SNPs throughout the entire genome and the SNPs used in the analyses are not only from loci associated with skin pigmentation. Second, you will only consider one allele at each locus for the two individuals explored in the activity. In some cases, people could be homozygous for a particular SNP in which case using only one SNP at a locus is accurate, but in other cases people may have two different alleles at a locus.

Table 2: Frequencies of various skin pigmentation alleles in different native populations. (Data from Strum R.A. 2009 and The 1000 Genomes Project Consortium.)

			Allele Frequency in Indigenous Population			
Gene	**SNP Locus**	**Allele**	European	Chinese	Japanese	African
TYR	rs1042602	C	0.583	1.0	1.0	1.0
		A	0.417	0.0	0.0	0.0
	rs1800422	G	0.604	1.0	1.0	0.935
		A	0.396	0.0	0.0	0.065
	rs1126809	G	0.783	1.0	1.0	1.0
		A	0.217	0.0	0.0	0.0
TYRP1	rs1408799	C	0.30	0.989	0.978	0.775
		T	0.70	0.011	0.022	0.225
	rs2733832	C	0.367	0.989	0.977	0.933
		T	0.633	0.011	0.023	0.067
OCA2	rs1800401	C	0.935	1.0	1.0	0.979
		T	0.065	0.0	0.0	0.021
	rs1800407	G	0.933	1.0	1.0	1.0
		A	0.067	0.0	0.0	0.0
	rs1800414	A	1.0	0.367	0.477	1.0
		G	0.0	0.633	0.523	0.0
	rs12913832	T	0.208	1.0	1.0	1.0
		C	0.792	0.0	0.0	0.0
SLC45A2	rs26722	G	1.0	0.611	0.591	0.95
		A	0.0	0.389	0.409	0.05
	rs16891982	G	0.017	0.989	1.0	1.0
		C	0.983	0.011	0.0	0.0
SLC24A5	rs1426654	G	0.0	0.989	0.989	0.975
		A	1.0	0.011	0.011	0.025
KITLG	rs642742	A	0.136	0.267	0.114	0.922
		G	0.864	0.733	0.886	0.0778

Profile 1 (below) shows the alleles present in those five genes in an individual. Use the information in Table 2 to determine the most likely predominant ancestry of that individual; in other words, is this individual of mostly European, Chinese, Japanese, or African ancestry?

To answer this question, look at each allele present in this individual and record its associated frequency in each of the four populations from Table 2. For example, this individual has allele C for the *TYR* SNP rs1042602. According to Table 1, the frequency of this allele in Europeans is 0.583, which means that 58.3% of European individuals have this allele. On the other hand, this allele's frequency is 100% among people of Chinese, Japanese, and African ancestry.

After you fill in the frequencies, circle the highest one(s) for each allele to help you conclude the most likely predominant ancestry for the individual profiled. Row 1 has been completed for you.

Profile 2

			Allele Frequency in Indigenous Population			
Gene	**SNP Locus**	**Allele**	European	Chinese	Japanese	African
TYR	rs1042602	C				
	rs1800422	G				
	rs1126809	G				
TYRP1	rs1408799	T				
	rs2733832	C				
OCA2	rs1800401	C				
	rs1800407	G				
	rs1800414	G				
	rs12913832	T				
SLC45A2	rs26722	A				
	rs16891982	G				
SLC24A5	rs1426654	G				
KITLG	rs642742	G				

10. What can you conclude about the predominant ancestry of this individual based on the data? Explain your answer.

11. When trying to distinguish ancestry, some SNP alleles are more helpful than others. Which is more useful in determining likely ancestry: a C allele at rs1042602 or a C allele at rs12913832? Explain your answer.