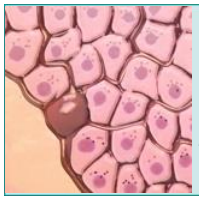


Understanding Variation in Human Skin Color

hhmi | BioInteractive

Activity
Educator Materials

OVERVIEW

The film *The Biology of Skin Color* walks viewers through the process by which Nina Jablonski came to propose an explanation for why humans living in different parts of the world have different natural skin colors. Specifically, students learn how patterns in variation for the *MC1R* gene provide evidence that dark skin is favored in environments that experience intense UV radiation. As mentioned briefly in the film, however, human skin color is a polygenic trait. In Part 1 of this activity, a simple mathematical model illustrates an idealized relationship between the number of genes involved in a trait and the number of phenotypes derived from the combination of alleles. In Part 2, students learn about the methods by which geneticists identify skin color genes and estimate heritability. Finally, in Part 3, students learn how geneticists analyze genetic variations to trace an individual's ancestry and draw conclusions about the predominant ancestry of two different individuals by comparing their genetic profiles against real allele frequency data.

KEY CONCEPTS AND LEARNING TARGETS

- Scientists use mathematical models to estimate the number of genes that affect a trait. Many different genes contribute to differences in human skin color.
- Changes to a gene's DNA sequence can affect the translation of the gene into amino acids, and ultimately, the function of a protein and the expression of a trait.
- By comparing an individual's DNA against a database of DNA sequences from different populations, scientists can infer ancestry.
- Both genetics and the environment can affect expression of a trait. Experiments suggest the degree to which differences in traits are inherited. Differences in human skin color are mostly controlled by genetics.
- Mathematical models can be developed to explore how the number of genes that influence a trait affect the possible number of phenotypes.
- SNP data can be used to evaluate and make evidence-based claims about the possible genetic ancestry of individuals.

CURRICULUM CONNECTIONS

Standards	Curriculum Connection
NGSS (April 2013)	HS-LS3-1, HS-LS3-3
AP Biology (2015)	3.A.1, 3.C.1, 4.C.2; SP 1, 5
IB Biology (2016)	3.1, 7.2
Common Core (2010)	Math.A-CED.1, MP4
Vision and Change (2009)	CC3, DP3

KEY TERMS

allele, allele frequency, gene, genotype, heritability, indigenous population, locus (pl: loci), phenotype, polygenic, single-nucleotide polymorphism (SNP)

TIME REQUIREMENTS

One 50-minute class period, longer with class discussion. Students will need to have viewed the film prior to the activity.

SUGGESTED AUDIENCE

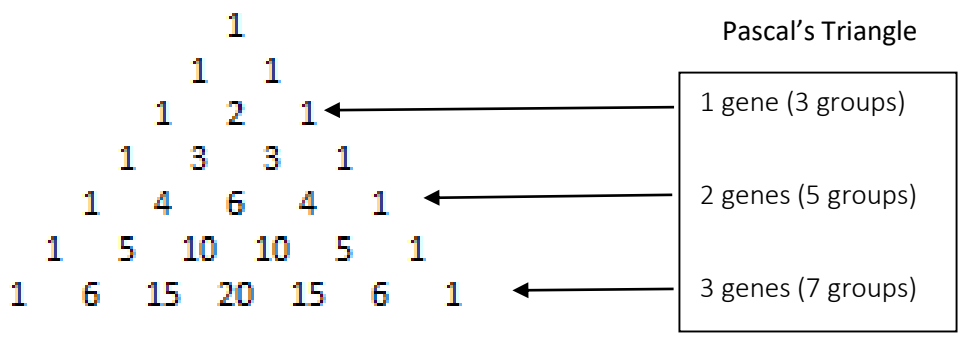
- High School: General Biology, AP Biology
- College: Introductory Biology

PRIOR KNOWLEDGE

- Students should have a basic understanding of Mendelian genetics, including the terms DNA, gene, and allele, and know that variations in some traits are inherited.
- It would be helpful for students to have prior knowledge of the terms *genotype* and *phenotype* and how to apply the terms to specific examples.
- Students should be comfortable generating and using a mathematical expression with two variables.

TEACHING TIPS

- Have students watch the short (19-minute) film *The Biology of Skin Color* before completing this activity. If you don't have sufficient in-class time, consider assigning it as homework. Have students write down any questions they have while they watch. Run through some of these questions as a warm-up or as a concluding discussion.
- Before beginning the lesson, consider reviewing genes and alleles with students. One way to accomplish this is to ask them in an open-ended class discussion to elicit everything they know about genes and alleles. Write down everything students say so all can see the list. At the end of the brainstorming session, highlight the following overarching concepts:
    - *Genes are inherited.* Genes are located on chromosomes. Chromosomes are inherited in pairs, one from each parent. Different versions of genes are called alleles. A single gene can have many alleles.
    - *Genes affect phenotypes.* Genes code for proteins, which are critical for thousands of functions within cells. The expression and action of proteins result in the distinguishable traits of an organism: its phenotypes.
- Make sure to emphasize the critical point that, while much of the data for skin color presented in this activity focuses on differences among people, comparison of genomic sequences from individuals around the world has revealed that all humans are closely related to one another and that individuals have much in common.
- Questions 1–4 ask students to answer questions about a simple mathematical model of genotype and phenotype, in which different genotypes result in identical phenotypes (for example, A^1A^0 is equivalent to A^0A^1). If students have covered the concept in mathematics, they may realize that, if they counted the number of possible combinations in each group, the pattern would be like every other row of Pascal's triangle. If you think it is appropriate, tell students that they could use this triangle to predict the number of individuals within each group if there was a cross between two parents that were heterozygous for each allele.



You may consider collaborating with a math teacher to reinforce the concepts of probability that can accompany learning about Pascal's triangle.

- When students begin Part 3, remind them that this activity is making some assumptions for simplicity. They will only explore profiles for 13 SNPs and all the SNPs are related to skin pigmentation. In reality, tracing ancestry requires looking at many SNPs throughout the entire genome and the chosen SNPs are not specific to loci involved in skin pigmentation. Second, students 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 is valid, but in other cases people may have two different alleles at a locus.

ANSWER KEY

PART 1: Using a model to understand skin color genetics

Completed Table 1, with colors corresponding to unique phenotypes. Black = 0 pigment alleles; Bright blue = 1; Red = 2; Green = 3; Purple = 4; Yellow = 5; Light blue = 6

Scenario	Possible Genotypes	Number of Unique Phenotypes
1: One gene (*A*)	A^1A^1 A^1A^0 A^0A^1 A^0A^0	3
2: Two genes (*A*, *B*)	$A^1A^1B^1B^1$ $A^1A^1B^0B^1$ $A^0A^1B^1B^1$ $A^0A^1B^0B^1$ $A^1A^1B^1B^0$ $A^1A^1B^0B^0$ $A^0A^1B^1B^0$ $A^0A^1B^0B^0$ $A^1A^0B^1B^1$ $A^1A^0B^0B^1$ $A^0A^0B^1B^1$ $A^0A^0B^0B^1$ $A^1A^0B^1B^0$ $A^1A^0B^0B^0$ $A^0A^0B^1B^0$ $A^0A^0B^0B^0$	5
3: Three genes (*A*, *B*, *C*)	$A^1A^1B^1B^1C^1C^1$ $A^1A^1B^1B^1C^0C^1$ $A^1A^1B^0B^1C^1C^1$ $A^1A^1B^0B^1C^0C^1$ $A^0A^1B^1B^1C^1C^1$ $A^0A^1B^1B^1C^0C^1$ $A^0A^1B^0B^1C^1C^1$ $A^0A^1B^0B^1C^0C^1$ $A^1A^1B^1B^1C^1C^0$ $A^1A^1B^1B^1C^0C^0$ $A^1A^1B^0B^1C^1C^0$ $A^1A^1B^0B^1C^0C^0$ $A^0A^1B^1B^1C^1C^0$ $A^0A^1B^1B^1C^0C^0$ $A^0A^1B^0B^1C^1C^0$ $A^0A^1B^0B^1C^0C^0$ $A^1A^1B^1B^0C^1C^1$ $A^1A^1B^1B^0C^0C^1$ $A^1A^1B^0B^0C^1C^1$ $A^1A^1B^0B^0C^0C^1$ $A^0A^1B^1B^0C^1C^1$ $A^0A^1B^1B^0C^0C^1$ $A^0A^1B^0B^0C^1C^1$ $A^0A^1B^0B^0C^0C^1$ $A^1A^1B^1B^0C^1C^0$ $A^1A^1B^1B^0C^0C^0$ $A^1A^1B^0B^0C^1C^0$ $A^1A^1B^0B^0C^0C^0$ $A^0A^1B^1B^0C^1C^0$ $A^0A^1B^1B^0C^0C^0$ $A^0A^1B^0B^0C^1C^0$ $A^0A^1B^0B^0C^0C^0$ $A^1A^0B^1B^1C^1C^1$ $A^1A^0B^1B^1C^0C^1$ $A^1A^0B^0B^1C^1C^1$ $A^1A^0B^0B^1C^0C^1$ $A^0A^0B^1B^1C^1C^1$ $A^0A^0B^1B^1C^0C^1$ $A^0A^0B^0B^1C^1C^1$ $A^0A^0B^0B^1C^0C^1$ $A^1A^0B^1B^1C^1C^0$ $A^1A^0B^1B^1C^0C^0$ $A^1A^0B^0B^1C^1C^0$ $A^1A^0B^0B^1C^0C^0$ $A^0A^0B^1B^1C^1C^0$ $A^0A^0B^1B^1C^0C^0$ $A^0A^0B^0B^1C^1C^0$ $A^0A^0B^0B^1C^0C^0$ $A^1A^0B^1B^0C^1C^1$ $A^1A^0B^1B^0C^0C^1$ $A^1A^0B^0B^0C^1C^1$ $A^1A^0B^0B^0C^0C^1$ $A^0A^0B^1B^0C^1C^1$ $A^0A^0B^1B^0C^0C^1$ $A^0A^0B^0B^0C^1C^1$ $A^0A^0B^0B^0C^0C^1$ $A^1A^0B^1B^0C^1C^0$ $A^1A^0B^1B^0C^0C^0$ $A^1A^0B^0B^0C^1C^0$ $A^1A^0B^0B^0C^0C^0$ $A^0A^0B^1B^0C^1C^0$ $A^0A^0B^1B^0C^0C^0$ $A^0A^0B^0B^0C^1C^0$ $A^0A^0B^0B^0C^0C^0$	7

1. Describe the basic relationship between the number of unique skin color phenotypes generated in the model and the number of genes responsible.

The number of skin color groups goes up as the number of genes involved in skin color increases.

2. Look at the number of unique phenotypes for one gene and two genes. The number of unique phenotypes for three genes is seven. Develop a mathematical expression to summarize the number of phenotypes (*P*) that can form from *N* number of genes.

   P = 2N + 1

3. Based on the mathematical expression, predict how many distinct phenotypes would result from six genes, each with two alleles, according to the model. *13*

4. 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. Based on the expression you developed above, and assuming that each of the 34 genes has two alleles, how many unique phenotypes would be generated?

   P = 2(34) + 1 = 69 distinct phenotypes

PART 2: Searching for skin color genes

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?

DNA is transcribed into mRNA that is "read," three nucleotides (a codon) at a time, by tRNA, which translates the codons to amino acids. Amino acids are strung together into a polypeptide, which is eventually processed into a functional protein. If a change to a single DNA nucleotide results in an mRNA codon corresponding to a different amino acid, then the overall structure and function of the protein might be changed. Students may also mention that a change in a regulatory sequence could alter the amount of protein being produced.

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

Since they have identical genes, any differences in phenotype will be due to environmental differences.

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

The value for heritability for skin color is reported to be 0.83. A heritability of 1.0 means all differences are because of genetics, and a value of 0.0 means all differences are because of the environment. 0.83 is closer to 1.0 than to 0.0, so the claim ("Differences in human skin color are caused primarily by differences in genetics") is supported.

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.

An environmental factor such as diet, environment during development, or any one of many other factors might inhibit or promote the activity of the proteins associated with pigment expression. Students may also mention how sun exposure can affect tanning, which is a temporary change in skin color. Sun exposure increases the amount of ultraviolet light striking skin cells, which causes DNA damage, which in turn results in changes in the expression of many genes.

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

Completed Profile 1:

			Allele Frequency in Indigenous Population			
Gene	**SNP Locus**	**Allele**	European	Chinese	Japanese	African
TYR	rs1042602	C	0.583	1.0	1.0	1.0
	rs1800422	G	0.604	1.0	1.0	0.935
	rs1126809	G	0.783	1.0	1.0	1.0
TYRP1	rs1408799	C	0.30	0.989	0.978	0.775
	rs2733832	C	0.367	0.989	0.977	0.933
OCA2	rs1800401	C	0.935	1.0	1.0	0.979
	rs1800407	G	0.933	1.0	1.0	1.0
	rs1800414	A	1.0	0.367	0.477	1.0
	rs12913832	T	0.208	1.0	1.0	1.0
SLC45A2	rs26722	G	1.0	0.611	0.591	0.95
	rs16891982	C	0.983	0.011	0.0	0.0
SLC24A5	rs1426654	G	0.0	0.989	0.989	0.975
KITLG	rs642742	A	0.136	0.267	0.114	0.922

9. Based on your analysis, is this individual of mostly European, Chinese, Japanese, or African ancestry? Explain your answer.

Based on this profile, the individual is most likely to be predominantly of Chinese ancestry. Nine of the 13 alleles are present in the highest frequency among indigenous Chinese populations, the C allele of rs16891982 is not present in the indigenous Japanese or African populations, and the G allele of rs1426654 is not present in the indigenous European population. You may wish to highlight that with these data alone, it is difficult to exclude Japanese ancestry for this individual. The differences between the Chinese and Japanese allele frequencies for these SNPs are about one percent for the three alleles that are higher in the Chinese population. Additionally, while the C allele of rs16891982 is not present in the indigenous Japanese population, the allele is only at one percent in the Chinese population. Sampling error may explain the difference between frequencies in the two groups.

Completed Profile 2:

			Allele Frequency in Indigenous Population			
Gene	**SNP Locus**	**Allele**	European	Chinese	Japanese	African
TYR	rs1042602	C	0.583	1.0	1.0	1.0
	rs1800422	G	0.604	1.0	1.0	0.935
	rs1126809	G	0.783	1.0	1.0	1.0
TYRP1	rs1408799	T	0.70	0.011	0.022	0.225
	rs2733832	C	0.367	0.989	0.977	0.933
OCA2	rs1800401	C	0.935	1.0	1.0	0.979
	rs1800407	G	0.933	1.0	1.0	1.0
	rs1800414	G	0.0	0.633	0.523	0.0
	rs12913832	T	0.208	1.0	1.0	1.0
SLC45A2	rs26722	A	0.0	0.389	0.409	0.05
	rs16891982	G	0.017	0.989	1.0	1.0
SLC24A5	rs1426654	G	0.0	0.989	0.989	0.975
KITLG	rs642742	G	0.864	0.733	0.886	0.0778

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

This individual is most likely of predominantly Japanese or Chinese ancestry. Nine of the 13 alleles are found at the highest frequencies among indigenous Chinese populations, and 10 of the 13 alleles are found at the highest frequencies among indigenous Japanese populations. In addition, neither of these indigenous populations can be excluded using this particular data set. Like the answer for the previous individual, it is difficult to distinguish between the Japanese and Chinese populations using this data set.

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.

A C allele at rs12913832 is more useful because it is found in 79.2% of individuals in indigenous European populations but 0% of individuals in indigenous Chinese, Japanese, or African populations. On the other hand, a C allele at rs1042602 is found among 100% of individuals in indigenous Japanese, Chinese, and African populations and more than half (58.3%) of people in indigenous European populations. The C allele is therefore so well conserved among these populations that it is not useful in helping to determine likely ancestry; individuals with ancestry of each of the four populations are very likely to have the allele.

12. How could additional SNP loci help to draw conclusions about the ancestry of the individual with Profile 2?

Additional SNP loci with alleles present at different frequencies among indigenous Japanese and Chinese populations could help distinguish the ancestry of the person with Profile 2.

13. If you were to develop a SNP chip using five of the loci from Table 2, which five would you choose, and why?

Answers will vary. It is most important that students justify their thinking and show that they understand that certain alleles can rule out a particular indigenous population; for example, the G allele at rs1426654 rules out

European ancestry. The SNP rs642742 is less helpful at determining ancestry because both alleles are present in all populations.

14. When a person gets their ancestry information back from the company, it is often not conclusive. Instead of stating "You are Japanese," one might be told that they were 23% European, 65% Japanese, 6% Chinese, and 6% African. How might an individual come to have such a varied ancestry?

An individual can have a varied ancestry if not all of their ancestors were from the same indigenous populations—that is, if some relatives were the result of an individual from one indigenous group having children with people from another group.

REFERENCES

Clark P., Stark A.E., Walsh R.J., Jardine R., Martin N.G. "A twin study of skin reflectance," *Annals of Human Biology* 8 (1981): 529–541.

Klug W.S., Cummings M.R., Spencer, C.A. *Concepts of Genetics*. (New York: Pearson, 2006).

Sturm R.A., Duffy D.L. "Human pigmentation genes under environmental selection," *Genome Biology* 13 (2012): 248. doi:10.1186/gb-2012-13-9-248.

Sturm R.A. "Molecular genetics of human pigmentation diversity," *Human Molecular Genetics* 18 (2009): R9–R17. doi:10.1093/hmg/ddp003.

The 1000 Genomes Project Consortium. "A global reference for human genetic variation," *Nature* 526 (2015): 68–74. doi:10.1038/nature15393.

NOTE

This activity is adapted from the Smithsonian Institution's Teaching Evolution through Human Examples project (NSF Grant No. 1119468) activity in the *Evolution of Human Skin Color* curriculum unit for AP Biology. View the full curriculum unit: http://humanorigins.si.edu/education/teaching-evolution-through-human-examples.

AUTHOR

Paul Beardsley, Cal Poly Pomona
Edited by K. David Pinkerton, educational consultant; Stephanie Keep, educational consultant; and Melissa Csikari, HHMI
Scientific Review by Rebecca Lamason, University of California, Berkeley; and Kathryn Jones, Howard Community College