



## Diet and the Evolution of Salivary Amylase

### OVERVIEW

This activity serves as a supplement to the film *Got Lactase? The Co-evolution of Genes and Culture*. The film focuses on the evolution of the ability to digest lactose, the main sugar in milk. This activity focuses on another recent human adaptation: the ability to digest foods rich in starch.

Students analyze data obtained from a research study in order to draw conclusions about the relationship between the number of copies of the salivary amylase (*AMY1*) gene and production of salivary amylase, which is the enzyme that breaks down starch in the mouth, and the relationship between *AMY1* gene copy number and dietary starch consumption among different populations. This activity involves analyzing research data and graphing, using scientific reasoning to make claims, and using statistics to support these claims.

### KEY CONCEPTS

- Compounds in food are sources of energy for cells in the body. These compounds first have to be broken down into simple molecules that can be absorbed and used by cells.
- Digestive enzymes, such as salivary amylase, facilitate the breakdown of food molecules, such as starch.
- Changes in human diet, such as starting to use milk products after the domestication of animals or eating increasing amounts of starch-rich foods, have resulted in new evolutionary adaptations. For example, the *AMY1* gene, which produces the enzyme amylase, has undergone duplication events in humans, resulting in copy number variations among human populations.

### STUDENT LEARNING TARGETS

- Graph research data and appropriately label all graph components, including title, axes, units, and legend;
- Interpret primary data from different research studies in order to draw conclusions;
- Make valid claims based on scientific data and support those claims using scientific reasoning;
- Analyze research data using correlation and regression tests;
- Use a statistical test, such as the *t*-test, to help support evidence-based claims by defining uncertainty.

### CURRICULUM CONNECTIONS

Standards	Curriculum Connection
NGSS (April 2013)	HS-LS1-3, HS-LS3-1, HS-LS3-3, HS-LS4-2, HS-LS4-3, HS-LS4-4
Common Core (2010)	ELA.RST.9-12.4, RST.9-12.7, WHST.9-12.1 Math.S-ID.1, S-ID.2, S-ID.4, S-ID.6, S-IC.1, S-IC.4, F-IF.7; MP2, MP3, MP4
AP Biology (2012–13)	1.A.2, 3.A.1, 3.A.3, 3.C.1, 4.A.1, 4.C.1, SP1, SP2, SP3, SP5, SP6
IB Biology (2016)	2.4, 2.7, 3.1, 5.2, D.1
IB Env Systems and Societies (2017)	5.2
Vision and Change (2009)	CC2, DP1, DP2

### KEY TERMS

amylase, enzyme, gene, null hypothesis, starch, uncertainty

## TIME REQUIREMENT

- One and a half 50-minute classroom periods with assignment of a small amount of homework. Viewing the short film prior to the activity requires an additional 15 minutes. The math extensions will require additional time in class or as homework.

## SUGGESTED AUDIENCE

- Advanced high school Biology (AP/IB)
- College-level biology courses.

## PRIOR KNOWLEDGE

Students should have

- prior experience with constructing graphs, performing statistics calculations (i.e., mean, standard deviation, correlation coefficient, and  $t$ -test), and making and justifying claims with evidence and scientific reasoning.
- a general understanding of enzymes, carbohydrates, and gene expression.

## MATERIALS

- Scientific calculator and/or a computer with a spreadsheet program like Excel; stats program for  $t$ -test
- Graphing paper if not using a spreadsheet program
- Colored pencils for graphing if not using a spreadsheet program
- Ruler for graphing if not using a spreadsheet program
- Extra resources (print or online) for statistics instruction and support if needed

## TEACHING TIPS

- Students should watch the short film *Got Lactase? The Co-evolution of Genes and Culture* (<http://www.hhmi.org/biointeractive/making-fittest-got-lactase-co-evolution-genes-and-culture>) either before or after doing this activity.
- You may wish to have your students work in pairs.
- This activity includes two extensions for students who are familiar with calculating correlation coefficients and performing  $t$ -tests. Students who are not familiar with these calculations may skip the extensions.
- If you are unfamiliar with selecting appropriate graphs and performing statistical tests, please refer to the Answer Key below for further explanation.
- For additional background in statistics, refer to the “Teacher Guide: Math and Statistics” (<http://www.hhmi.org/biointeractive/teacher-guide-math-and-statistics>). You may print out selected pages from the guide for your students.
- As a further extension to this activity, ask students to design a study to further investigate the evolutionary relationship between diet and amylase production. Their plan should include
  - a new research question;
  - the hypothesis you are testing;
  - the variables you will measure, including units; and
  - a measurable prediction.

## ADDITIONAL BACKGROUND ABOUT AMYLASE

- All vertebrate species have genes that encode amylase. One version of the gene is expressed in the pancreas, and amylase is secreted into the small intestine. A few species of mammals, including some species of bats, rabbits, rodents, and primates, also produce amylase in their saliva.

- Starch is an insoluble semicrystalline structure. It cannot be dissolved in water or in stomach acid, which is why an enzyme like amylase is needed for hydrolysis. Salivary amylase can start breaking down starch as soon as it is eaten.
- The amylase enzyme accounts for 40% to 50% of protein in human saliva.
- The number of copies of the salivary amylase gene varies greatly among individuals. Humans are diploid organisms, meaning that they have two copies of any gene (one from each parent), except for genes on the sex chromosomes. But when it comes to salivary amylase, people have anywhere from two to 15 copies of the gene on just one chromosome.
- The closest living relative to humans, the chimpanzee, has two diploid copies of the *AMY1* gene. Unlike humans, chimpanzees don't exhibit copy number variation in the gene.
- One of the conclusions that students will be able to draw from this activity is that the number of diploid copies of the *AMY1* gene in people and the amount of salivary amylase they produce are correlated. The correlation is not strong, but it is statistically significant. At the end of the activity, ask your students what other factors, genetic and/or environmental, might affect amylase production.

### ANSWER KEY

1. List the enzymes involved in starch digestion, including where they work and what they do.

*Amylases use hydrolysis to break covalent bonds. Salivary amylase works in the mouth and pancreatic amylase works in the small intestine. Maltase, which works in the small intestine, breaks maltose (a glucose-glucose disaccharide) into glucose.*

2. Name the gene that produces salivary amylase. How many copies of the gene do humans have?

*The *AMY1* gene on chromosome 1 produces salivary amylase. Humans have between 2 and fifteen diploid copies of the gene.*

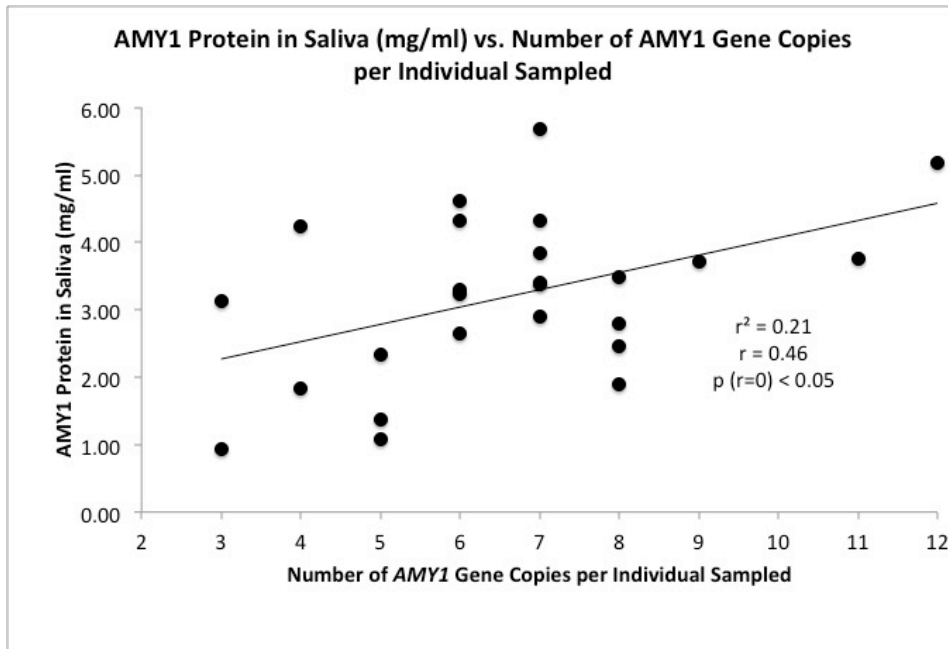
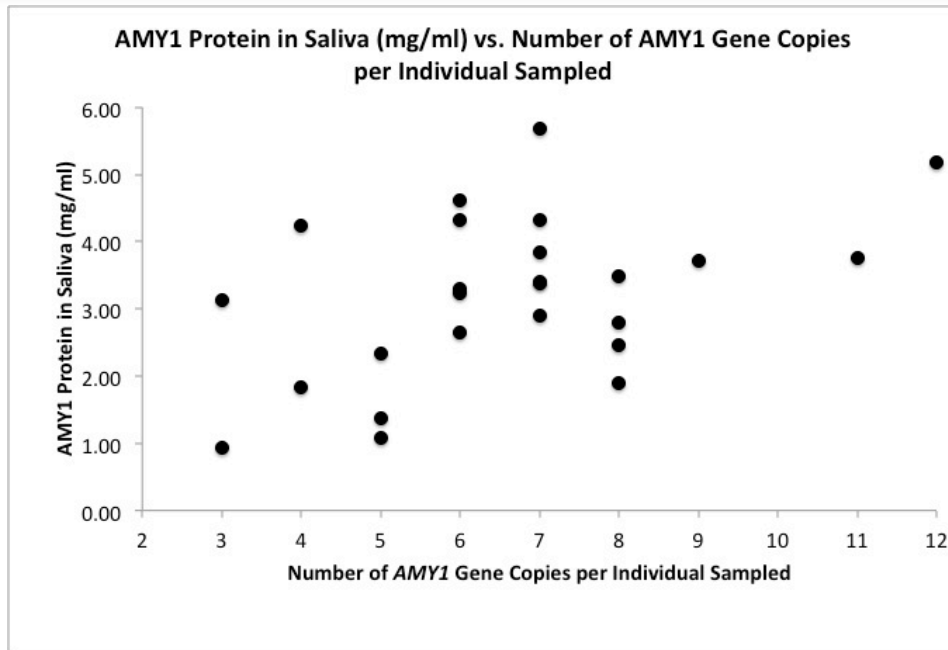
3. The researchers hypothesized that increased consumption of starch-rich foods during human evolutionary history gave individuals with multiple copies of the *AMY1* gene an "advantage." In this context, what does advantage mean?

*Having more copies of the *AMY1* gene might mean that those individuals produce more salivary amylase, are able to digest more starch, and are able to obtain more energy from starches than individuals with less amylase. Students might also expand their answer to suggest that these individuals are more likely to survive and reproduce, passing on their genes.*

### Part A: Relationship Between *AMY1* Copy Number and Amylase Production

1. The data in Table 1 show the number of *AMY1* gene copies and milligrams of amylase protein per milliliter of saliva in 25 adults. On a separate sheet of paper, construct and label a graph that illustrates the relationship between these two measured variables. Include a title for your graph and labels for the x- and y-axes.

*The proper graph for this exercise is a scatterplot since students are comparing one measured variable against another to look for trends or associations. Some students may be tempted to average the two sets of data and compare them with a bar graph. This is not the appropriate comparison because the numbers in each data set have completely different identities (units). The following graph represents the concentration of *AMY1* protein (mg/ml) in saliva as a function of the number of *AMY1* gene copies in humans from cultures with a traditionally high-starch diet. A second graph is included below the first, showing a regression line (also known as a trend line or line of best fit). That graph also includes the correlation calculation results. (See the math extension below for an explanation of the correlation calculations.)*



- From the graph, do the two variables appear to be associated? **Yes** Explain your answer.  
*AMY1 protein concentration in saliva appears to increase as the number of AMY1 gene copies increases. (Note to teachers: Students may be tempted to conclude that the graph is “proof” that an increase in gene copy number causes more AMY1 protein to be produced. Make it clear to students that “correlation does not necessarily imply causation.” However, an observed correlation can suggest hypotheses that can be tested.)*
- Based on the graph, make a claim regarding the number of AMY1 gene copies and how they relate to the concentration of AMY1 protein in saliva. Explain the claim using scientific reasoning.  
*Based on these data, it appears that the more copies of the AMY1 gene a person has, the more amylase will be produced and measured in each milliliter of saliva. (Note that unless we do a linear regression test, we cannot say whether the correlation between these two variables is statistically significant.)*

4. Based on your claim, suggest a hypothesis to explain why more or fewer copies of a gene would affect the amount of protein produced.

**Answers may vary. One example: The more copies a person has of a particular gene, the more messenger RNA (mRNA) can be transcribed in a given amount of time from that gene, and the more protein can be synthesized by translating that mRNA.**

5. Suppose that you analyzed the number of copies of the *AMY1* gene in two individuals. Individual A has four diploid copies of the *AMY1* gene, and individual B has eight diploid copies of the *AMY1* gene. Can you say with certainty who will have more amylase enzyme in their saliva? Use evidence from your graph and/or Table 1 to support your answer.

**There is a positive correlation between the number of *AMY1* gene copies and the amount of amylase in saliva, but that does not mean that someone with eight copies of the *AMY1* gene will have more amylase in their saliva than someone with four copies of the gene. In the graph, one individual has four copies of the gene and more protein in their saliva than anyone in the sample with eight copies of the gene. Therefore, there must be other factors that affect amylase production.**

### Part B: Relationship Between *AMY1* Copy Number and Dietary Starch

1. The data in Table 2 represent the number of *AMY1* gene copies in two groups of populations with different diets. For each diet-profile population, determine the sample sizes and then calculate the means, standard deviations, and 95% confidence interval for the data. Enter your answers in Table 3 below.

**Table 3.** Calculations for Sample Groups from Table 2

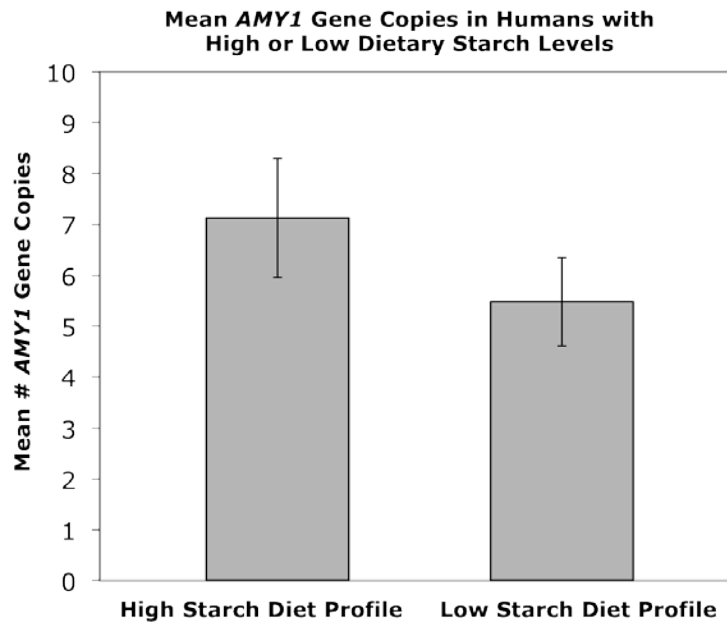
	High-Starch Diet Profile	Low-Starch Diet Profile
Sample Size ( <i>n</i> )	25	25
Mean ( $\bar{x}$ )	7.12	5.48
Standard Deviation ( <i>s</i> )	2.92	2.18
95% Confidence Interval ( $\frac{2s}{\sqrt{n}}$ )	1.168	0.872

Note to instructors: The formula provided for calculating 95% confidence interval (CI) is the standard deviation multiplied by two divided by the square root of the sample size. This calculation provides an estimate of the 95% CI. The actual CI calculations are done relative to the appropriate t-statistic, which is approximately 2 for large datasets. Depending on the class you are teaching, you can have your students calculate standard error of the mean instead of 95% confidence interval. The formula for SEM is  $\left(\frac{s}{\sqrt{n}}\right)$ .

2. On a separate sheet of paper, construct and label a graph that summarizes the data in Table 3. Include 95% confidence interval error bars, as well as a title for your graph and labels for the x- and y-axes.

**The proper graph for this exercise is a bar graph because it is a comparison of two samples [or sample groups]. Since the variable (number of *AMY1* gene copies) is measured, the best graph to represent a summary of the data is a bar graph of the means of the two samples with error bars indicated.**

Note to instructors: The following graph summarizes the data in Table 3 and includes error bars. Note that the error bars are the 95% confidence intervals in this particular graph. These error bars are often explained as being 95% certain that the confidence interval contains the true mean for the population. However, that's technically incorrect. What 95% confidence intervals means is that if you took repeated random samples from a population and calculated the mean and confidence limits for each sample, the confidence interval for 95% of your samples would include the population mean. So, in other words, there is a good chance that the 95% confidence interval of your sample contains the true population mean.



3. From the graph, how do the two diet profiles compare? Explain your answer.

*There appears to be a correlation between a higher dietary starch level and an increased number of AMY1 gene copies. Likewise, there also appears to be a correlation between low dietary starch and fewer copies of the AMY1 gene. Again, correlation does not necessarily indicate causation.*

4. Suggest a scientific hypothesis (a testable explanation) to explain your answer to question 3 above.

*Answers may vary. One example: Individuals who get the majority of their calories from starchy foods may have an evolutionary advantage over other individuals if they have extra copies of the AMY1 gene and therefore can produce more salivary amylase. This advantage comes from the increased ability to begin the chemical breakdown of energy-rich starches in the mouth. Even the smallest increase in survival probability, especially in times when other food sources are scarce, can result in an increase in the mean number of copies of the AMY1 gene in a population over time—a classic example of evolution by natural selection.*

## MATH EXTENSIONS

### Math Extension for Part A

1. Use Pearson's correlation coefficient from a linear regression test to test the null hypothesis that these two variables listed in Table 1 are not correlated ( $r = 0$ ). What do you conclude based on your calculations?

*Table 1 in the student activity presents data for independent ( $x$ ; number of AMY1 gene copies) and dependent ( $y$ ; AMY1 protein concentration in saliva) variables. Simple linear regression tests the null statistical hypothesis that there is no relationship between the  $x$  and  $y$  variables and any observed relationship occurred by chance. The statistic to use in this case is the Pearson's correlation coefficient ( $r$ ), which can range from  $-1$  to  $1$ . If the null hypothesis is true, then  $r = 0$ .*

*Critical  $r$ -values can be found online or in statistics books in  $r$ -value tables. Critical  $r$ -values vary depending on the sample size of the data set and the previously determined rejection level, which is usually set at  $p = 0.05$ . For the data in Part A,  $r = 0.46$  and the probability ( $p$ ) of getting an  $r$ -value as large as  $0.46$  by chance is less than  $0.05$ . It is therefore safe to reject the null statistical hypothesis, thereby adding strength to the claim that multiple copies of the AMY1 gene can lead to increased production of salivary amylase.*

2. Use a statistical test to determine the strength of the correlation between AMY1 gene copy number and salivary amylase production. What do you conclude based on your calculations?

The coefficient of determination ( $r^2$ ) tells us the strength of the relationship between  $x$  and  $y$  and can be interpreted in two ways. First, if  $r^2 = 0.94$ , then you can say that the independent variable ( $x$ ) predicts the dependent variable ( $y$ ) with 94% accuracy. Second, you can say that the pairs of coordinates have 94% of their variance in common. In other words, 94% of the observed variance in  $y$ -values is associated with the variance in  $x$ -values. In this case,  $r^2$  is 0.21, which means it is not a very strong correlation (but it is statistically significant).

### Math Extension for Part B

1. Perform a  $t$ -test to find out whether this difference is real and not simply due to chance. Explain your results.

*The  $t$ -test is used to test the null statistical hypothesis ( $H_0$ ) that the mean number of gene copies is equal for individuals consuming high-starch and low-starch diets, and any observed differences are likely to have occurred by chance. The  $t$ -test for the Perry data returns a  $t_{\text{obs}}$  ( $t$  stat) of 2.250. This is higher than the critical  $t$ -value of 2.015 at a rejection level ( $\alpha$ ) of 0.05. The probability ( $p$ ) of getting a  $t_{\text{obs}}$  as large as 2.250 by chance alone is 0.030. We can therefore reject the idea that  $H_0$  is true and conclude with some confidence that the differences between the two means are not likely to have occurred by chance and that the differences observed are very likely real. It is important to remember that absolute certainty is not possible, but the statistical test allows us to estimate our uncertainty. Students may notice that the 95% confidence interval error bars overlap. The 95% confidence interval error bars only show uncertainty in the mean from which they are calculated. Based on the relationship between two error bars, we may wonder if two means are statistically distinguishable or not. The  $t$ -test gives us that answer.*

2. Based on the data and statistics, make a claim regarding what might explain the difference in *AMY1* copy number between the two groups of populations.

*The two groups of populations are very likely different. This may be due to differences in diet that could lead to selective evolutionary pressure for a higher number of *AMY1* gene copies if an individual or population has a high-starch diet and a lower number of *AMY1* gene copies if an individual or population has a low-starch diet.*

### REFERENCES

- Perry, G. H. et al. 2007. Diet and the evolution of human amylase gene copy number variation. *Nature Genetics* 39:1256–1260.
- Laden, G., and R. Wrangham. 2005. The rise of the hominids as an adaptive shift in fallback foods: plant underground storage organs (USOs) and australopith origins. *Journal of Human Evolution* 49:482–498.
- McGeachin, R. L., and J. R. Akin. 1982. Amylase levels in the tissues and body fluids of several primate species. *Comparative Biochemistry and Physiology. A, Comparative Physiology* 72:267–269.
- Meisler, M. H., and C.-N. Ting. 1993. The remarkable evolutionary history of the human amylase genes. *Critical Reviews in Oral Biology and Medicine* 4:503–509.

### AUTHOR

Written by Paul Strode, PhD, Fairview High School, and Laura Bonetta, PhD, HHMI.

Edited by Susan Dodge, consultant.

Reviewed by Brad Williamson, University of Kansas, and Nathaniel Dominy, PhD, Dartmouth University.

Field Tested by David Knuffke, Deer Park High School; David Prescott, St. John's Ravenscourt School; Karin Marcotullio, Ballston Spa High School; Kim Hayen, Heritage High School; Mark Little, Broomfield High School; Stacie Deyglio, Linden Hill.