# ScienceLIVE

## Pika Population Genetics

Grades: 9-12

## LESSON DESCRIPTION:

The goal of this lesson is to answer the question: Are pikas living at low elevations genetically different than those at high elevations? Students investigate this question in two ways, both commonly used by modern population geneticists. First, students investigate general gene flow in these populations by looking at the proportion of individuals that are heterozygous in a population. Second, they will look at single nucleotide polymorphism (SNP) data to identify DNA regions (loci) that may be undergoing natural selection and will use their creativity to guess possible functions of these genes. Finally, students will develop methods for testing their gene's function and consider the ability of low elevation pikas to survive in the future.

## LEARNING GOALS:

1. Reaffirm Mendelian genetics concepts and be able to test a population for inbreeding and understand the consequences of inbreeding on a population
2. Understand what a SNP is and how scientists use SNPs to understand natural selection
3. Improve math and spreadsheet skills, as well as graphing and interpreting data trends
4. Develop a scientific proposal for studying pika adaptation at different elevations

## PREREQUISITES

This lesson is intended for an advanced biology audience. Students should be familiar with the structure of DNA, alleles, genes, loci, heterozygosity, inbreeding, and have a general sense for how natural selection works prior to this lesson. ScienceLIVE's Pika Body Size and Adaptation lesson (available at www.science-live.org/teachers) is a nice adaptation introduction lesson.

## MATERIALS

- This is a computer-based lesson that should be conducted in a computer lab in which students can work individually or in small groups. The lesson uses Google Spreadsheets, however a Google account is not necessary to use that data. You can log in using scienceliveoutreach, password: k*12class if needed.
- Students will receive a worksheet and an instruction sheet, available at www.science-live.org/teachers.

## STANDARDS

**High School Next Generation Science Standards Addressed*:**

- HS-LS3-3. Variation and distribution of expressed traits in a population.
- HS-LS4-2. Evolution.
- HS-LS4-3. Advantageous heritable traits.
- HS-LS4-4. Natural selection.
    - LS4.C: Adaptation

## Part 1: Inbreeding on an Elevational Gradient

1. To begin this exercise, it is helpful to start with some basic Punnett squares to demonstrate the following (these can be for any sample trait you choose):
    a. Tt x Tt = 1:2:1 ratio, with 50% of offspring heterozygous
    b. TT x tt = All heterozygous offspring
    c. TT x Tt = 50% heterozygous
    d. tt x Tt = 50% heterozygous
    e. TT x TT = no heterozygotes
    f. tt x tt = no heterozygotes

2. This demonstration should remind students that heterozygotes are made through many different parental combinations, and in a healthy, non-inbred population, heterozygosity should be relatively high. The lower the heterozygosity, the more likely the population has been subject to inbreeding.

3. Next, lead students in a discussion for what might cause inbreeding in a population. The goal here is to lead students to the idea that inbreeding occurs due to reproductive isolation of some kind. This can be because populations are small, far apart, or both (along with many other social reasons, but we'll focus on these two).

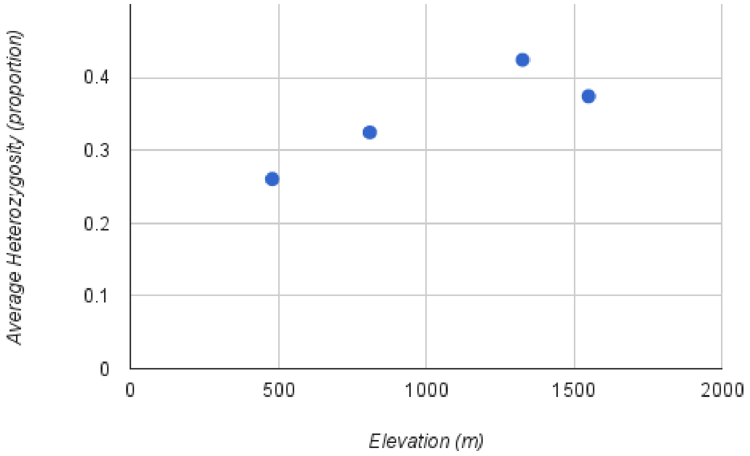
4. Usually we talk about heterozygosity at particular loci (locations) or genes (like Mendel's yellow and green peas). In this lesson we will focus on specific loci that are the sites of single nucleotide polymorphisms (SNPs, or snips). The University of Utah has put together a nice description and demonstration of SNPs, which you can find here: http://learn.genetics.utah.edu/content/pharma/snips/ If your students aren't familiar with SNPs, having them run through the "What is a SNP" chapter on this site should help. This can also be a nice chance to review mutations and the errors in cellular processes that could lead to an initial nucleotide substitution in an individual.

5. You can now introduce students to pikas via this video: (https://vimeo.com/122452931)

6. The first question students will be answering today is whether pika populations at all elevations in the North Cascades are healthy, or if some populations are too small and fragmented, causing inbreeding. To answer this, students will follow the instruction sheet to summarize and graph heterozygosity data in Google Spreadsheets.

7. This part of the lesson is fairly cookie-cutter, leading students step-by-step through the analysis and graphing exercise. We STRONGLY recommend that you run through this lesson on your own before class, so you are ready to troubleshoot with students.

8. Their final graph should be something like this:



9. *Question 15:* Students should notice that heterozygosity increases with elevation, meaning low elevation populations are the most inbred. This is likely due to a lack of population connectivity at low elevations, since pikas are more suited to cooler, wetter climates at higher elevations. The less suitable habitat at low elevations probably leads to less successful movement between populations, as well as smaller populations, both of which increase the likelihood of mating with a close relative.

### Part 2: Identifying adaptive loci (SNPs)

We often teach genetics as Mendel first studied the subject – we know there is a "heritable unit" for green peas and a "heritable unit" for yellow peas, one is dominant over the other, etc. But Mendel didn't know there were genes! Others figured out that DNA was behind the scenes of these phenotypes. Now we know there are genes at different loci, comprised of alleles that represent different nucleotide sequences at a specific locus. Genetic analyses have gotten more complicated since Mendel's day. One reason for this is that we often have to work backwards to figure out what a gene does. Often, genetic analyses are done as follows:

- Step One: Find regions of the genome that don't look like the rest – regions that show signs of selection because their allelic frequencies are too high or low, relative to the rest of the genome (remember ~95% of our genetic code doesn't directly code for proteins). Because the non-coding regions of DNA are changing at a background, normal rate of mutation, any region (locus) that shows substantially more or fewer changes than this background rate gives us a sign that it may be undergoing selection.
- Step Two: Once outlier loci have been identified, scientists then experiment to figure out what those regions do through correlation of different phenotypes. The Utah SNP site has a nice description of this process as well, under the "What is a Haplotype" and "Applying SNP Profiles to Drug Choices" chapters of the SNP site.

Today we are going to see if we can pick out regions of the pika genome in North Cascades National Park that appear to be undergoing selection based on elevation. You can begin this section of the lesson by showing this overview of ecological genetics research in the North Cascades: https://vimeo.com/122543923

1. After the very recipe-like Part 1, Part 2 of this lesson is designed to give students more freedom to develop an analysis plan and theories about elevational adaptation. Students should read the introductory page (Page 5 of the Instruction Sheet) and develop a plan for calculating allele frequencies. You may want to have them discuss their plan before enacting it.

2. They can achieve allele frequencies fairly easily, by using the =COUNT() and =COUNTIF() functions in Google Spreadsheets. The most efficient way of doing this is to use this equation:

   =COUNTIF(B2:C9,”1”)/COUNT(B2:C9)

   and

   =COUNTIF(B2:C9,”2”)/COUNT(B2:C9)

   This will calculate the proportion of alleles present in low elevation populations that are Allele 1 and Allele 2, respectively. Note that the sum of the frequency of Allele 1 and Allele 2 should equal 1.

3. Once students have calculated all their frequencies, they should get the following:

| | SNP 1306 | | SNP 3453 | | SNP 3606 | |
|---|---|---|---|---|---|---|
| | Allele 1 | Allele 2 | Allele 1 | Allele 2 | Allele 1 | Allele 2 |
| Low | 0 | 1 | 0.5 | 0.5 | 1 | 0 |
| Mid-Lo | 0.5 | 0.5 | 0.5 | 0.5 | 1 | 0 |
| Mid-Hi | 0.4 | 0.6 | 0.5 | 0.5 | 1 | 0 |
| High | 0.95 | 0.05 | 0.5 | 0.5 | 0.95 | 0.05 |

4. These three SNP loci show very different patterns. SNP 1306 shows clear elevational selection, with Allele 1 becoming more prevalent as you move up the mountain, and Allele 2 showing the opposite trend. Students should note this on their worksheets, indicating that Allele 1 appears to be a high-altitude adaptation of some sort, while Allele 2 is a low-altitude adaptation (though again, these are two sides of the same coin).

5. For *Question 3* on the students' worksheet, they can come up with any number of gene functions. These can be creative and fun, but we would encourage you to steer students toward what a gene actually does: coding for protein. So if students think the gene impacts fur thickness, perhaps there is a stronger keratin protein created by this gene. If students think that Allele 1 of the gene aids pikas in acquiring oxygen, perhaps it codes for one of hemoglobin's subunits. Again, there are myriad possibilities here!

6. For *Question 4* on the worksheet, students should think about how to figure out what a gene of interest does. Most scientists begin by collecting lots of data – in this case you would need to get lots of pika genetic samples and, if we go with the hair thickness hypothesis, they would need hair samples from each of them as well. A scientists would then analyze the keratin content of all the hair samples and run statistics to determine if the presence of Allele 1 on SNP 1306 correlates to higher keratin levels in hair. Again, there are many ways to achieve this, but in all cases the animals' phenotypes (e.g. hair thickness) will need to be tested based on genotype.

7. What about the other two SNPs? In *Question 5*, students should note that neither of the other two loci appear to be adaptive for elevation. In the case of SNP 3453, nearly every individual is heterozygous at the locus, while for SNP 3603, only one individual has Allele 2. This could mean that Allele 2 is either a very new mutation or has been strongly selected against throughout these pika populations.

8. For *Question 6*, there is no one right answer. Our inbreeding analyses indicate that low elevation populations aren't doing as well as high elevations, and yet finding an adaptive locus like SNP 1306 is a sign of hope that there may be adaptive potential in these populations. It really comes down to what the genes at this adaptive locus and others like it do. If they represent minor changes in pikas' phenotypes, it may not be enough to save these populations, but if the adaptation represents a large improvement in survival or reproductive success at low elevations, they may be able to sustain further environmental change.

## Student Instruction Sheet

**Reminder of Key Genetics Terms:**

Gene: A unit of inheritance usually occurring at specific locations (loci) on a chromosome. A sequence of nucleotides that specify the order of amino acids in a protein or part of a protein.

Alleles: Alternate forms or varieties of a gene, e.g. alleles for attached or unattached earlobes.

Locus (plural = loci): A location on a chromosome.

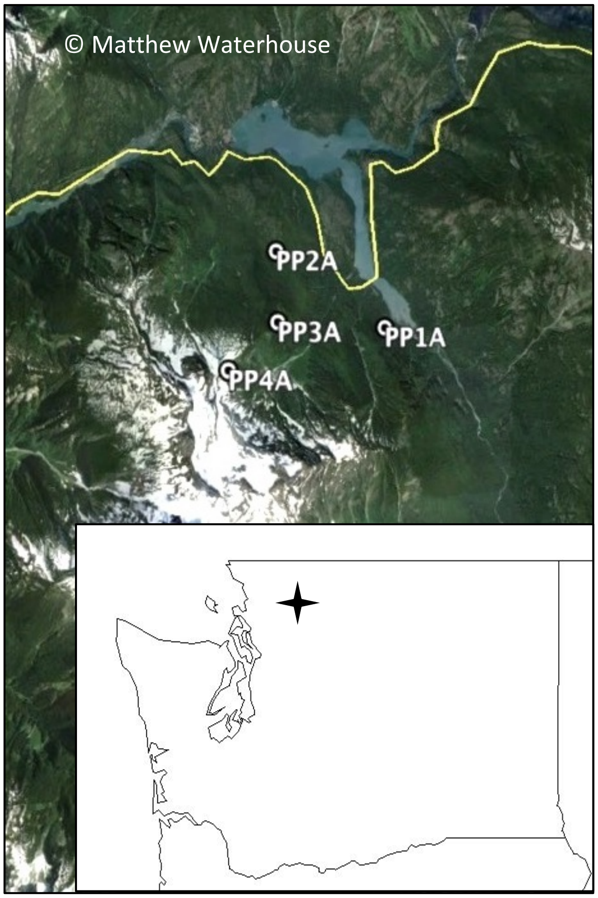
Heterozygous: Having two different alleles for a particular gene (e.g. Aa).

Inbreeding: The breeding of two individuals who are closely related to one another.

SNP: Single nucleotide polymorphism – a sequence of DNA in which a single nucleotide has been substituted (e.g. A**T**AGGCCTTA vs. A**A**AGGCCTTA).

### Part 1: Inbreeding on an Elevational Gradient

Are low- and high-elevation pika populations equally healthy? To answer this question, we'll be looking at the proportion of individuals that are heterozygous in four populations, each at a different elevation in the mountains of North Cascades National Park in Washington. All of the sites are on Pyramid Peak, and are numbered from the lowest elevation (PP1A; 480 m) to highest elevation (PP4A; 1550 m).



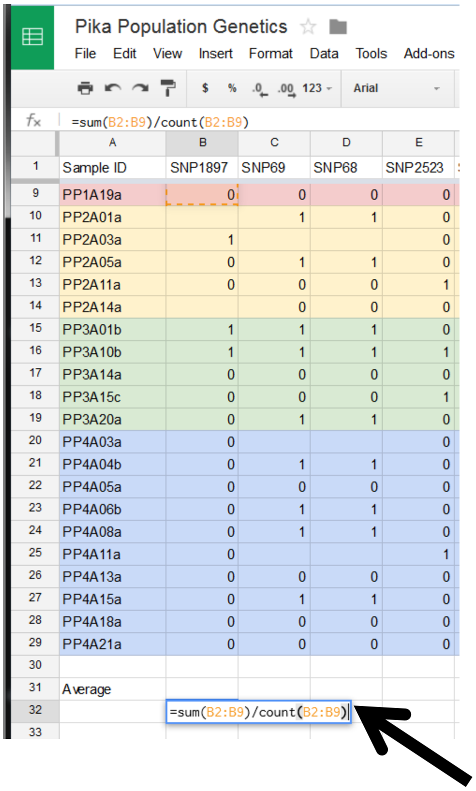
1. Before you start your analysis, develop a prediction. Based on your class discussions and what you know about pikas, do you think pika populations will be healthier (have higher heterozygosity) at high, middle, or low elevations? Why? (Answer on your worksheet)

2. Calculating heterozygosity is just as it sounds: we need to figure out the proportion of individuals in each population that are heterozygous at each locus. This can be intimidating task, so this lesson's creators have cut a few corners for you.

3. Open the Pika Population Genetics Google Spreadsheet (https://docs.google.com/spreadsheets/d/1l3BALWWwZz6gQbRVN94TahNrj1iuy7YYuGPuKc7Y1jw/edit?usp=sharing).

4. IMMEDIATELY resave this document in your own Google Drive account by clicking on **File →Make a Copy** and rename it with your own name. If you don't have your own Google account, ask your teacher for the ScienceLIVE login.

5. To begin our investigation, start in the "Heterozygosity_RawData" tab. At first this sheet may look like a LOT of data (and it is!), but it isn't as intimidating as it looks.
    - The first column (Column A) includes all the pikas sampled. The first three letters of each name tells you the pika's home population (e.g. pikas starting with PP1… are from the low elevation site). The populations are also color-coded.
    - The rest of the columns (B through AG) each represent a SNP locus. In the columns below each SNP name, cells are either blank, 0s, or 1s.
        - 1 = heterozygous at the locus
        - 0 = homozygous at the locus
        - Blank cells = individuals for which the locus couldn't be sequenced
    - Remember, our goal is to calculate the proportion of individuals in each population that are heterozygous at each locus. We'll start by calculating heterozygosity (H) in the low elevation population for SNP1879. Since we've given heterozygous individuals a value of 1, we can calculate H by entering the following equation in cell B32: "=sum(B2:B9)/count(B2:B9)". Hit enter.
    - Notice that the cells we referenced (B2:B9) are the heterozygosity values from SNP1897 for the low elevation.
    - The equation we created divides the number of heterozygous individuals (all the 1s) by the total number of individuals.

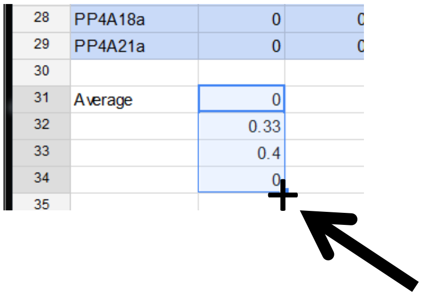


=sum(B2:B9)/count(B2:B9)

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | Sample ID | SNP1897 | SNP69 | SNP68 | SNP2523 |
| 9 | PP1A19a | 0 | 0 | 0 | 0 |
| 10 | PP2A01a | | 1 | 1 | 0 |
| 11 | PP2A03a | 1 | | | 0 |
| 12 | PP2A05a | 0 | 1 | 1 | 0 |
| 13 | PP2A11a | 0 | 0 | 0 | 1 |
| 14 | PP2A14a | | 0 | 0 | 0 |
| 15 | PP3A01b | 1 | 1 | 1 | 0 |
| 16 | PP3A10b | 1 | 1 | 1 | 1 |
| 17 | PP3A14a | 0 | 0 | 0 | 0 |
| 18 | PP3A15c | 0 | 0 | 0 | 1 |
| 19 | PP3A20a | 0 | 1 | 1 | 0 |
| 20 | PP4A03a | 0 | | | 0 |
| 21 | PP4A04b | 0 | 1 | 1 | 0 |
| 22 | PP4A05a | 0 | 0 | 0 | 0 |
| 23 | PP4A06b | 0 | 1 | 1 | 0 |
| 24 | PP4A08a | 0 | 1 | 1 | 0 |
| 25 | PP4A11a | 0 | | | 1 |
| 26 | PP4A13a | 0 | 0 | 0 | 0 |
| 27 | PP4A15a | 0 | 1 | 1 | 0 |
| 28 | PP4A18a | 0 | 0 | 0 | 0 |
| 29 | PP4A21a | 0 | 0 | 0 | 0 |
| 30 | | | | | |
| 31 | Average | | | | |
| 32 | | =sum(B2:B9)/count(B2:B9) | | | |
| 33 | | | | | |

6. Repeat this process for the other three populations for SNP1987.

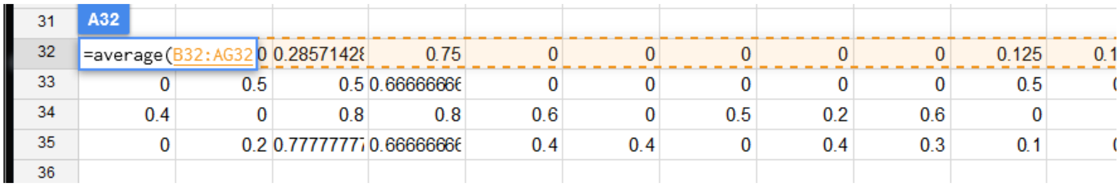
7. Rather than typing in the formulas for all of the other SNPs, there is a shortcut! Highlight your four calculated H values. Notice that once those four cells are highlighted, there is a blue square in the bottom right of the bottom-most cell.



| | | |
|---|---|---|
| 28 | PP4A18a | 0 |
| 29 | PP4A21a | 0 |
| 30 | | |
| 31 | Average | 0 |
| 32 | | 0.33 |
| 33 | | 0.4 |
| 34 | | 0 |
| 35 | | |

When you hover your mouse over that blue square, it will change from an arrow to a plus sign (see bottom of Page 2). Drag this plus sign to the right, all the way to the column AG. If all went well, your H values should have been calculated for you! If this didn't work, ask your teacher for help.

8. One more step: calculate the average H across all loci. There is a function in Sheets for this. Type "=average(B32:AB32)" to calculate the average of all low elevation loci.



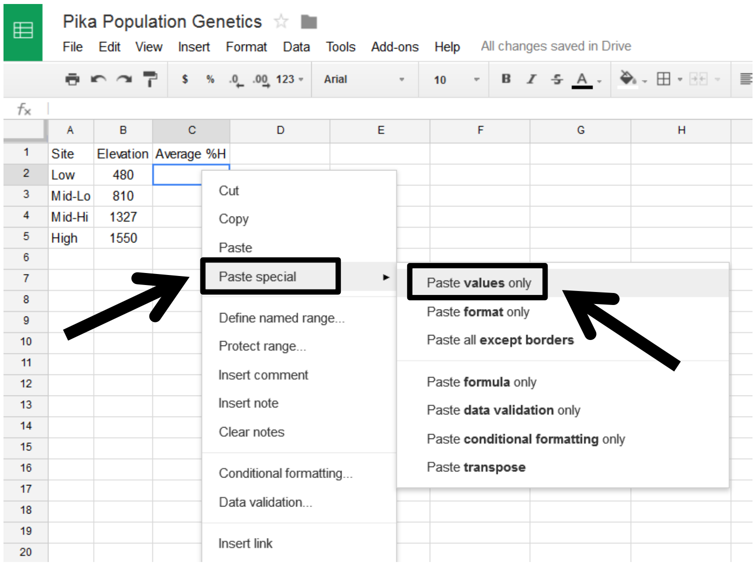
| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 31 | | | | | | | | | | | |
| 32 | =average(B32:AG32 | 0 | 0.2857142 | 0.75 | 0 | 0 | 0 | 0 | 0 | 0.125 | 0.1 |
| 33 | 0 | 0.5 | 0.5 | 0.6666666 | 0 | 0 | 0 | 0 | 0 | 0.5 | |
| 34 | 0.4 | 0 | 0.8 | 0.8 | 0.6 | 0 | 0.5 | 0.2 | 0.6 | 0 | |
| 35 | 0 | 0.2 | 0.7777777 | 0.6666666 | 0.4 | 0.4 | 0 | 0.4 | 0.3 | 0.1 | |
| 36 | | | | | | | | | | | |

9. Repeat for the other three populations (you can use the corner drag trick you used in #7 again).

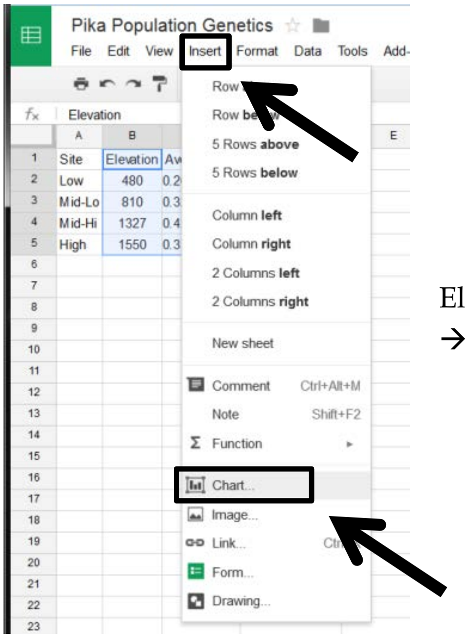
10. We now have average heterozygosity for our four populations. How does it change with elevation? Let's graph it to visualize this pattern.

11. To graph this data, highlight and copy the four averages you just calculated.

12. Click on the Heterozygosity_Summary" tab and right-click in the first cell below Average %H (cell C2). Select "Paste special" → "Paste values only."



| | A | B | C |
|---|---|---|---|
| 1 | Site | Elevation | Average %H |
| 2 | Low | 480 | |
| 3 | Mid-Lo | 810 | |
| 4 | Mid-Hi | 1327 | |
| 5 | High | 1550 | |



| | A | B | |
|---|---|---|---|
| 1 | Site | Elevation | Av |
| 2 | Low | 480 | 0.2 |
| 3 | Mid-Lo | 810 | 0.3 |
| 4 | Mid-Hi | 1327 | 0.4 |
| 5 | High | 1550 | 0.3 |

13. To graph your data, highlight cells B1:C5 (the Elevation and Average %H columns) and click "Insert" → "Chart"

14. Since the independent (x-axis) variable is continuous, we should plot this as an x-y scatter plot. Next to "Recommended charts", click "More>>" →"Scatter" and select the top scatterplot option. Click the "Customize" tab to add axis labels, a trendline, and make any other custom changes to your graph. When you're ready, click the blue "Insert" button and have your teacher come check your graph.

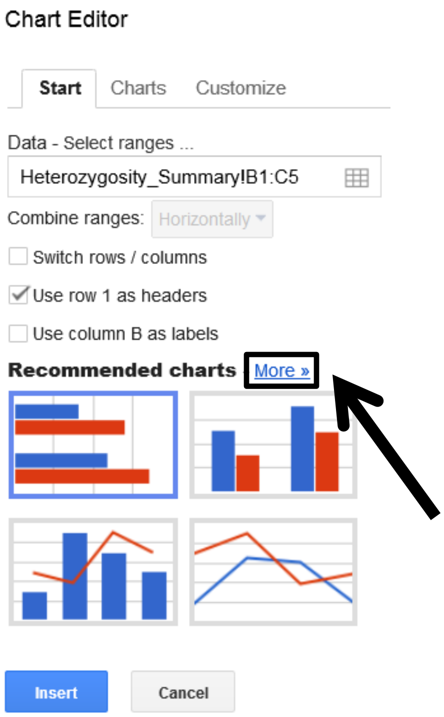


Chart Editor

Data - Select ranges ...
Heterozygosity_Summary!B1:C5
Combine ranges: Horizontally
☐ Switch rows / columns
☑ Use row 1 as headers
☐ Use column B as labels

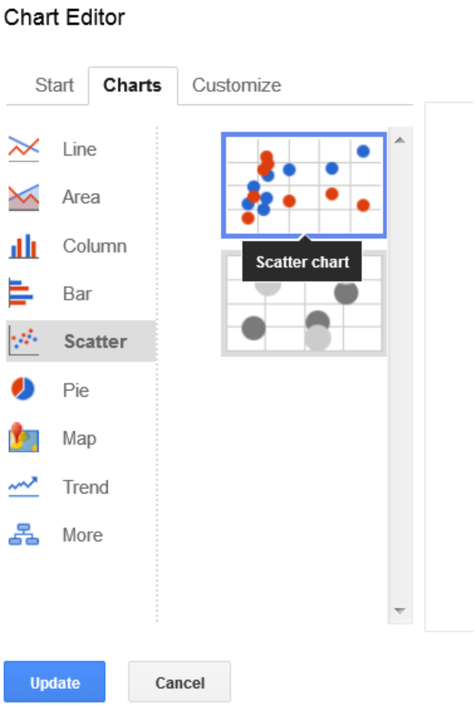


Chart Editor

15. What do you see? Revisit your prediction from # 1. Were you correct? Why/why not? What do you think is driving this pattern? Think about the ideal habitat and climate constraints of the pika, remembering that they don't do well in hot temperatures. (Record your answer on your worksheet.)

## Part 2: Identifying adaptive loci (SNPs)

With each replication and transcription event, DNA has the potential to mutate. Changes occur constantly in our genetic codes, though often our cellular machinery (e.g. DNA polymerase) catches these errors. The mutations that do slip through our cellular proofreaders are often neutral, meaning they don't affect the proteins that our DNA encodes. Sometimes these mutations do affect proteins, however, and they can either be beneficial or deleterious. It is up to natural selection to act on these mutations – if the trait is beneficial and increases an organism's chances of survival or reproduction, it is likely to become more common in the population over time. If the opposite is true, and the mutation leads to lower survival or reproduction, then the trait is likely to be selected against.

To figure out which genes are being acted on by natural selection, scientists often use SNPs. The SNPs that are being acted on by natural selection are often called **adaptive** loci, because they represent parts of the genetic code that are helping or hindering an organism's ability to adapt to its environment. Scientists identify these loci by looking at **allele frequencies**. If an allele for a SNP is in either very high or very low proportions in a location relative to another population, it is probably being selected for or against by natural selection or some other selective force. We have identified three potential SNP loci from the North Cascades pika populations for you to evaluate.

For this part of the lesson, you will use your Excel skills to address the question, "Are there signs of genetic adaptation to living on an elevational gradient in North Cascades National Park?" To do this, we have to figure out if any SNP alleles are favored at some elevations and not others by calculating their frequency in each population. Allele frequencies are just proportions: of all the possible alleles for a given SNP in a given population, what proportion of them is Allele 1? Allele 2? This can be represented by the equation:

$$Frequency\ of\ Allele\ 1 = \frac{\#\ of\ Allele\ 1}{Total\ \#\ of\ Alleles}$$

Given your Excel function knowledge, calculate the Frequency of Allele 1 and Allele 2 for each of the three SNPs in the Adaptive_Loci tab of the Google Spreadsheet. Once you have done this, answer the corresponding questions on your worksheet.

Some useful Excel functions:

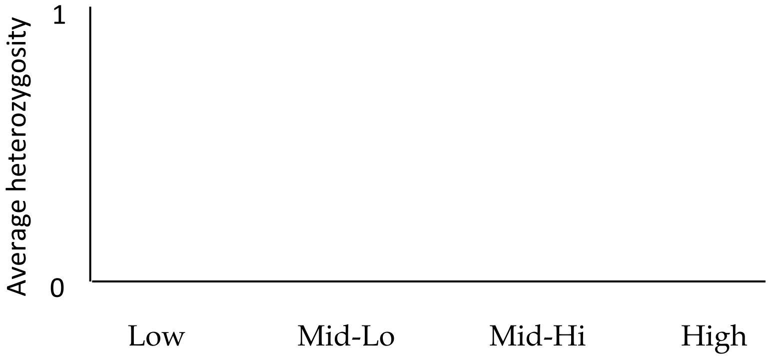
Count: "=COUNT()" counts all cells with values present

Countif: "=COUNTIF()" allows you to set conditions for counting. For example, "=COUNTIF(B2:C9,"2") will count all the cells between B2 and C9 that have a 2 in them.

# Student Worksheet

## Part 1: Inbreeding on an Elevational Gradient

1. Before you start your analysis, develop a prediction. Based on your class discussions and what you know about pikas, do you think pika populations will be healthier (have higher heterozygosity) at high, middle-high, middle-low, or low elevations? Sketch what you think an average heterozygosity bar graph would look like at these four elevations if your prediction is true.



Why do you predict this?

15. What do you see? Revisit your prediction from # 1. Were you correct? Why/why not? What do you think is driving this pattern? Think about the ideal habitat and climate constraints of the pika, remembering that they don't do well in hot temperatures.

**Part 2: Identifying adaptive loci (SNPs)**

1. What are the allele frequencies for each SNP in each population?

| | SNP1306 | | SNP3453 | | SNP3603 | |
|---|---|---|---|---|---|---|
| | Allele 1 | Allele 2 | Allele 1 | Allele 2 | Allele 1 | Allele 2 |
| Low | | | | | | |
| Mid-Lo | | | | | | |
| Mid-Hi | | | | | | |
| High | | | | | | |

2. Looking at these three SNP loci, which appears to be adaptive (undergoing selection)? Which allele for this SNP may represent a low-elevation adaptation, which a high-elevation adaptation? How can you tell?

3. There are many different possible functions for an elevation-adapted gene region like this. Do you have any guesses on what this gene might do? Use your creativity to guess the function of this gene.

4. How would you figure out if you're right about the function of this gene? Describe an experiment (or observational study) that you could use to figure out this function.

5. What about the other two SNPs you analyzed? Do they appear to be adaptive? Why/why not? Describe the allele patterns you see in these two loci and why you might see the pattern you do.

6. Based on the results of both Part 1 and Part 2 of this lesson, do you think low elevation pikas are adapted to their warmer, drier environment? Or are these populations going to disappear too?