**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**

© Matthew Waterhouse

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

+

1. 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. 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.
2. 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.
3. Repeat for the other three populations (you can use the corner drag trick you used in #7 again).
4. We now have average heterozygosity for our four populations. How does it change with elevation? Let’s graph it to visualize this pattern.

1. To graph this data, highlight and copy the four averages you just calculated.
2. Click on the Heterozygosity\_Summary” tab and right-click in the first cell below Average %H (cell C2). Select “Paste special” 🡪 “Paste values only.”

1. To graph your data, highlight cells B1:C5 (the Elevation and Average %H columns) and click “Insert” 🡪 “Chart”
2. 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.



1. 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.