



Calculating Probability of Identification in Northern Pacific Humpback Whales

Standards

- **HS.S-IC.1** Understand statistics as a process for making inferences about population parameters based on a random sample from that population.
- **MP.1** Make sense of problems and persevere in solving them.
- **MP.5** Use appropriate tools strategically.

Prior Knowledge

This lesson is an extension of what students did in the [CSI Wildlife](#) from HHMI “Frequency Primer.” In the Frequency Primer, students learned how DNA can be used to connect elephant tusks to poached elephants and how to determine the probability that this identification is accurate. This lesson leads students through calculating relative frequencies of alleles in known humpback whale populations and using those relative frequencies to calculate the probability of identification for individual whales. Whether this lesson is being used within this Save the Whales biology unit or as a part of a larger math unit on probability, the science background in the Frequency Primer and the CSI Wildlife from HHMI Case 1 preceding it is essential for contextual understanding.

Materials Needed

- [NP Humpback Whale Genotype Matches Data](#)
- Calculators

Objectives

- Students will be able to calculate the relative frequencies of key alleles in humpback whales populations from given genotype data.
- Students will be able to use these relative frequencies to calculate the probability of identification for individual whales.

Procedure

- As a class, examine the *NP Breeding Ground Reference* tab in the dataset [NP Humpback Whale Genotype Matches Data](#), ensuring that students know how to read the data in the spreadsheet.
 - a. For each allele, there are two columns--one showing the STR inherited from the mom and one showing the STR inherited from the dad.
 - b. The whales are grouped into sets of ten based on the sampling region.
- 2. Discuss the concept of relative frequency: the number of times a particular event occurs compared to the sample size. In this case, that would be how many times each STR appears in the data for all 30 whales.
$$\text{relative frequency} = \# \text{ of times STR appears} / 60$$
- 3. Divide up the data among small groups and have students calculate the relative frequency for each STR. Record the relative frequencies on a large chart or shared digital document.
- 4. STRs with a higher relative frequency are the more common STRs. Those with a lower relative frequency are less common and can help connect a whale to its breeding ground.
 - a. Identify the STRs with the lowest relative frequencies for each allele. Which are unique to specific populations?
 - b. Why would it be important to be able to connect an individual whale to its population? How does this relate to the problem of entanglement?
- 5. As researchers are tracking whale migration patterns, they use these genetic profiles to track the individual whales and connect them to their breeding grounds. The relative frequencies of STRs can be used to calculate the probability of identification for a sample whale. To calculate this probability, follow the same process used in the Frequency Primer: Calculate the probability of the individual whale being homozygous or heterozygous for each identified allele and then multiply those probabilities together.
 - a. If the data shows that the individual whale is homozygous for an allele, calculate the probability of that occurrence by squaring the relative frequency of the STR.

For example, Sample Whale 1 is homozygous for allele EV14, showing an STR of 131. To calculate this probability, square the relative frequency of STR 131.
 - b. If the data shows that the individual whale is heterozygous for an allele, calculate the probability of that occurrence by doubling the product of the relative frequencies of the STRs.

For example, Sample Whale 1 is heterozygous for allele EV37, showing STRs of 198 and 200. To calculate this probability, multiply $2(\text{relative frequency of 198})(\text{relative frequency of 200})$.

- c. After calculating the probability of being homozygous or heterozygous for each identified allele in the given whale, multiply each of these probabilities together. This is the probability of another whale having the same genetic profile as the one from which the sample was taken.
- d. Dividing 1 by this probability tells us that there is a 1 in _____ chance of another whale having this same genetic profile.
- e. How can these calculations be helpful for researchers? How does this relate to the problem of entanglement?