HANDS-ON ACTIVITY
Predicting Evolutionary Relationships

In this lab, you will design a process to analyze the molecular data of various organisms, and then classify the organisms according to the evolutionary relationship of the protein cytochrome C.

BACKGROUND

Organisms can be classified by physical traits. This is known as phylogenetics. However, this system has limitations, because many organisms may have similar physical appearances but be very diverse on a molecular level. For example, outside of the classroom you have seen different types of birds and insects and you may have seen a bat. You likely observed that all of these organisms had wings and could fly. However, upon closer examination, these organisms have many differences, not just in their physical appearance, but on a molecular level as well. Therefore, in addition to the comparison of physical characteristics, modern classification includes the analysis of molecular data as a basis for decisions about evolutionary relationships.

Scientists use empirical evidence, which is evidence obtained from observations and experiments, to classify organisms. The evidence may come from a detailed description of the physical characteristics of an organism, which may focus on the structural characteristics one species has in common with another. The description may be based on logical reasoning as well as observations made both in the field and in a laboratory setting. Today, as the result of technological advancements, evidence also includes DNA analysis, in which scientists compare the number of DNA sequences two species may have in common. These pieces of evidence, along with others, help scientists determine how closely related species are to each other.

The biochemical structures of organisms are used frequently to decipher species' relatedness. One approach of biochemical analysis is to determine the DNA sequence of genes, or the amino acid sequence of proteins. As you know, DNA encodes the information for making proteins. DNA is passed on from parents to offspring, so the DNA (and the proteins encoded) of the parent organisms and their offspring is highly similar. Over time, random mutations in DNA can occur, giving rise to a change in the amino acid sequence of the encoded proteins. These changes in the chemical structures of amino acids can be tracked and used to determine relatedness among species. The more similar the amino acid sequences of two species, the more closely related the species are likely to be.

Cytochrome C is a protein found in most eukaryotes. It is essential to energy production at the cellular level. The complete sequence of cytochrome C for three organisms is shown in the Amino Acid Sequence Alignment chart on the next page. Rather than the three-letter code to which you may be accustomed, each amino acid can be indicated by one letter. This convention is called, simply, a single-letter code.

AMINO ACID SEQUENCE ALIGNMENT OF CYTOCHROME C

POSITION	10	20	30	40	50	
HUMAN	--------GD VEKGKKIFIM KCSQCHTVEK GGKHKTGPNL HGLFGRKTGQ APGYS					
WHEAT	ASFSEAPPGN PDAGAKIFKT KCAQCHTVDA GAGHKQHGPN LGLFGRQSGT TAGYS					
YEAST	---TEFKAGS AKKGATLFKT RCELCHTVEK GGPHKVGPNL HGIFGRHSGQ AQGYS					
	60	70	80	90	100	110
HUMAN	YTAAN KNKGIIWGED TLMEYLE**NPK KYIPGTKMIF** VGIKKKEERA DLIAYLKKAT NE					
WHEAT	YSAAN KNKAVEWEEN TLYDYLL**NPK KYIPGTKMVF** PGLKKPQDRA DLIAYLKKAT SS					
YEAST	YTDAN IKKNVLWDEN NMSEYLT**NPK KYIPGTKMAF** GGLKKEKDRN DLITYLKKAC E-					

A cytochrome C molecule consists of a chain of 104 amino acids. The longest amino acid sequence that is common among the three organisms can be seen in bold-faced type. These identical sequences were aligned so that the remainder of the amino acids can be compared for differences. You will be an evolutionary detective as you research and analyze the cytochrome C sequence of several organisms and use the data to determine how closely the organisms are related.

PROCEDURE

1. Come up with a process plan. In your Evidence Notebook, write out the plan for your analysis. As you plan the process, use logical reasoning to make the following decisions:

    •   Select which organisms you will compare to the human sequence. You can select organisms from Table 1 below or use library or Internet resources to research cytochrome C amino acid sequence for other organisms not listed in Table 1.

    •   Form a hypothesis about which organisms are more closely related to humans and which organisms are less closely related to humans.

    •   Decide on a method to compare the amino acid sequences to determine how close the evolutionary relationships of the chosen organisms are to humans.

2. Have your teacher approve your plan.

3. Organize the results of your analysis into appropriate data tables in your Evidence Notebook. Be certain that the tables are constructed and labeled properly. Compare the different amino acid sequences that were analyzed. Table 2 provides information on the structure and properties of amino acids. This information could be helpful in your process of determining the types of changes that occurred in the amino acids. Analyze your data to identify patterns in the evolutionary relationships among the chosen organisms.

4. Share your results with other teams.

ANALYZE

1. Summarize your observations, including an analysis of data tables or graphs that you made.

   _____

   _____

   _____

   _____

   _____

2. Describe how the evolutionary relationships among species correlate with the molecular data you collected from the amino acid sequence analysis.

   _____

   _____

   _____

   _____

   _____

CONCLUDE

1. What conclusions can you draw from your results? What conclusions can you draw from your classmates' results?

   _____

   _____

   _____

   _____

   _____

2. Was your analytical process a good method for determining the evolutionary relationships among different organisms? Explain why or why not, and give examples of what might be missing from your analytical process.

   _____

   _____

   _____

   _____

   _____

TABLE 1: CYTOCHROME C AMINO ACID SEQUENCES

AMINO ACID NUMBER	HUMAN	CHICKEN	FROG	HORSE	MONKEY	RABBIT	SHARK	TURTLE
42	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN
43	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA
44	PRO	GLU	ALA	PRO	PRO	VAL	GLN	GLU
46	TYR	PHE	PHE	PHE	TYR	PHE	PHE	PHE
47	SER	SER	SER	THR	SER	SER	SER	SER
49	THR	THR	THR	THR	THR	THR	THR	THR
50	ALA	ASP	ASP	ASP	ALA	ASP	ASP	GLU
53	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS
54	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN
55	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS
56	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY
57	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE
58	ILE	THR	THR	THR	ILE	THR	THR	THR
60	GLY	GLY	GLY	LYS	GLY	GLY	GLN	GLY
61	GLU	GLU	GLU	GLU	GLU	GLU	GLN	GLU
62	ASP	ASP	ASP	GLU	ASP	ASP	GLU	GLU
63	THR	THR	THR	THR	THR	THR	THR	THR
64	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU
65	MET	MET	MET	MET	MET	MET	ARG	MET
66	GLU	GLU	GLU	GLU	GLU	GLU	ILE	GLU
100	LYS	ASP	SER	LYS	LYS	LYS	LYS	ASP
101	ALA	ALA	ALA	ALA	ALA	ALA	THR	ALA
102	THR	THR	CYS	THR	THR	THR	ALA	THR
103	ASN	SER	SER	ASN	ASN	ASN	ALA	SER
104	GLU	LYS	LYS	GLU	GLU	GLU	SER	LYS

TABLE 2: STRUCTURES AND PROPERTIES OF AMINO ACIDS

AMINO ACID	SINGLE-LETTER CODE	STRUCTURAL FORMULA	PROPERTIES
Alanine	A	H_3C—CH—COOH with NH$_2$ below CH	Neutral and hydrophobic (nonpolar)
Arginine	R	HN=, H_2N—C—NH—CH_2—CH_2—CH_2—CH—COOH with NH$_2$ below CH	Basic and hydrophilic (polar)
Asparagine	N	H_2N—C—CH_2—CH—COOH with O below C and NH$_2$ below CH	Neutral and polar
Aspartic acid	D	HO—C—CH_2—CH—COOH with O below C and NH$_2$ below CH	Acidic and nonpolar
Cysteine	C	HS—CH_2—CH—COOH with NH$_2$ below CH	Neutral and slightly polar
Glutamic acid	E	HO—C—CH_2—CH_2—CH—COOH with O below C and NH$_2$ below CH	Acidic and polar
Glutamine	Q	H_2N—C—CH_2—CH_2—CH—COOH with O below C and NH$_2$ below CH	Acidic and polar
Glycine	G	H_2C—COOH with NH$_2$ below	Neutral and amphiphilic (can exist in any surroundings)
Histidine	H	HC=C—CH_2—CH—COOH with N, NH, NH$_2$ below and CH below	Basic and polar
Isoleucine	I	H_3C above, H_3C—CH_2—CH—CH—COOH with NH$_2$ below	Neutral and nonpolar

TABLE 2: STRUCTURES AND PROPERTIES OF AMINO ACIDS, *CONTINUED*

Leucine	L		Neutral and nonpolar (very hydrophobic)
Lysine	K	$H_2N-CH_2-CH_2-CH_2-CH_2-CH-COOH$ with NH_2 below	Basic and polar (also causes kinks in protein chain)
Methionine	M	$H_3C-S-CH_2-CH_2-CH-COOH$ with NH_2 below	Neutral and nonpolar
Phenylalanine	F		Neutral and nonpolar
Proline	P		Neutral and nonpolar
Serine	S	$CH_2-CH-COOH$ with OH and NH_2 below	Neutral and polar
Threonine	T	$H_3C-CH-CH-COOH$ with OH and NH_2 below	Neutral and polar
Tryptophan	W		Neutral and slightly polar
Tyrosine	Y		Neutral and polar
Valine	V		Neutral and nonpolar