Species Identification: Penguins 7. It's Not All Black and White!

Name: ____

Objective

Using DNA sequence, distinguish species in the same genus from one another.

Background

In this activity, we will observe and analyze three sets of data, each of which is a DNA segment from a different species of penguin in the genus *Spheniscus*. The obvious

factor that sets these three species apart is their geographic distribution. With species in this genus spread across South America and across the Atlantic in Africa, we are interested in their genetic relationships since they are so far apart. It is also important that we understand the genetic differences to successfully breed these birds in captivity and to maintain a diverse and healthy gene pool.

Procedure

- 1. At each lab table, there should be four packets. Each student should take one packet. One person will have the master copy, and the other three students will each represent one of the penguin species.
- 2. First, look at the master copy and observe the left column. This is a 660 base segment of a gene sequence that is 1051 bases long, called ND2, and is a working gene. This is a segment of the circular DNA found in the mitochondrion. (There is also a nonfunctional copy of ND2, known as a pseudogene, located in the nuclear DNA.) They are labeled according to individual penguin, gene, and location of the gene. For example,

MALAGAMT/ND2: Malaga = Name of Individual African Penguin mt = Mitochondrial DNA ND2 = Gene That Has Been Isolated

- 3. Your first job will be to review the master copy together as a group and find all the differences between the three species of penguins. You will complete this task by looking at the individuals to the left and categorizing them by species. This information can be found on the cover of each data set packet. Write this information on your data sheet after you have found base differences by comparing each line of sequence.
- 4. After you know which species is which, go through the rows of bases, writing down each place where there is a difference. Do this on your data sheet. Be sure to mark which species has the substitution and the base location in the gene sequence. For example, at base 370, there is a difference that sets apart the Magellanic sample from the others.
- 5. You should now look on each species' data set and find the differences between the restriction enzymes from species to species. Keep in mind that we will need to cut each sequence in a place on the gene that will differ from species to species. Therefore, we will use an enzyme that will cut the gene in a place unique to that species. This is how we get different fragment lengths that can be seen in gel electrophoresis. Write the name of these restriction enzymes you would use on your data sheet.
- 6. Revisit the master copy and find the series of bases that will be cut by the different enzymes. Be sure to write these sequences on the data sheet. When you find their location, you may also write the number where the base sequence to be cut will begin. Do the same for the amino acid sequence found in the master copy. (See the data sheet.)



Analysis:

Below, write a step-by-step procedure for determining if two gene sequences actually represent two different species. Be sure to number your steps by placing numbers in the left margin

Data Sheet

Identify the species of penguin here:

Species:

Restriction Enzyme Differences:

Below, write the restriction enzymes that are unique to each species and that we can use for identification purposes.

S. demersus:	
S. humboldti:	
S manellanicus:	
0. mazeitunieus.	

Restriction Enzymes:

	Restriction Fragment Leng	ths	Sequence
Enzyme	Alu I	Rsa I	
S. demersus:			Alu I AG!CT
			Rsa I GT!AC
S. humboldti:			Alu I AG!CT
			Rsa I GT!AC
S. magellanicus:			Alu I AG!CT
			Rsa I GT!AC

Using the master copy's amino acid sequence and the amino acid chart at your table, note below the differences in the amino acid sequence by writing the species with the different amino acid, its location, and the amino acid name that is unique to that species at that location.

Species:

Location (No.):

Amino Acid: _____

SOLUTION

Data Sheet

Identify the species of penguin here:

Individual's Name/Number:	Species:		
POGO	African Penguin		
F445	Humboldt Penguin		
CHANCE	Magellanic Penguin		

Base Substitutions:

Below, write the location (number) of each base difference.

S. demersus:	116											
S. humboldti:	213	322	360	377	416	421	509	531	609	621	623	
S. magellanicus:	. 135	369	440	584	646							

Restriction Enzyme Differences:

Below, write the restriction enzymes that are unique to each species and that we can use for identification purposes.

S. demersus:	None of these cut in a unique place.					
S. humboldti:	Alu I					
S. magellanicus.	Rsa I					

Restriction Enzymes:

SOLUTION

	Restriction Fragment Leng	ths	Sequence
Enzyme	Alu I	Rsa I	
S. demersus:	240	135	Alu I AG!CT
	175	160	Rsa I GT!AC
	245	365	
S. humboldti:	240	135	Alu I AG!CT
	420	160	Rsa I GT!AC
		365	
S. magellanicus:	240	295	Alu I AG!CT
	175	365	Rsa I GT!AC
	245		

Using the master copy's amino acid sequence and the amino acid chart at your table, note below the differences in the amino acid sequence by writing the species with the different amino acid, its location, and the amino acid name that is unique to that species at that location.

Species	Location (No.)	Amino Acid
African	39	Serine
Humboldt	108	Threonine
Humboldt	126	Methionine
Humboldt	139	Valine
Humboldt	141	Threonine
Magellanic	147	Aspartic Acid
Humboldt	170	Serine
Magellanic	195	Methionine
Humboldt	208	Leucine
Magellanic	216	Alanine

Master Copy

Genus: Spheniscus



Gene: ND2

Gene Location: mtDNA

Data Sets: mtDNA, and Amino Acid Sequence

F445mt/ND2 CHANCEmt/ND2 POGOmt/ND2	10 1 ATGAACCCAC 1 ATGAACCCAC 1 ATGAACCCAC	20 ACACAAAATT ACACAAAATT ACACAAAATT	30 AATCTCCTAC AATCTCCTAC AATCTCCTAC	40 CTAAGCCTCC CTAAGCCTCC CTAAGCCTCC	50 TTCTAGGAAC TTCTAGGAAC TTCTAGGAAC	50 50 50
F445mt/ND2 CHANCEmt/ND2 POGOmt/ND2	60 51 AAATCAACAC 51 AAATCAACAC 51 AAATCAACAC	70 CCTCGCTATT CCTCGCTATT CCTCGCTATT	80 ATCCCCCTCA ATCCCCCTCA ATCCCCCTCA	90 ТСТСААААТС ТСТСААААТС ТСТСААААТС	100 CCACCACCCC CCACCACCCC CCACCACCCC	100 100 100
F445mt/ND2 CHANCEmt/ND2 POGOmt/ND2	110 101 CGAGCCATCG 101 CGAGCCATCG 101 CGAGCCATCG	120 AAGCAACAAT AAGCAACAAT AAGCAGCAAT	130 САААТАТТТС САААТАТТТС САААТАТТТС	140 CTAGTACAAG CTAGCACAAG CTAGTACAAG	150 CCACTGCCTC CCACTGCCTC CCACTGCCTC	150 150 150
F445mt/ND2 CHANCEmt/ND2 POGOmt/ND2	160 151 CACACTCATA 151 CACACTCATA 151 CACACTCATA	170 СТСТТСТСАА СТСТТСТСАА СТСТТСТСАА	180 GCATAACATA GCATAACATA GCATAACATA	190 ATGCCTGATT ÁTGCCTGATT ATGCCTGATT	200 CACAGGACAA CACAGGACAA CACAGGACAA	200 200 200
F445mt/ND2 CHANCEmt/ND2 POGOmt/ND2	210 201 TGAGACATCA 201 TGAGACATCA 201 TGAGACATCA	220 CCTAACTACC CCCAACTACC CCCAACTACC	230 GTCCTGCCTC GTCCTGCCTC GTCCTGCCTC	240 CTATTATCAG CTATTATCAG CTATTATCAG	250 CTGCAATTGC CTGCAATTGC CTGCAATTGC	250 250 250
F445mt/ND2 CHANCEmt/ND2 POGOmt/ND2	260 251 AATTAACTAT 251 AATTAACTAT 251 AATTAACTAT	270 GGACTGGTCC GGACTGGTCC GGACTGGTCC	280 CATTCCACTT CATTCCACTT CATTCCACTT	290 CTGATTCCCA CTGATTCCCA CTGATTCCCA	300 GAAGTACTTC GAAGTACTTC GAAGTACTTC	300 300 300
F445mt/ND2 CHANCEmt/ND2 POGOmt/ND2	310 301 AAGGTTCCCC 301 AAGGTTCCCC 301 AAGGTTCCCC	320 TATTAGCCAC TATTAGCCAC TATTAGCCAC	330 TACCCTTCTA TGCCCTTCTA TGCCCTTCTA	340 CTAGCAACAA CTAGCAACAA CTAGCAACAA	350 TAATGAAATT TAATGAAATT TAATGAAATT TAATGAAATT	350 350 350
F445mt/ND2 CHANCEmt/ND2 POGOmt/ND2	360 351 CCCCCCAATT 351 CCCCCCAATC 351 CCCCCCAATC	370 ACCATCCTCT ACCATCCTTT ACCATCCTCT	380 TCATAATATC TCATAACATC TCATAACATC TCATAACATC	390 CCACACACTT CCACACACTT CCACACACTT	400 AATCCCCTAC AATCCCCTAC AATCCCCTAC	400 400 400
F445mt/ND2 CHANCEmt/ND2 POGOmt/ND2	410 401 TACTAACCTC 401 TACTAACCTC 401 TACTAACCTC	420 CATAGTTATC CATAGCTATC CATAGCTATC	430 ACCTCAGCAG GCCTCAGCAG GCCTCAGCAG	440 CCCTAGGGGG CCCTAGGGGA CCCTAGGGGG	450 CTGAATAGGA CTGAATAGGA CTGAATAGGA	450 450 450
F445mt/ND2 CHANCEmt/ND2 POGOmt/ND2	460 451 CTAAACCAGA 451 CTAAACCAGA 451 CTAAACCAGA	470 CACAAATCCG CACAAATCCG CACAAATCCG	480 ААААААТСТТА ААААААТСТТА ААААААТСТТА	490 GCCTTCTCAT GCCTTCTCAT GCCTTCTCAT	500 CTATCTCCCA CTATCTCCCA CTATCTCCCA	500 500 500
F445mt/ND2 CHANCEmt/ND2 POGOmt/ND2	510 501 TCTAGGCTCA 501 TCTAGGCTGA 501 TCTAGGCTGA	520 ATAGCCATCA ATAGCCATCA ATAGCCATCA	530 TCATCATCTA TCATCATCTA TCATCATCTA	540 TAGCCCTAAA CAGCCCTAAA CAGCCCTAAA	550 CTCACCCTAC CTCACCCTAC CTCACCCTAC	550 550 550
F445mt/ND2 CHANCEmt/ND2 POGOmt/ND2	560 551 TAACCTTCTA 551 TAACCTTCTA 551 TAACCTTCTA	570 CCTATACTCC CCTATACTCC CCTATACTCC	580 СТААТААССА СТААТААССА СТААТААССА	590 TCACAACATT TCATAACATT TCACAACATT	600 TCTCACCCTC TCTCACCCTC TCTCACCCTC	600 600 600
F445mt/ND2 CHANCEmt/ND2 POGOmt/ND2	610 601 AATACAACCA 601 AATACAACTA 601 AATACAACTA	620 АААСССТААА АААСССТААА АААСССТААА	630 ACTATCCACT GCCATCCACT GCCATCCACT	640 ATAATAACCT ATAATAACCT ATAATAACCT	650 CATGAACAAA CATGAGCAAA CATGAACAAA	650 650 650

Penguin Data Set 1

Species: African Penguin

A.K.A.: Black-Footed Penguin

Spheniscus demersus



Individuals in Data Set:

Pogo

Breeding Range: Namibia and South Africa

Length: 68 cm

World Population: 70,000 Breeding Pairs

POGOmt/ND2

ATGAACCCAC ACACAAAATT AATCTCCTAC CTAAGCCTCC TTCTAGGAAC AAATCAACAC CCTCGCTATT ATCCCCCTCA TCTCAAAATC CCACCACCCC CGAGCCATCG AAGCAGCAAT CAAATATTTC CTAGTACAAG CCACTGCCTC CACACTCATA CTCTTCTCAA GCATAACATA ATGCCTGATT CACAGGACAA TGAGACATCA CCCAACTACC GTCCTGCCTC CTATTATCAG CTGCAATTGC AATTAACTAT GGACTGGTCC CATTCCACTT CTGATTCCCA GAAGTACTTC AAGGTTCCCC TATTAGCCAC TGCCCTTCTA CTAGCAACAA TAATGAAATT CCCCCCAATC ACCATCCTCT TCATAACATC CCACACACTT AATCCCCTAC TACTAACCTC CATAGCTATC GCCTCAGCAG CCCTAGGGGGG CTGAATAGGA CTAAACCAGA CACAAATCCG AAAAATCTTA GCCTTCTCAT CTATCTCCCA TCTAGGCTGA ATAGCCATCA TCATCATCTA CAGCCCTAAA CTCACCCTAC TAACCTTCTA CCTATACTCC CTAATAACCA TCACAACATT TCTCACCCTC ΑΑΤΑCAACTA ΑΑΑCCCTAAA GCCATCCACT ΑΤΑΑΤΑΑCCT CATGAACAAA ΑΑΑCAACTCT

Penguin Data Set 2

Species: Humboldt Penguin

Spheniscus humboldti





Individuals in Data Set:

F445

Breeding Range: Northern Chile and Peru

Length: 70 cm

World Population: 12,000 Breeding Pairs

F445mt/ND2 ATGAACCCAC ACACAAAATT AATCTCCTAC CTAAGCCTCC TTCTAGGAAC AAATCAACAC CCTCGCTATT ATCCCCCTCA TCTCAAAATC CCACCACCCC CGAGCCATCG AAGCAACAAT CAAATATTTC CTAGTACAAG CCACTGCCTC CACACTCATA CTCTTCTCAA GCATAACATA ATGCCTGATT CACAGGACAA TGAGACATCA CCTAACTACC GTCCTGCCTC CTATTATCAG CTGCAATTGC AATTAACTAT GGACTGGTCC CATTCCACTT CTGATTCCCA GAAGTACTTC AAGGTTCCCC TATTAGCCAC TACCCTTCTA CTAGCAACAA TAATGAAATT CCCCCCAATT ACCATCCTCT TCATAATATC CCACACACTT AATCCCCTAC TACTAACCTC CATAGTTATC ACCTCAGCAG CCCTAGGGGG CTGAATAGGA CTAAACCAGA CACAAATCCG AAAAATCTTA GCCTTCTCAT CTATCTCCCA TCTAGGCTCA ATAGCCATCA TCATCATCTA TAGCCCTAAA CTCACCCTAC TAACCTTCTA CCTATACTCC CTAATAACCA TCACAACATT TCTCACCCTC AATACAACCA AAACCCTAAA ACTATCCACT ATAATAACCT CATGAACAAA AAACAACTCT

Penguin Data Set 3

Species: Magellanic Penguin

Spheniscus magellanicus





Individuals in Data Set: Chance

Breeding Range: Chile, Argentina, and the Falkland Islands

Length: 70 cm

World Population: 1,800,000 Breeding Pairs

CHANCEmt/ND2

ATGAACCCAC ACACAAAATT AATCTCCTAC CTAAGCCTCC TTCTAGGAAC AAATCAACAC CCTCGCTATT ATCCCCCTCA TCTCAAAATC CCACCACCCC CGAGCCATCG AAGCAACAAT CAAATATTTC CTAGCACAAG CCACTGCCTC CACACTCATA CTCTTCTCAA GCATAACATA ATGCCTGATT CACAGGACAA TGAGACATCA CCCAACTACC GTCCTGCCTC CTATTATCAG CTGCAATTGC AATTAACTAT GGACTGGTCC CATTCCACTT CTGATTCCCA GAAGTACTTC AAGGTTCCCC TATTAGCCAC TGCCCTTCTA CTAGCAACAA TAATGAAATT CCCCCCAATC ACCATCCTTT TCATAACATC CCACACACTT AATCCCCTAC TACTAACCTC CATAGCTATC GCCTCAGCAG CCCTAGGGGA CTGAATAGGA CTAAACCAGA CACAAATCCG AAAAATCTTA GCCTTCTCAT CTATCTCCCA TCTAGGCTGA ATAGCCATCA TCATCATCTA CAGCCCTAAA CTCACCCTAC TAACCTTCTA CCTATACTCC CTAATAACCA TCATAACATT TCTCACCCTC AATACAACTA AAACCCTAAA GCCATCCACT ATAATAACCT CATGAGCAAA AAACAACTCT

Amino Acid Sequence Comparison (Optional)

F445mt/ND2/A CHANCEmt/ND2 POGOmt/ND2/A	1 1 1	10 WNPHTKLISY MNPHTKLISY MNPHTKLISY	20 LSLLLGTNQH LSLLLGTNQH LSLLLGTNQH	30 PRYYPPHLKI PRYYPPHLKI PRYYPPHLKI	40 PPPPSHRSNN PPPPSHRSNN PPPPSHRSSN	50 QMFPSTSHCL QMFPSTSHCL QMFPSTSHCL	50 50 50
F445mt/ND2/A CHANCEmt/ND2 POGOmt/ND2/A	51 51 51	60 HTHTLLKHNM HTHTLLKHNM HTHTLLKHNM	70 MPDSQDNETS MPDSQDNETS MPDSQDNETS	80 PHYRPASYYQ PNYRPASYYQ PNYRPASYYQ	90 LQLQLTHDMS LQLQLTHDMS LQLQLTHDMS	100 HSTSDSQKYF HSTSDSQKYF HSTSDSQKYF	100 100 100
F445mt/ND2/A CHANCEmt/ND2 POGOmt/ND2/A	101 101 101	110 KVPLLATTLL KVPLLATALL KVPLLATALL	120 LATNNKFPPI LATNNKFPPI LATNNKFPPI	130 TILFWMSHTL TILFWTSHTL TILFWTSHTL	140 NPLILTSMVI NPLLLTSMAI NPLLLTSMAI	150 TSAALGGNMG ASAALGDMMG ASAALGGMMG	150 150 150
F445mt/ND2/A CHANCEmt/ND2 POGOmt/ND2/A	151 151 151	160 LNQTQIRKIL LNQTQIRKIL LNQTQIRKIL	170 AFSSISHLGS AFSSISHLGW AFSSISHLGW	180 MAIIIIYSPK MAIIIIYSPK MAIIIIYSPK	190 LTLLTFYLYS LTLLTFYLYS LTLLTFYLYS	200 LWTITTFLTL LWTIMTFLTL LWTITTFLTL	200 200 200
F445mt/ND2/A CHANCEmt/ND2 POGOmt/ND2/A	201 201 201	210 NTTKTLKLST NTTKTLKPST NTTKTLKPST	220 MNTSØTKNNS MNTSØTKNNS NNTSØTKNNS	230	240	250	250 250 250

Amino Acid Sequences for ND2 Gene

POGOmt/ND	2/44					
10 MNPHTKLISY	20 LSLLLGTNQH	30 PRYYPPHLKI	40 PPPPSHRSSN	50 QMFPSTSHCL	60 HTHTLLKHNM	
70 MPDSQDNETS	80 PNYRPASYYQ	90 LQLQLTMDWS	100 Hstsdsqkyf	110 KVPLLATALL	120 LATMMKFPPI	
130 TILFMTSHTL	140 NPLLLTSMAI	150 ASAALGGWMG	160 LNQTQIRKIL	170 AFSSISHLGW	180 MAIIIIYSPK	
190 LTLLTFYLYS	200 LMTITTFLTL	210 NTTKTLKPST	220 MMTSWTKNNS	230	240	
CHANCEmt/	ND2/AA					
10 MNPHTKLISY	20 LSLLLGTNQH	30 PRYYPPHLKI	40 PPPPSHRSNN	50 QMFPSTSHCL	60 HTHTLLKHNM	

(Continued on page 88.)

Amino Acid Sequences For ND2 Gene (Continued)

F445mt/ND2/AA								
10	20	30	40	50	60			
MNPHTKLISY	LSLLLGTNQH	PRYYPPHLKI	PPPPSHRSNN	QMFPSTSHCL	HTHTLLKHNM			
70	80	90	100	110	120			
MPDSQDNETS	PNYRPASYYQ	LQLQLTMDWS	HSTSDSQKYF	KVPLLATTLL	LATMMKFPPI			
130	140	150	160	170	180			
TILFMMSHTL	NPLLLTSMVI	TSAALGGWMG	LNQTQIRKIL	AFSSISHLGS	MAIIIIYSPK			
190 LTLLTFYLYS	200 LMTITTFLTL	210 NTTKTLKLST	220 MMTSWTKNNS	230	240			

Single-Letter Amino Acid Symbols

Amino Acid	Three-Letter Symbol	One-Letter Symbol
Alanine	Ala	А
Arginine	Arg	R
Asparagine	Asn	Ν
Aspartic Acid	Asp	D
Cysteine	Cys	С
Glutamine	Gln	Q
Glutamic Acid	Glu	Е
Glycine	Gly	G
Histidine	His	Н
Isoleucine	Ile	Ι
Leucine	Leu	L
Lysine	Lys	К
Methionine	Met	М
Phenylalanine	Phe	F
Proline	Pro	Р
Serine	Ser	S
Threonine	Thr	Т
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V
***	End	*
Xxx	Unidentified	X