STOMACH SOUP: USING GENETICS TO IDENTIFY THE ORGANISMS CONSUMED BY WHITE MARLIN (*Kajikia albida*)

Quinn Girasek
Virginia Institute of Marine Science

Grade Level
High School

Subject Area
Biology

VA SEA is a collaborative project between the Chesapeake Bay National Estuarine Research Reserve, the Virginia Institute of Marine Science's Marine Advisory Program, and Virginia Sea Grant. The VA SEA project is made possible through funding from the National Science Foundation and William & Mary's Society of 1918 Endowment.

Title:

Stomach Soup: Using Genetics to Identify the Organisms Consumed by White Marlin (*Kajikia albida*)

Focus:

Students will use randomly selected genetic sequences to identify digested fish and invertebrates. Students will also learn how to use NCBI GenBank and BLAST, which are part of the National Library of Medicine and used in genetics research.

Grade Level:

High School Biology 9-10

Virginia Standards of Learning:

BIO.5 The student will investigate and understand that there are common mechanisms for inheritance. Key ideas include

a) DNA has structure and is the foundation for protein synthesis

BIO.7 The student will investigate and understand that populations change through time. Key ideas include

b) genetic variation, reproductive strategies, and environmental pressures affect the survival of populations

c) natural selection is a mechanism that leads to adaptations and may lead to the emergence of new species

Learning Objectives:

- Students will simulate the process by which we extract DNA from stomach contents
- Students will identify species using NCBI BLAST
- Students will assess the diversity of organisms consumed by top oceanic predators
- Students will create and infer phylogenetic trees (if time allows)
- Students will transcribe and translate sequences (for advanced students)

Total length of time required for the lesson:

30-40 minutes

Vocabulary:

- DNA: Deoxyribonucleic acid
- Transcription: Changing DNA to RNA
- Translation: Changing RNA to Proteins
- Amino Acid: Organic compounds that form proteins
- NCBI: National Center for Biotechnology Information
- NCBI GenBank: “NIH genetic sequence database, an annotated collection of all publicly available DNA sequence”
- NCBI BLAST: Basic Local Alignment Search Tool that is part of NCBI
- Bioinformatics: “A scientific field that uses computers and other tools to analyze biological data, such as DNA and amino acid sequences”

Background Information:

Genetics is a useful tool for many different fields, from studying molecular structures to managing fish species in fisheries. In genetic research, scientists often use resources like NCBI GenBank and BLAST to analyze and understand the DNA they have sequenced.

Genetics can be used in studying trophic ecology, which helps scientists learn about a species’ diet. Knowing what a species eats is important for understanding its biology and how it interacts with other species in the ecosystem. Food provides energy for essential functions like growth, bone health, and overall well-being. By studying diet, we can learn more about a species' role in the ecosystem, such as its ecological niche and how its population changes over time. We can also identify which areas of the environment need protection or restoration, since a species' diet can reveal important details about its habitat and how it moves through its environment to find food. Predators are also more likely to accumulate harmful chemicals through their diet, which can affect humans too. From understanding what predators eat, we can learn more about these processes, such as bioaccumulation and biomagnification. Additionally, species can influence each other through competition or predator-prey relationships.

There are three main ways to study a species' diet: morphology, genetics, and stable isotopes. This lesson focuses on using genetics to analyze diet. The first step is to choose a species and collect stomach samples. To make sure the results are meaningful, researchers perform a power analysis to figure out the minimum number of samples needed. A power analysis helps determine the minimum sample size for statistically significant results. After collecting the stomachs, the contents are brought to a lab, where they are mixed into a single liquid solution. Next, DNA is extracted from the solution using a special DNA extraction kit. Then, a process called polymerase chain reaction (PCR) is used to make many copies of specific DNA segments. Using these DNA copies, a genomic library is created, and DNA sequencing is done. The final DNA sequences are cleaned up and analyzed using tools like NCBI BLAST and R. This lesson teaches students about the importance of diet studies and how to use NCBI BLAST, a tool scientists use for their research.

Materials & Supplies:

- Paper bag
    - 1 for each group, groups can be up to 3-4 students
    - For a class size of 24: 6 or 8
- Fish Cutouts
    - There are 9 fish, and each bag will have 3 fish
    - For six bags: 18 fish
    - For eight bags: 24 fish
    - Note: these are all fish skeletons, but some are invertebrates. They all look the same alike because with digested organisms, it is hard to identify to the species level without genetics! The eyes are colored and align with the answer key.
- Student handouts
    - There are two handouts, one for everyone and the second is for advanced students or students who find the work interesting and are highly motivated but work slower but than their peers
    - The second handout may be given out as homework or an optional assignment during class

Teacher Preparation:

- Color print and cut out paper
    - Dashed lines will indicate other places to cut to make the fish like a puzzle
    - Laminate paper after cutting the pieces so it can easily be reused
- Place paper fish into brown paper bags and fish are color coded by eyes
    - Fish with red, orange, and yellow eyes should be placed in one bag
    - Fish with light green, dark green, and dark blue eyes should be placed in one bag
    - Fish with pink, purple, and light blue eyes should be placed in one bag
    - Repeat so that there are enough bags for each group and each group has three fish of different eye colors in their bag
    - You can mix the fish however you would like
- Print out the handout

Procedure:

Background + Activity Introduction (around 15 min)

- Use PowerPoint slides. There are speaker notes that expand on the slides

Activity (around 20 min)

- Give each group a paper bag with the cut-up fish bones
- Students will take out the fish and match up the pieces to complete the fish
    - This is analogous to extracting and sequencing the DNA
- Students will use NCBI BLAST to answer questions on the worksheet
    - BLAST has a nucleotide and a protein part

  - Students will use the nucleotide BLAST
  - Refer to slides 15-19 for a tutorial
  - For the discussion, an extra explore slide focuses on creating and interpreting phylogenetic trees using the https://phylot.biobyte.de website (slide 22)
- Advanced and/or highly motivated students may be given worksheet number 2, which may be done in class or outside of class
  - These will require slides 23-24 on the PowerPoint

Post activity (around 5 min)

- Student groups can compare results and talk about what fish their groups found
  - Refer to slide 21-22 for more discussion topics and questions

Assessment:

See attached handouts. Participation in groups and in the discussions should also be considered.

References:

N/A

Handouts/Worksheets:

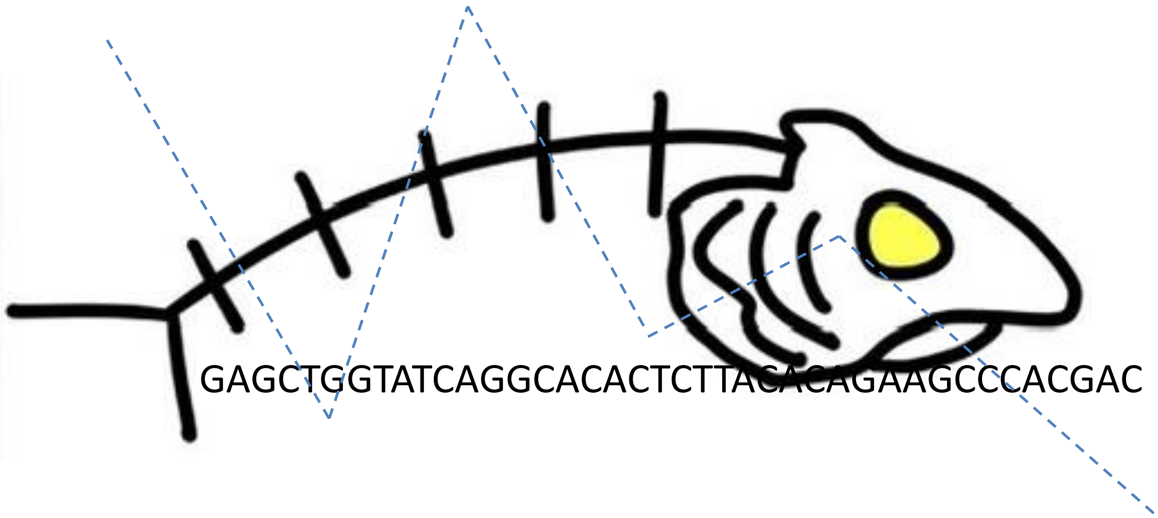
1. Fish to cut out (pages 5-7)
2. Worksheet 1. Student Worksheet (page 8)
3. Worksheet 2. Advanced Student Worksheet Supplement (pages 9)
4. Worksheet 1 answer key (pages 10-13)
5. Worksheet 2 answer key (pages 14-25)

NOTE: For more advanced work, students may transcribe and translate the DNA to RNA and Proteins, in addition to the base worksheet. Then they may use Protein BLAST and learn about the proteins that come up. The steps for using the Protein Blast are the same as Nucleotide BLAST but students must input the protein sequences they determined. Refer to slide 21 for the one letter symbol of proteins.

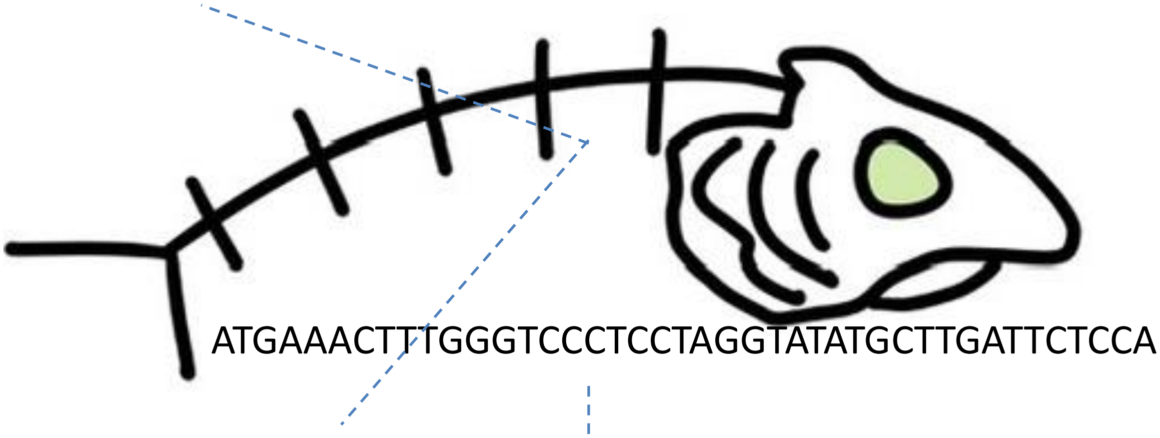
Handout 1. Fish to Cut Out

CATAGTGCATATATGTATTTACACCATACATATATATTAAACAT

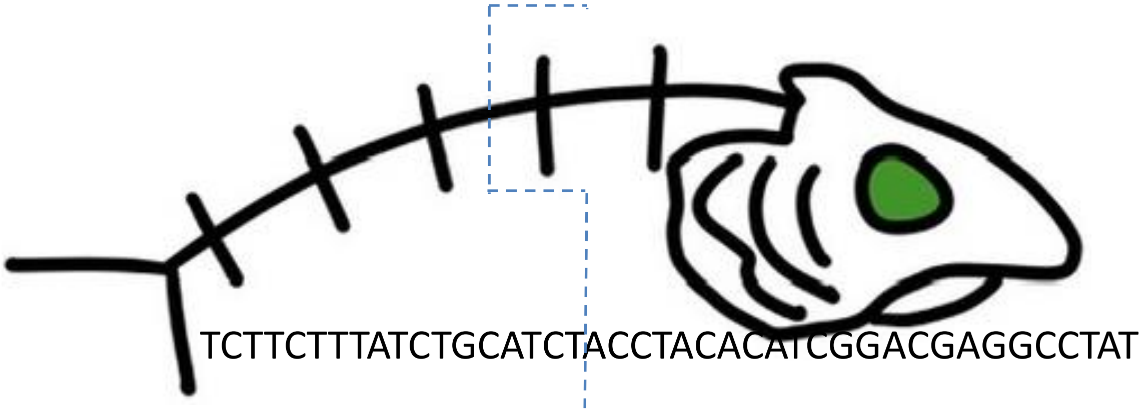
TAATTCTCTTAACACTTGGCCTGATTTACGAATGACTACAAG



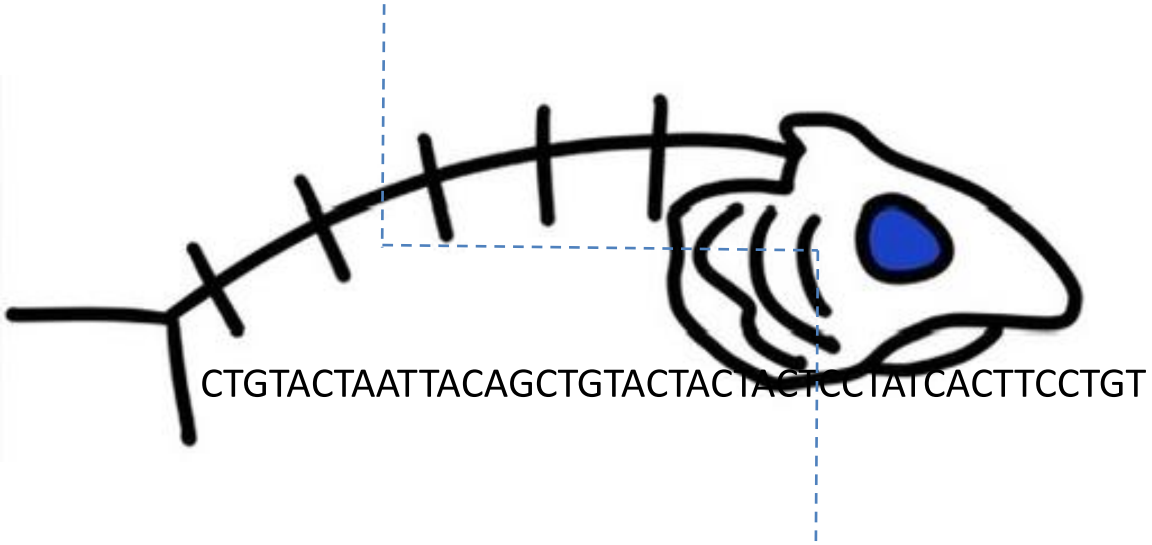
GAGCTGGTATCAGGCACACTCTTACACAGAAGCCCACGAC



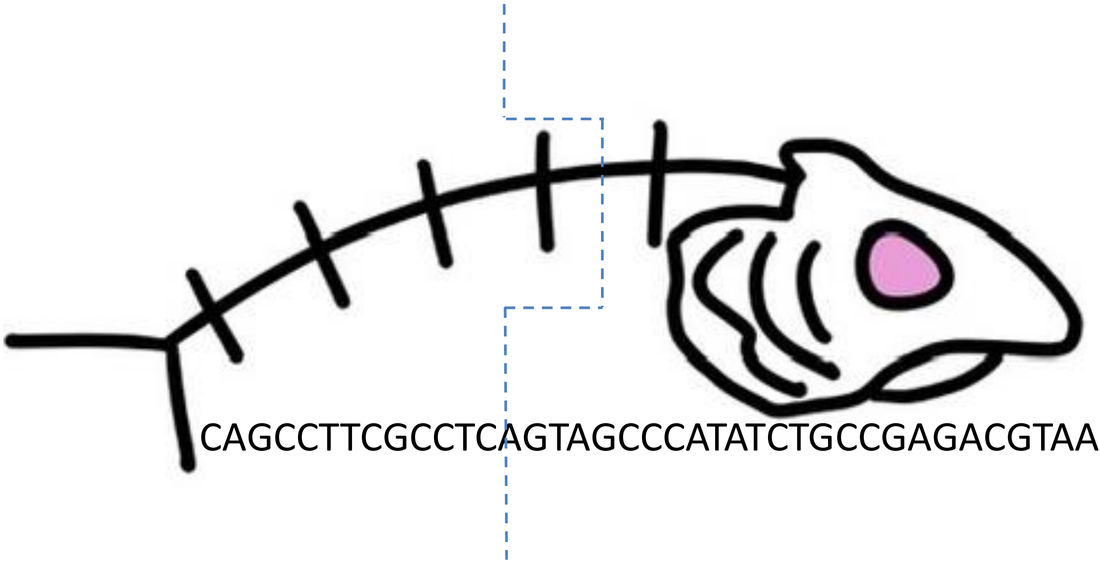
ATGAAACTTTGGGTCCCTCCTAGGTATATGCTTGATTCTCCA



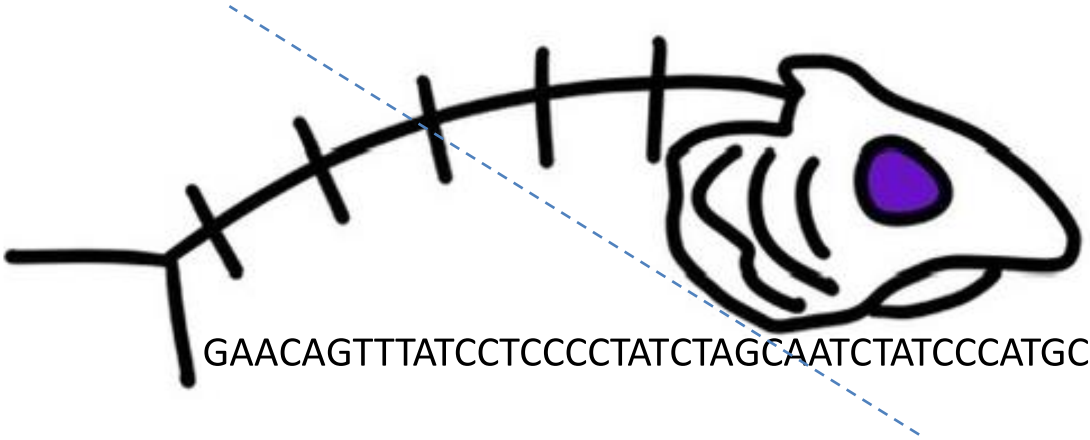
TCTTCTTTATCTGCATCTACCTACACATCGGACGAGGCCTAT



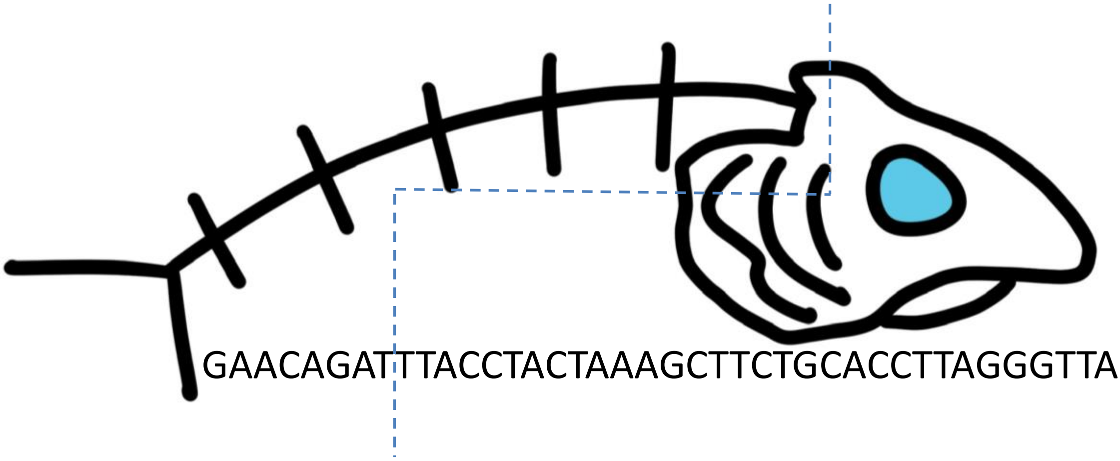
CTGTACTAATTACAGCTGTACTACTACTCCTATCACTTCCTGT



CAGCCTTCGCCTCAGTAGCCCATATCTGCCGAGACGTAAAC



GAACAGTTTATCCTCCCCTATCTAGCAATCTATCCCATGCAGG



GAACAGATTTACCTACTAAAGCTTCTGCACCTTAGGGTTACTTTAA

Worksheet 1. Student Worksheet

Name Assessment Class period

Date Eye color of organism

1. What is the DNA sequence you picked?

NCBI BLAST website: https://blast.ncbi.nlm.nih.gov/Blast.cgi

Go to NCBI GenBank, nucleotide BLAST. Search the sequence and answer the following questions:

2. What is the first scientific name that comes up? They are in the second column.

3. Are the first three scientific names the same species?

4. Keep scrolling down the page. Are there other species listed that are different from the top result? If not, list the English common name of your species. If yes, list the other scientific names and determine the English common name for those species, in addition to your species.

5. If there are other species that are listed in the results, why do you think the results show other species besides the correct species identification?

6. In what ways can BLAST be used to investigate evolutionary relationships?

Worksheet 2. Advanced Student Worksheet Supplement – use the DNA sequence from the organism in Worksheet 1

Name Assessment Class period

Date Eye color of organism

1. What is the transcribed sequence? (DNA to RNA)

2. What is the translated sequence? (RNA to protein)

Go to NCBI GenBank, protein BLAST. Search the protein sequence and answer the following questions:

3. What are the first three proteins that come up?

4. Pick one of them. Click on one of the description rows with the name of the protein (first column). What is the Query sequence and the Subject of the protein?

5. For the protein you picked, what is the main function of this gene? Tip: use www.uniprot.org or https://www.ebi.ac.uk/interpro/entry/InterPro/#table to learn more about the protein. Google scholar is also a good source to find information about proteins.

6. In these searches where you are learning more about the proteins, there are new vocabulary words. List three of them and provide the definition.

Answer Keys: There are two answer keys, one for the regular worksheet and one for the supplemental worksheet that builds upon the first worksheet and is for more advanced students. The second one can also be given as homework or optional material for students who express more interest in the topics.

Answer Key: Worksheet 1. Student Worksheet

1. What is the DNA sequence you picked?

Red species: CATAGTGCATATATGTATTTACACCATACATATATATTAAACAT

Orange species: TAATTCTCTTAACACTTGGCCTGATTTACGAATGACTACAAG

Yellow species: GAGCTGGTATCAGGCACACTCTTACACAGAAGCCCACGAC

Light Green species: ATGAAACTTTGGGTCCCTCCTAGGTATATGCTTGATTCTCCA

Dark Green species: TCTTCTTTATCTGCATCTACCTACACATCGGACGAGGCCTAT

Dark Blue species: CTGTACTAATTACAGCTGTACTACTACTCCTATCACTTCCTGT

Pink species: CAGCCTTCGCCTCAGTAGCCCATATCTGCCGAGACGTAAAC

Purple species: GAACAGTTTATCCTCCCCTATCTAGCAATCTATCCCATGCAGG

Light Blue species: GAACAGATTTACCTACTAAAGCTTCTGCACCTTAGGGTTACTTTAA

NCBI BLAST website: **https://blast.ncbi.nlm.nih.gov/Blast.cgi**

Go to NCBI GenBank, nucleotide BLAST. Search the sequence and answer the following questions:

2. What is the first scientific name that comes up? They are in the second column.

Red species: Frigate Tuna (*Auxis thazard*)

Orange species: Little Tunny (*Euthynnus alletteratus*)

Yellow species: Atlantic Chub Mackerel (*Scomber colias*)

Light Green species: Planehead Filefish (*Stephanolepis hispidus*)

Dark Green species: Bluefish (Pomatomus saltator)

Dark Blue species: mahi mahi, dolphinfish, dorado – they have three common names! (*Coryphaena hippurus*)

Pink species: Skipjack tuna (*Katsuwonus pelamis*)

Purple species: Atlantic bird squid (*Ornithoteuthis antillarum*)

Light Blue species: Neon flying squid (*Ommastrephes bartramii*)

3. Are the first three scientific names the same species?

Red species: Yes

Orange species: Yes

Yellow species: Yes

Light Green species: Yes

Dark Green species: Yes

Dark Blue species: Yes

Pink species: Yes

Purple species: Yes

Light Blue species: Yes

4. Keep scrolling down the page. Are there other species listed that are different from the top result? If not, list the English common name of your species. If yes, list the other scientific names and determine the English common name for those species, in addition to your species.

Red species: Yes. *Auxis thazard* (Frigate tuna) is the species and *Euthynnus affinis* (Mackerel tuna) is the other one that comes up.

Orange species: No. *Euthynnus alletteratus* (Little tunny).

Yellow species: Yes. *Scomber colias* (Atlantic chub mackerel) is the species and *Scomber japonicus* (Pacific chub mackerel) is the other one that comes up.

Light Green species: No. *Stephanolepis hispidus* (Planehead Filefish).

Dark Green species: Yes: *Pomatomus saltator* (Bluefish) is the species and the others that come up are *Steindachnerina brevipinna* (Characin) and Xenopus longipes (Lake Oku clawed frog) and *Xenopus eysoole* (Bamiléké clawed frog). There are others that are a 95% instead of 100% for query cover so I am not including them in the answer key.

Dark Blue species: No. *Coryphaena hippurus* (Mahi mahi, dolphinfish, dorado – they have three common names!).

Pink species: No. *Katsuwonus pelamis* (Skipjack tuna) is the species and the others that come up are *Auxis thazard* (Frigate Tuna) and *Euthynnus affinis* (Mackerel tuna).

Purple species: No. *Ornithoteuthis antillarum*) (Atlantic bird squid) is the species and *Hyaloteuthis pelagica* (Glassy flying squid or glass squid) is the other one that comes up. There are others that are a 95% instead of 100% for query cover so I am not including them in the answer key.

Light Blue species: Yes. *Ommastrephes bartramii* (Neon flying squid) is the species and the others that come up are *Ommastrephes caroli* (Webbed flying squid) and *Lolliguncula brevis* (Atlantic brief squid). There are others that are a £97.83% instead of 100% for query cover so I am not including them in the answer key.

5. If there are other species that are listed in the results, why do you think the results show other species besides the correct species identification?

The other species listed are generally closely related species so it would make sense for them to also pop up in the search results. Closely related species share a lot of the genetic code together so the probability of finding only one species and not a close relative is low, depending on what part of the gene is used to identify the species. This depends on if the region is conserved or if it is a highly variable part of the genome.

6. In what ways can BLAST be used to investigate evolutionary relationships?

BLAST has a function to create a distance tree of all results, which can be used to assess how similar the BLAST results are. There are also other websites (e.g., https://phylot.biobyte.de) and programs (e.g., MacVector and MEGA - Molecular Evolutionary Genetics Analysis) where you can import FASTA files and create phylogenetic trees to assess evolutionary relationships.

Answer Key: Worksheet 2. Advanced Student Worksheet Supplement

1. What is the transcribed sequence? (DNA to RNA)

Red species: CAUAGUGCAUAUAUGUAUUUACACCAUACAUAUAUAUUAAACAU

Orange species: UAAUUCUCUUAACACUUGGCCUGAUUUACGAAUGACUACAAG

Yellow species: GAGCUGGUAUCAGGCACACUCUUACACAGAAGCCCACGAC

Light Green species: AUGAAACUUUGGGUCCCUCCUAGGUAUAUGCUUGAUUCUCCA

Dark Green species: UCUUCUUUAUCUGCAUCUACCUACACAUCGGACGAGGCCUAU

Dark Blue species: CUGUACUAAUUACAGCUGUACUACUACUCCUAUCACUUCCUGU

Pink species: CAGCCUUCGCCUCAGUAGCCCAUAUCUGCCGAGACGUAAAC

Purple species: GAACAGUUUAUCCUCCCCUAUCUAGCAAUCUAUCCCAUGCAGG

Light Blue species: CGUAAUUUGUUUAGAGAGUUCUUAUUGAAAAACGAGAUUGCGACC

2. What is the translated sequence? (RNA to protein) *Note that some of the protein sequences have periods. That is because the RNA sequences do not translate to proteins nicely. Most of the species do not have them, and if/when students are able to do the others more easily, encourage them to try the more challenging ones. Some of the proteins that come up are hypothetical proteins or do not have a lot of research done on them. I kept them in because it is important for students to understand that there are gaps and limitations to scientific databases.*

Red species: HSAYMYLHHTYILN

Orange species: .FS.HLA.FTNDYK

Yellow species: ELVSGTLLHRSPR

Light Green species: MKLWVPPRYMLDSP

Dark Green species: SSLSASTYTSDEAY

Dark Blue species: LY.LQLYYYSYHFL

Pink species: QPSPQ.PISAET.

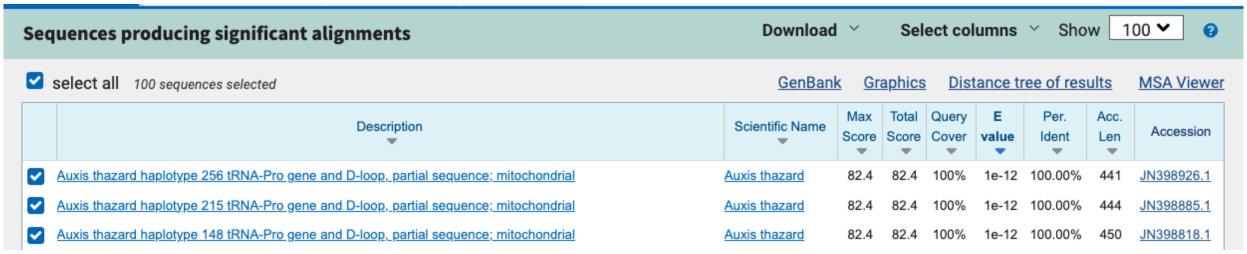
Purple species: EQFILPYLAIYPMQ

Light Blue species: RNLFREFLLKNEIAT

Go to NCBI GenBank, <u>protein</u> BLAST. Search the protein sequence and answer the following questions:

3. What are the first three proteins that come up? *These are screenshots of the results when I inputted the sequences.*

Red species:

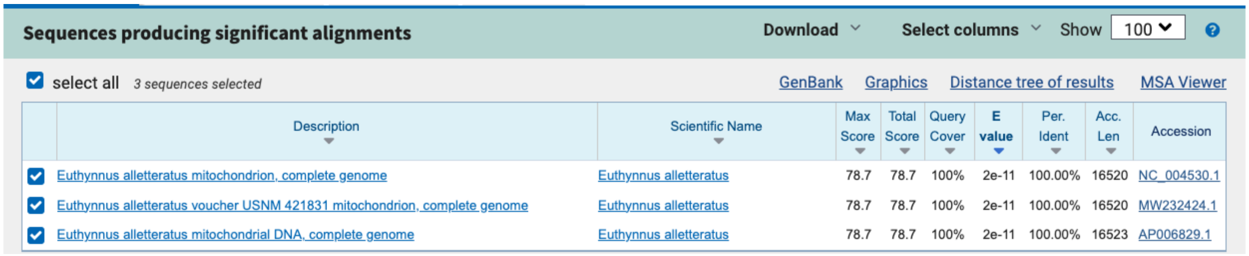


Sequences producing significant alignments Download Select columns Show 100

☑ select all *100 sequences selected* GenBank Graphics Distance tree of results MSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Auxis thazard haplotype 256 tRNA-Pro gene and D-loop, partial sequence; mitochondrial	Auxis thazard	82.4	82.4	100%	1e-12	100.00%	441	JN398926.1
☑	Auxis thazard haplotype 215 tRNA-Pro gene and D-loop, partial sequence; mitochondrial	Auxis thazard	82.4	82.4	100%	1e-12	100.00%	444	JN398885.1
☑	Auxis thazard haplotype 148 tRNA-Pro gene and D-loop, partial sequence; mitochondrial	Auxis thazard	82.4	82.4	100%	1e-12	100.00%	450	JN398818.1

Orange species:

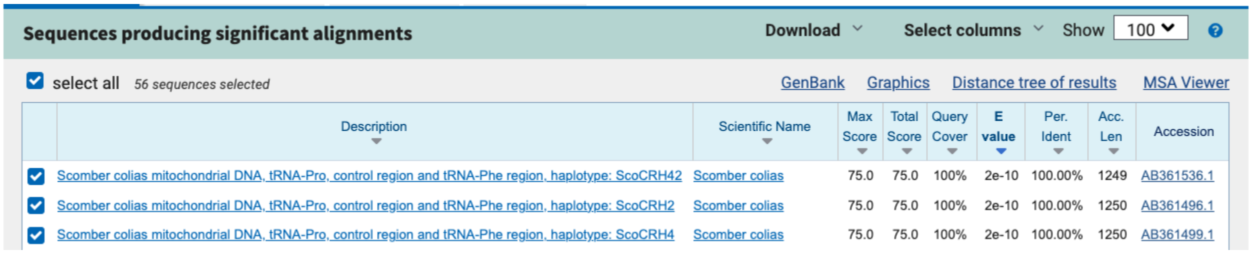


Sequences producing significant alignments Download Select columns Show 100

☑ select all *3 sequences selected* GenBank Graphics Distance tree of results MSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Euthynnus alletteratus mitochondrion, complete genome	Euthynnus alletteratus	78.7	78.7	100%	2e-11	100.00%	16520	NC_004530.1
☑	Euthynnus alletteratus voucher USNM 421831 mitochondrion, complete genome	Euthynnus alletteratus	78.7	78.7	100%	2e-11	100.00%	16520	MW232424.1
☑	Euthynnus alletteratus mitochondrial DNA, complete genome	Euthynnus alletteratus	78.7	78.7	100%	2e-11	100.00%	16523	AP006829.1

Yellow species:

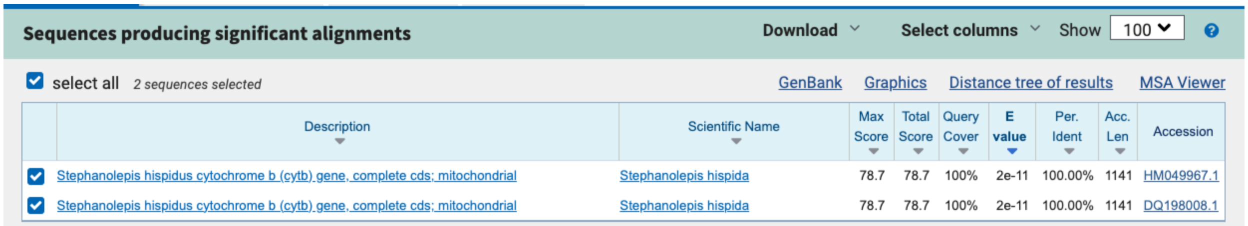


Sequences producing significant alignments Download Select columns Show 100

☑ select all *56 sequences selected* GenBank Graphics Distance tree of results MSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Scomber colias mitochondrial DNA, tRNA-Pro, control region and tRNA-Phe region, haplotype: ScoCRH42	Scomber colias	75.0	75.0	100%	2e-10	100.00%	1249	AB361536.1
☑	Scomber colias mitochondrial DNA, tRNA-Pro, control region and tRNA-Phe region, haplotype: ScoCRH2	Scomber colias	75.0	75.0	100%	2e-10	100.00%	1250	AB361496.1
☑	Scomber colias mitochondrial DNA, tRNA-Pro, control region and tRNA-Phe region, haplotype: ScoCRH4	Scomber colias	75.0	75.0	100%	2e-10	100.00%	1250	AB361499.1

Light Green species:

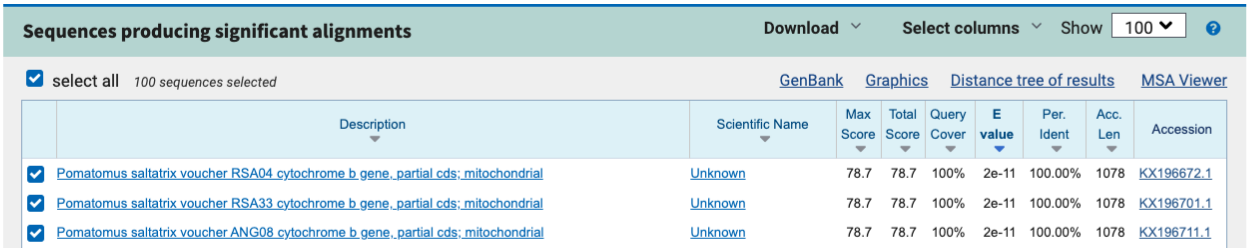


Sequences producing significant alignments Download Select columns Show 100

☑ select all *2 sequences selected* GenBank Graphics Distance tree of results MSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Stephanolepis hispidus cytochrome b (cytb) gene, complete cds; mitochondrial	Stephanolepis hispida	78.7	78.7	100%	2e-11	100.00%	1141	HM049967.1
☑	Stephanolepis hispidus cytochrome b (cytb) gene, complete cds; mitochondrial	Stephanolepis hispida	78.7	78.7	100%	2e-11	100.00%	1141	DQ198008.1

Dark Green species:

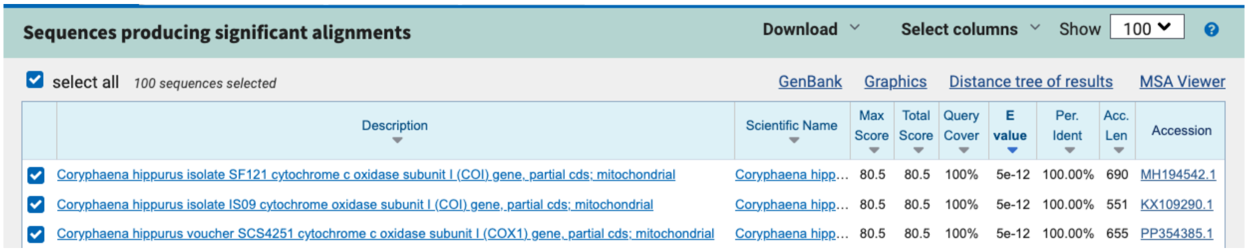


Sequences producing significant alignments Download Select columns Show 100

☑ select all *100 sequences selected* GenBank Graphics Distance tree of results MSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Pomatomus saltatrix voucher RSA04 cytochrome b gene, partial cds; mitochondrial	Unknown	78.7	78.7	100%	2e-11	100.00%	1078	KX196672.1
☑	Pomatomus saltatrix voucher RSA33 cytochrome b gene, partial cds; mitochondrial	Unknown	78.7	78.7	100%	2e-11	100.00%	1078	KX196701.1
☑	Pomatomus saltatrix voucher ANG08 cytochrome b gene, partial cds; mitochondrial	Unknown	78.7	78.7	100%	2e-11	100.00%	1078	KX196711.1

Dark Blue species:

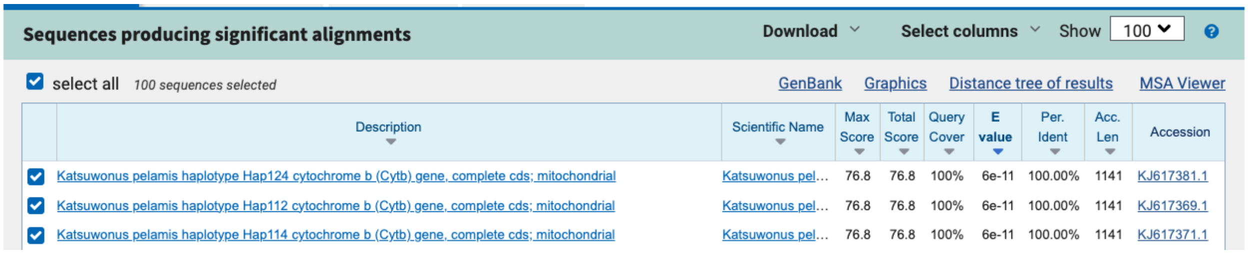


Sequences producing significant alignments

☑ select all *100 sequences selected*

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Coryphaena hippurus isolate SF121 cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondrial	Coryphaena hipp...	80.5	80.5	100%	5e-12	100.00%	690	MH194542.1
☑	Coryphaena hippurus isolate IS09 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial	Coryphaena hipp...	80.5	80.5	100%	5e-12	100.00%	551	KX109290.1
☑	Coryphaena hippurus voucher SCS4251 cytochrome c oxidase subunit I (COX1) gene, partial cds; mitochondrial	Coryphaena hipp...	80.5	80.5	100%	5e-12	100.00%	655	PP354385.1

Pink species:

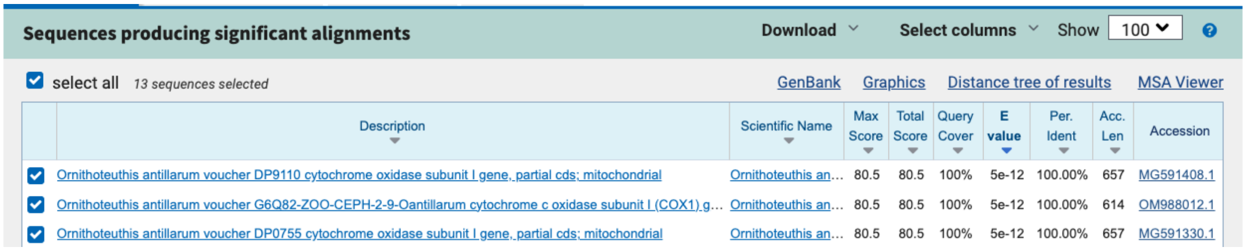


Sequences producing significant alignments

☑ select all *100 sequences selected*

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Katsuwonus pelamis haplotype Hap124 cytochrome b (Cytb) gene, complete cds; mitochondrial	Katsuwonus pel...	76.8	76.8	100%	6e-11	100.00%	1141	KJ617381.1
☑	Katsuwonus pelamis haplotype Hap112 cytochrome b (Cytb) gene, complete cds; mitochondrial	Katsuwonus pel...	76.8	76.8	100%	6e-11	100.00%	1141	KJ617369.1
☑	Katsuwonus pelamis haplotype Hap114 cytochrome b (Cytb) gene, complete cds; mitochondrial	Katsuwonus pel...	76.8	76.8	100%	6e-11	100.00%	1141	KJ617371.1

Purple species:

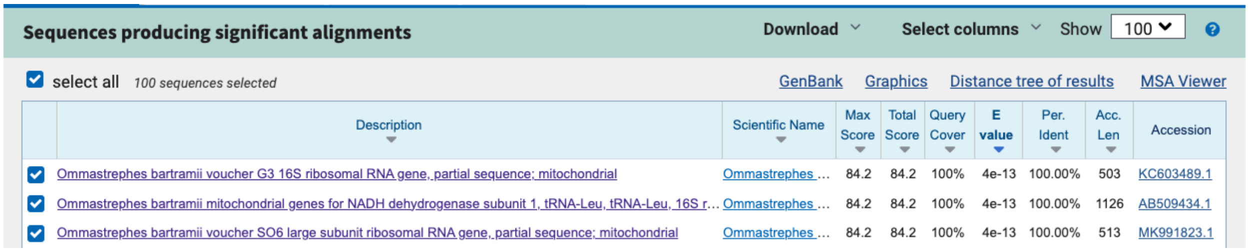


Sequences producing significant alignments

☑ select all *13 sequences selected*

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Ornithoteuthis antillarum voucher DP9110 cytochrome oxidase subunit I gene, partial cds; mitochondrial	Ornithoteuthis an...	80.5	80.5	100%	5e-12	100.00%	657	MG591408.1
☑	Ornithoteuthis antillarum voucher G6Q82-ZOO-CEPH-2-9-Oantillarum cytochrome c oxidase subunit I (COX1) g...	Ornithoteuthis an...	80.5	80.5	100%	5e-12	100.00%	614	OM988012.1
☑	Ornithoteuthis antillarum voucher DP0755 cytochrome oxidase subunit I gene, partial cds; mitochondrial	Ornithoteuthis an...	80.5	80.5	100%	5e-12	100.00%	657	MG591330.1

Light Blue species:



Sequences producing significant alignments

☑ select all *100 sequences selected*

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Ommastrephes bartramii voucher G3 16S ribosomal RNA gene, partial sequence; mitochondrial	Ommastrephes ...	84.2	84.2	100%	4e-13	100.00%	503	KC603489.1
☑	Ommastrephes bartramii mitochondrial genes for NADH dehydrogenase subunit 1, tRNA-Leu, tRNA-Leu, 16S r...	Ommastrephes ...	84.2	84.2	100%	4e-13	100.00%	1126	AB509434.1
☑	Ommastrephes bartramii voucher SO6 large subunit ribosomal RNA gene, partial sequence; mitochondrial	Ommastrephes ...	84.2	84.2	100%	4e-13	100.00%	513	MK991823.1

4. Pick one of them. Click on one of the description rows with the name of the protein (first column). What is the Query sequence and the Subject of the protein?

Red species:

Name: hypothetical protein CTI12_AA335920 [Artemisia annua] *This is a hypothetical protein that students will not be choosing

Name: ubiquitin-like modifier-activating enzyme 1 [Anoplopoma fimbria]

Query: SAYMYLHHTYIL

Subject: SAYMYLPVTYIL

Name: Cyclic nucleotide-binding-like [Phytophthora cactorum]

Query: YMYLHHTYI

Subject: YMYLHHTYV

Orange species:

Name: hypothetical protein [Candidatus Acidiferrum sp.] *This is a hypothetical protein that students will not be choosing

Name: LOW QUALITY PROTEIN: probable E3 ubiquitin-protein ligase HERC3 [Trematomus bernacchii]

Query: FSHLAFTN--DYK

Subject: FEHLAFTNPSDYK

Name: E3 ISG15--protein ligase HERC5 [Dissostichus eleginoides]

Query: FSHLAFTN--DYK

Subject: FEHLAFTNPSDYK

Yellow species:

Name: HEAT repeat domain-containing protein [Gemmatimonadales bacterium]

Query: ELVSGTLLHRSPR

Subject: ELLAETLLHRSPR

Name: hypothetical protein BJ322DRAFT_1009786 [Thelephora terrestris] *This is a hypothetical protein that students will not be choosing

Name: hypothetical protein VaNZ11_006378 [Volvox africanus] *This is a hypothetical protein that students will not be choosing

Light Green species:

Name: related to RTM1 protein [Rhynchosporium secalis]

Query: MK-LWVPPRYM

Subject: MKHLWIPPRYM

Name: related to RTM1 protein [Rhynchosporium commune]

Query: MK-LWVPPRYM

Subject: MKHLWIPPRYM

Name: related to RTM1 protein [Rhynchosporium agropyri]

Query: MK-LWVPPRYM

Subject: MKHLWIPPRYM

Dark Green species:

Name: adenosylcobalamin-dependent ribonucleoside-diphosphate reductase [Dyadobacter arcticus]

Query: SSLSASTYTSDEAY

Subject: SSQSAATYTTDEAY

Name: adenosylcobalamin-dependent ribonucleoside-diphosphate reductase [Dyadobacter sp. CY323]

Query: SSLSASTYTSDEAY

Subject: SSQSARTYTTDEAY

Name: adenosylcobalamin-dependent ribonucleoside-diphosphate reductase [Dyadobacter beijingensis]

Query: SSLSASTYTSDEAY

Subject: SSQSERTYTSDEAY

Dark Blue species:

Name: TetR/AcrR family transcriptional regulator [Ktedonobacteraceae bacterium]

Query: LYLQLYYYSY-HF

Subject: LFLQLFYYSYEHF

Name: MATE family efflux transporter [Kineothrix sp. MSJ-39]

Query: LYLQLYYYSYHFL

Subject: LYLQLYYLGYPFL

Pink species:

Name: hypothetical protein Vlu01_18160 [Micromonospora lutea] *This is a hypothetical protein that students will not be choosing

Name: thiamine pyrophosphate-dependent enzyme

Query: QPSPQPISAET

Subject: QPAPQPVSAET

Name: 4Fe-4S dicluster domain-containing protein

Query: QPSPQPISAET

Subject: QPAPQPISGET

Purple species:

Name: hypothetical protein PV327_008079 [Microctonus hyperodae] *This is a hypothetical protein that students will not be choosing

Name: DUF4396 domain-containing protein [Alkalibacillus salilacus]

Query: QFILPYL--------AIYPMQ

Subject: EFILAYLFGIIFQFLAIYPMQ

Name: DUF4396 domain-containing protein [Gracilibacillus salinarum]

Query: QFILPYL--------AIYPMQ

Subject: QFILAYLFGILFQFYAIYPMH

Light Blue species:

Name: DegT/DnrJ/EryC1/StrS family aminotransferase [Bacteroidales bacterium]

Query: RNLFREFLLKNEIAT

Subject: RDLFREFLLKHDIET

Name: thiamine pyrophosphate-dependent enzyme [Algoriphagus namhaensis]

Query: FREFLLKNEIAT

Subject: FREFILKNEIAT

Name: DegT/DnrJ/EryC1/StrS family aminotransferase [Muribaculaceae bacterium]

Query: RNLFREFLLKNEIAT

Subject: RDLFREFLLKHGIET

5. For the protein you picked, what is the main function of this gene? Tip: use www.uniprot.org or https://www.ebi.ac.uk/interpro/entry/InterPro/#table to learn more about the protein. Google scholar is also a good source to find information about proteins. *Some of the functions are copy-pasted from the websites. Students only need to write a bullet point type list for the functions rather than the whole paragraphs found on the websites. I linked the information at the end of the function lists.*

Red species:

Name: ubiquitin-like modifier-activating enzyme 1 [Anoplopoma fimbria]

Function: Catalyzes the first step in ubiquitin conjugation to mark cellular proteins for degradation through the ubiquitin-proteasome system (PubMed:1447181, PubMed:1606621, PubMed:33108101).
Activates ubiquitin by first adenylating its C-terminal glycine residue with ATP, and thereafter linking this residue to the side chain of a cysteine residue in E1, yielding a ubiquitin-E1 thioester and free AMP (PubMed:1447181).
Essential for the formation of radiation-induced foci, timely DNA repair and for response to replication stress. Promotes the recruitment of TP53BP1 and BRCA1 at DNA damage sites (PubMed:22456334). https://www.uniprot.org/uniprotkb/P22314/entry

Name: Cyclic nucleotide-binding-like [Phytophthora cactorum]

Function: Cyclic nucleotide-binding domain is a group of proteins that bind cyclic nucleotides. https://www.ebi.ac.uk/interpro/entry/InterPro/IPR000595/

Orange species:

Name: LOW QUALITY PROTEIN: probable E3 ubiquitin-protein ligase HERC3 [Trematomus bernacchii]

Function: protein ubiquitination, ubiquitin-dependent protein catabolic process https://www.uniprot.org/uniprotkb/Q15034/entry

Name: E3 ISG15--protein ligase HERC5 [Dissostichus eleginoides]

Function: Major E3 ligase for ISG15 conjugation (PubMed:26355087, PubMed:27534820, PubMed:27564865, PubMed:34572049, PubMed:37279284).

Acts as a positive regulator of innate antiviral response in cells induced by interferon. Functions as part of the ISGylation machinery that recognizes target proteins in a broad and relatively non-specific manner. Catalyzes ISGylation of IRF3 which results in sustained activation, it attenuates IRF3-PIN1 interaction, which antagonizes IRF3 ubiquitination and degradation, and boosts the antiviral response. Mediates ISGylation of the phosphatase PTEN leading to its degradation, thus alleviating its suppression of the PI3K-AKT signaling pathway and promoting the production of cytokines that facilitate bacterial clearance (PubMed:37279284).

Interferes with the function of key viral structural proteins such as ebolavirus structural protein VP40 or HIV-1 protein GAG (PubMed:22093708, PubMed:34572049).

Catalyzes ISGylation of influenza A viral NS1 which attenuates virulence; ISGylated NS1 fails to form homodimers and thus to interact with its RNA targets. Catalyzes ISGylation of papillomavirus type 16 L1 protein which results in dominant-negative effect on virus infectivity. Physically associated with polyribosomes, broadly modifies newly synthesized proteins in a cotranslational manner. In an interferon-stimulated cell, newly translated viral proteins are primary targets of ISG15. Promotes parkin/PRKN ubiquitin E3 ligase activity by suppressing the intramolecular interaction that maintains its autoinhibited conformation (PubMed:27534820).13 publications

(Microbial infection) Functions as an E3 ligase for ISGylation of hepatitis B virus protein X leading to enhanced viral replication due to increased interferon resistance
https://www.uniprot.org/uniprotkb/Q9UII4/entry

Yellow species:

Name: HEAT repeat domain-containing protein [Gemmatimonadales bacterium]

Function: It has been noted that many HEAT repeat-containing proteins are involved in intracellular transport processes.
https://www.ebi.ac.uk/interpro/entry/InterPro/IPR000357/

Light Green species:

Name: related to RTM1 protein [Rhynchosporium secalis]

Function (of the RTM1 protein): Carbohydrate binding
https://www.uniprot.org/uniprotkb/A0A1P8APH7/entry , toxicity resistance
https://pmc.ncbi.nlm.nih.gov/articles/PMC1206678/

Name: related to RTM1 protein [Rhynchosporium commune]

Function (of the RTM1 protein): Carbohydrate binding https://www.uniprot.org/uniprotkb/A0A1P8APH7/entry , toxicity resistance https://pmc.ncbi.nlm.nih.gov/articles/PMC1206678/

Name: related to RTM1 protein [Rhynchosporium agropyri]

Function (of the RTM1 protein): Carbohydrate binding https://www.uniprot.org/uniprotkb/A0A1P8APH7/entry , toxicity resistance https://pmc.ncbi.nlm.nih.gov/articles/PMC1206678/

Dark Green species:

Name: adenosylcobalamin-dependent ribonucleoside-diphosphate reductase [Dyadobacter arcticus]

Function: Catalytic activity, cobalamin and nucleotide binding, DNA replication https://www.uniprot.org/uniprotkb/Q59490/entry , responsible for the conversion of the ribose sugar of RNA into the deoxyribose sugar of DNA. This is the rate-limiting step of DNA biosynthesis https://www.ebi.ac.uk/interpro/entry/InterPro/IPR013344/

Name: adenosylcobalamin-dependent ribonucleoside-diphosphate reductase [Dyadobacter sp. CY323]

Function: Catalytic activity, cobalamin and nucleotide binding, DNA replication https://www.uniprot.org/uniprotkb/Q59490/entry , responsible for the conversion of the ribose sugar of RNA into the deoxyribose sugar of DNA. This is the rate-limiting step of DNA biosynthesis https://www.ebi.ac.uk/interpro/entry/InterPro/IPR013344/

Name: adenosylcobalamin-dependent ribonucleoside-diphosphate reductase [Dyadobacter beijingensis]

Function: Catalytic activity, cobalamin and nucleotide binding, DNA replication https://www.uniprot.org/uniprotkb/Q59490/entry , responsible for the conversion of the ribose sugar of RNA into the deoxyribose sugar of DNA. This is the rate-limiting step of DNA biosynthesis https://www.ebi.ac.uk/interpro/entry/InterPro/IPR013344/

Dark Blue species:

Name: TetR/AcrR family transcriptional regulator [Ktedonobacteraceae bacterium]

Function: DNA and metal ion binding, negative regulation of DNA-templated transcription, response to antibiotic https://www.uniprot.org/uniprotkb/A0A848KDK9/entry

Name: MATE family efflux transporter [Kineothrix sp. MSJ-39]

Function: The MATE family consists of probable efflux proteins including a functionally characterized multi drug efflux system from Vibrio parahaemolyticus, a putative ethionine resistance protein of Saccharomyces cerevisiae, and the functionally uncharacterized DNA damage-inducible protein F (DinF) of E. coli https://www.ebi.ac.uk/interpro/entry/ncbifam/TIGR00797/

Pink species:

Name: thiamine pyrophosphate-dependent enzyme

Function: Thiamine pyrophosphate is a vital coenzyme for transketolase, pyruvate dehydrogenase, and ketoglutarate dehydrogenase through controlling the carbohydrate metabolism and nucleic acids synthesis. The reduced levels of cerebral thiamine dysfunction the activities of thiamine pyrophosphate-dependent enzymes, and subsequently, AD occurrence https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/thiamine-pyrophosphate

Name: 4Fe-4S dicluster domain-containing protein

Function: metal ion binding, iron-sulfur cluster binding https://www.uniprot.org/uniprotkb/Q9ACK9/entry

Purple species:

Name: DUF4396 domain-containing protein [Alkalibacillus salilacus]

Function: This family of proteins is functionally uncharacterized but is a conserved protein domain. https://www.uniprot.org/uniprotkb/A0A6A7LLP5/entry#A0A6A7LLP5-1 https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=cl16817

Name: DUF4396 domain-containing protein [Gracilibacillus salinarum]

Function: This family of proteins is functionally uncharacterized but is a conserved protein domain. https://www.uniprot.org/uniprotkb/A0A6A7LLP5/entry#A0A6A7LLP5-1 https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=cl16817

Light Blue species:

Name: DegT/DnrJ/EryC1/StrS family aminotransferase [Bacteroidales bacterium]

Function: polysaccharide biosynthetic process, involved in 4-amino-6-deoxy-monosaccharide D-perosamine synthesis (perosamine is important in the glycosylation), involved in regulation of cell wall biogenesis https://www.uniprot.org/uniprotkb/B0JS13/entry https://www.ebi.ac.uk/interpro/entry/InterPro/IPR000653/

Name: thiamine pyrophosphate-dependent enzyme [Algoriphagus namhaensis]

Function: Thiamine pyrophosphate is a vital coenzyme for transketolase, pyruvate dehydrogenase, and ketoglutarate dehydrogenase through controlling the carbohydrate metabolism and nucleic acids synthesis. The reduced levels of cerebral thiamine dysfunction the activities of thiamine pyrophosphate-dependent enzymes, and subsequently, AD occurrence https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/thiamine-pyrophosphate

Name: DegT/DnrJ/EryC1/StrS family aminotransferase [Muribaculaceae bacterium]

Function: polysaccharide biosynthetic process, involved in 4-amino-6-deoxy-monosaccharide D-perosamine synthesis (perosamine is important in the glycosylation), involved in regulation of cell wall biogenesis https://www.uniprot.org/uniprotkb/B0JS13/entry https://www.ebi.ac.uk/interpro/entry/InterPro/IPR000653/

6. In these searches where you are learning more about the proteins, there are likely new vocabulary words. List three of them and provide the definition.

Students should choose words that are new to them. A short definition should be appropriate. Optional: have students cite where they got the definitions.