2019

Sequence CSI: The Nitrogen Cycle. Subjects: Life Science, Environmental Science, Marine/Ocean Science - Grades: 9-12

Stephanie Wilson
Virginia Institute of Marine Science

SEQUENCE CSI:
THE NITROGEN CYCLE

Stephanie Wilson
Virginia Institute of Marine Science

Grade Level
High School

Subject area
Biology

The VA SEA project was made possible through initial funding from the National Estuarine Research Reserve System Science Collaborative, which supports collaborative research that addresses coastal management problems important to the reserves. The Science Collaborative is funded by the National Oceanic and Atmospheric Administration and managed by the University of Michigan Water Center. VA SEA is currently supported by the Chesapeake Bay National Estuarine Research Reserve, Virginia Sea Grant, and the Virginia Institute of Marine Science Marine Advisory Program.

Title Sequence CSI: The Nitrogen Cycle

Focus Molecular techniques: Identifying organisms within the nitrogen cycle.

Topics covered:

- Technology in genetic/molecular analysis
- DNA, sequencing
- Gel electrophoresis
- Cycles / relationships

Grade Levels/Subject

High School Biology (9-12)

VA Science Standards

- BIO.1 i) appropriate technology including computers, graphing calculators, and probeware, is use for gathering and analyzing data, communicating results, modeling concepts, and simulating experimental conditions;
- BIO.5 g) the structure, function, and replication of nucleic acids;
- BIO.5 j) exploration of the impact of DNA technologies;
- Bio.8 b) nutrient cycling with energy flow through ecosystems;

Learning Objectives

- Students will apply practical scientific molecular methodology in the classroom to solve a problem.
- Students will utilize gel electrophoresis pictures in order to identify samples.
- Students will identify DNA sequences using a practical, scientific database: NCBI.

Total length of time required for the lesson

45 minutes to 1 hour

Key words, vocabulary

- **DNA:** Deoxyribonucleic acid, the double-stranded carrier of genetic information, is present in almost all living organisms. It genetically determines the makeup of an organism such as their hair or eye color. There is a unique code for every organism to make up the double helix of DNA.
- **DNA Sequence:** The code that makes up DNA is its sequence, it consists of four nucleotides adenine (A), cytosine (C), guanine (G), and thymine (T). The order of the nucleotides codes for specific proteins and characteristic in an organism.
- **Gel Electrophoresis:** A molecular analysis technique that separates DNA out by size using an electric charge. Depending on how long a sequence is, how many basepairs or nucleotides are in

it, it will move a certain distance along an agarose gel. DNA is a negatively charged molecule and therefore it can be pulled through a gel with smaller DNA fragments moving further than larger ones.

- **Gene:** A gene is encoded by a section of an organism's DNA, genes will give rise to characteristics in an organism and can be heritable (passed from parent to offspring).
- **Nitrogen Cycle:** The nitrogen cycle is the movement of nitrogen throughout an ecosystem – the full nitrogen cycle moves nitrogen through the geosphere, the atmosphere, and the biosphere. Nitrogen can change forms, for example it can be a gas, or it can be fixed to form ammonium, or it can be found in other larger compounds such as amino acids.
- **Scientific Standard:** A standard is a known measurement that can be utilized in scientific analyses; the standard is a reference point that is used to calibrate sample measurements.
- **Sequencing:** A molecular analysis technique in which the code of DNA, made up of its nucleotide sequence, is read. This can be used to identify organisms because each organism has a unique sequence.
- **Sequence Database:** A database that can be used to compare sequences to those that have already been identified in order to identify an unknown. These databases can also be used to search for certain sequences.

Background Information

See background PowerPoint.

My research focuses on biogeochemistry and, more specifically, the cycling of nitrogen in coastal ecosystems. To answer questions about the nitrogen cycle, I do a lot of molecular analyses to identify and quantify microorganisms because the majority of the nitrogen cycle is mediated by microbes. To identify microorganisms, I utilize DNA sequencing of a gene called the 16S ribosomal RNA gene that is present in all bacteria and archaea. This allows me to compare gene sequences to databases of previously identified organisms and identify the organisms in my own sample. Identifying which organisms are present can allow me to infer what potential biogeochemical reactions may be occurring in the environment. Nitrogen and its cycling is extremely important; all organisms require nitrogen, it can promote primary production in aquatic environments where it can be a limiting nutrient for phytoplankton or as fertilizer in agriculture, and it can be a pollutant when in excess for example when there are high levels of nitrogen in wastewater effluent.

This lesson focuses on identification of organisms using their DNA, because every organism has a genetic code. Using biotechnology, scientists can identify organisms to help them answer questions about the environment. For example, gel electrophoresis which separates DNA fragments by size can help to identify organisms. DNA sequencing is another avenue of identification in which previously identified organisms whose sequences have been deposited in databases can be used to identify unknown samples. These techniques are used in cutting edge research labs for identifying organisms, examining genetic variation, and assessing ancestry. Once organisms are identified relationships between them can be assessed. For example, most of the nitrogen cycle is completed

by microorganisms. Identifying which microorganisms are present in a sample can allow scientists to make inferences as to what cycling reactions may occur.

Handouts and references: see table below

-Envelopes to put "samples" in

1. CSI Sequence: Sequence Identification Worksheet (App. 1a)
2. Gel Electrophoresis Results (App. 2a-d)
3. NCBI Search Sequences (App. 3a, 4 sequences) & KEY (App.4)
4. How to Use NCBI website & Nucleotide BLAST Search – Instructions (App.4)
5. CSI Sequence: Connections Worksheet (App. 5a)
6. CSI Sequence: Organism Information Sheets (App. 6a-d)

Title	For Students (worksheets & references)	For Teacher (references or KEY)
1. CSI Sequence: Sequence Identification Worksheet	**Appendix 1a:** Student worksheet	**Appendix 1b:** **Teacher's ANSWER KEY** for Sequence ID worksheet
2. Gel Electrophoresis Results	**Appendices 2a – 2 d:** Sample A Sample B Sample C Sample D	**Appendices 2a – 2 d:** Sample A = bacteria Sample B = bacteria Sample C = plant Sample D = mammal
3. NCBI Search Sequences Better to have these already loaded on computers or on a flash drive so that students can copy and paste their assigned sequence into the NCBI website	**Appendices 3a** (4 sequences): Sample A Sample B Sample C Sample D	**Appendices 4** (4 sequences): **Teacher's ANSWER KEY** for Appendices 2a – 2 d Sample A = N-fixing bacteria Sample B = Denitrifying bacteria Sample C = Soybean plant Sample D = Cow, mammal
4. How to Use NCBI website & the Nucleotide BLAST Search	**Appendix 3b**	
5. Organism Information Sheets	**Appendices 5 a – d:** Sample A = *Rhizobium lemnae* (N-fixing bacteria) Sample B = *Rhodanobacter denitrificans* (denitrifying bacteria) Sample C = *Glycine max* (soybean) Sample D = *Bos taurus* (cow)	
6. Sequence Connections Worksheet	**Appendix 6a:** Student worksheet	**Appendix 6b:** Teacher's ANSWER KEY for Sequence Connections worksheet

Materials and supplies

-Projector/Screen for introductory PowerPoint

-Computer for each team (or each student) to access NCBI Website

-NCBI Website:
https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome

Classroom Set-Up

Tables arranged in groups for teams of 2-4 students.

Procedure

Introduction:

Attention Grabber:

- Have any of you ever written in code? Or have you ever had to decipher a code to see what it means?
    - The PPT has a code they can try to decode
- The introduction PowerPoint can be used to review concepts about DNA.
    - This will also be an introduction to molecular techniques such as gel electrophoresis and DNA sequencing.
- PowerPoint Script:

Slide 1: Presentation Title

Slide 2: Attention Grabber

Students should try to decode the message – once they figure it out you can relate this to how DNA encodes information, just like the code they read!

Slide 3: My research

These are picture of me in the lab and in the field:
I am a graduate student that studies the nitrogen cycle in coastal environments. In these pictures I am exploring the nitrogen cycle in the lab and in the field. The first picture is a picture of me at my study site with some of my colleagues we are collecting water and sediment samples in order to look at biogeochemical processes occurring at this beach; The Gloucester Point Beach in Virginia. The second photo is one of me in the lab performing molecular analyses on my samples from the field; here I am extracting DNA! This DNA will be used to determine what organisms are living in the water and sand of the Gloucester Point Beach.

Slide 4: DNA is like a code!

Almost every organism on the planet has DNA. It encodes the genetic material of the organism and serves as the blueprint for the organism's characteristics. Each organism's genetic code is unique and therefore they can be identified using that code – for

example DNA from crime scenes can be used to identify suspects or DNA from an organism can be used to identify what kind of organism it is.

Slide 5: DNA Review

Remember that DNA stands for deoxyribonucleic acid and it is a double stranded molecule. The code is made up of four nucleotide bases Adenine, Thymine, Cytosine, and Guanine. Does anyone remember which bases pair to with each other? A-T and C-G! The two strands of DNA, made up of these bases, are held together by hydrogen bonds. The nucleotides make up the code that we can use to identify organisms!
If you want students to spend more time learning about DNA there is an interactive website with tons of information: http://www.dnai.org/index.htm

Slide 6: The Chemical Structure of DNA

This is an optional slide from https://www.compoundchem.com/2015/03/24/dna/ which can be another source for reviewing DNA and its structure with the class.

Slide 7: Molecular Techniques or Biotechnology

Does anyone know how we can identify organisms using their DNA? Molecular techniques or biotechnology can be used to identify organisms. And we can use techniques such as Gel electrophoresis – shown in the picture or DNA sequencing! So let's looks at both of these a bit closer.

Slide 8: Gel Electrophoresis

This is a method of identifying DNA by separating DNA by size. A picture of a gel is shown again in the picture. Gel electrophoresis can allow for an identification or an inference as to what a DNA sample might be when it is compared to standards. Let's watch a quick video to learn a little bit more about how gel electrophoresis works.

Slide 9: DNA Sequencing

Sequencing lets us read the code of DNA and maybe identify the organism that DNA originates from. We can do this by breaking up the DNA from a long strand into shorter fragments which are then copied and replicated introducing fluorescent tags. The fluorescent tags indicate which base pairs are present and this can be read by a sequencing machine to determine the sequence. This can be used to identify unknown DNA samples, look at the genetic variation in a population, or to look at ancestry. To get some more details let's watch this video.

Slide 10: Sequence Databases

After sequencing DNA from a known organism we can keep that sequence for use later in database. Just like you when look up someone's phone number in a phonebook or online, sequence databases allow you to look up sequences and find out who they belong to. This can be extremely useful when identifying organisms in the environment such as microbes. Scientists interested in identifying microorganisms utilize sequencing databases often to determine what microbes are present in various environments.

A. Sample Identification Activity:

1. Students will break up into groups of 2-4 students. Each group will receive a color coded mystery sample (in an envelope) which they will have to try and identify.
2. Have the students follow the CSI Sequence Identification handout (App. 1a) and answer questions as they go.
3. Students should answer the first two questions and think-pair-share with their group.
4. The groups then ask for their sample's gel electrophoresis results (App. 2a-d) to analyze. The gel will give them a preliminary answer as to what their mystery organism is. Due to their low resolution result, students will have to think about how they can use other techniques to identify the organism with greater accuracy.
5. Students will then have to use their organism's sequence listed on the Sample Sequence Handout (App. 3a) in order to determine what it really is and distinguish between the two potential answers. It will be easiest if the sequences are loaded onto a thumb-drive or if the files are already on the available computer so that they can copy and paste the sequences into the BLAST search website.
6. Use the BLAST search to identify the sample per the instructions sheet: How to NCBI Search (App. 4) that has an example which identifies *E.coli* from a sequence.
7. Next, students can research about their organism on their own (I have provided some links about each sample organism below to start them off exploring the web) or they can get basic information from an info sheet provided for each sample in Appendix 5a-d.
   - Sample A: Britannica on Nitrogen Fixing Bacteria
   - https://www.britannica.com/science/nitrogen-fixing-bacteria
   - Sample B: Britannica on Denitrifying Bacteria
   - https://www.britannica.com/science/denitrifying-bacteria
   - Sample C: Encyclopedia of life Soybean overview
   - http://eol.org/pages/641527/overview/
   - Sample D: Encyclopedia of life Cow overview
   - http://eol.org/pages/328699/overview

B. Making Connections/Relationship Activity:

1. Students will form new groups of four (jigsaw structure) that include at least one student from each team from Activity A (one of each color). This should create a group where each student, or every few students, had a different organism to identify. They will work together on the CSI Sequence Connections handout (App 6a).

   Students should start by telling the other members of the group about the organism they researched earlier ("their organism").
2. Students should try to figure out what connects each of their organisms based on the knowledge they have about that organism (i.e. Nitrogen cycling), the key for this worksheet in Appendix 6b.
   The group should work together to sketch out how they believe their organisms relate to one another in a flow diagram (for example a cow might eat a plant and a plant needs nitrogen to grow which it will get from the bacteria that provide it, and so on).
3. An extension to this portion of the lesson would be to ask students to characterize the relationships and discuss potential consequences of losing a link within the cycle. Example: What might that mean for agriculture if we lost N-fixing bacteria?

4. Groups can then share how they think their organisms are related to one another with the rest of the class before going over the correct answer. This will give students an opportunity to share what their group has determined to be the cycle and compare that with what other students think as well.
5. The lesson can be used as an introduction to the Nitrogen cycle and if the class wants to explore the nitrogen cycle in more detail there is an interactive learning portal through PBS Learning Media at the following link: https://www.pbslearningmedia.org/resource/lsps07.sci.life.eco.nitrogen/the-nitrogen-cycle/

C. Closing/Summary:

- Pose a question to the class: What other ways can the identification techniques you used to identify your sample be used in the real world?

- Examples: DNA sequencing for paternity, crime scene DNA testing,

- The class may also use the Nitrogen cycle as a review activity – going over the steps of the nitrogen cycle to connect the investigation they completed with sequences to the cycle.

- Nitrogen cycling is an integral part of ecosystem function and in order to understand how nitrogen flows through a system we need to be able to identify what organisms are utilizing nitrogen in its different forms. DNA sequencing can help scientists do this by identifying the organisms present so that their relationship to the nitrogen cycle can be evaluated.

Assessment

The students will be required to fill out a handout as they go through the activity to mark their progress. At the end of the activity students will have to create a diagram in order to show how they believe their organisms relate to one another.

The groups can each present to the class how they set up their flow diagram and explain how they determined their organisms are related before the teacher goes over the correct answer. This will allow each group to present their ideas and explain them to the other students together.

References

"Ayrshire." *Encyclopedia of Life*, eol.org/pages/328699/overview.

"BLAST: Basic Local Alignment Search Tool." *National Center for Biotechnology Information*, U.S. National Library of Medicine, 2018, blast.ncbi.nlm.nih.gov/Blast.cgi.

"CBR3 Carbonyl Reductase 3 [Bos Taurus (Cattle)] - Gene - NCBI." *Current Neurology and Neuroscience Reports.*, U.S. National Library of Medicine, 2018, www.ncbi.nlm.nih.gov/gene/516036.

"Denitrifying Bacteria." *Encyclopædia Britannica*, Encyclopædia Britannica, Inc., 14 Dec. 2011, www.britannica.com/science/denitrifying-bacteria.

Elsik, Christine, et al. “Taxonomic Revision of Rochefortia Sw. The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution.” *Science*, 2009, doi:10.3897/bdj.4.e7720.figure2f.

Green, Stefan J., et al. “Denitrifying Bacteria Isolated from Terrestrial Subsurface Sediments Exposed to Mixed-Waste Contamination.” *Applied and Environmental Microbiology*, American Society for Microbiology, 15 May 2010, aem.asm.org/content/76/10/3244.

“Nitrogen-Fixing Bacteria.” *Encyclopædia Britannica*, Encyclopædia Britannica, Inc., 26 Oct. 2018, www.britannica.com/science/nitrogen-fixing-bacteria.

“Rhizobium Lemnae Strain JK4-Gm 16S Ribosomal RNA Gene, Partial Sequenc - Nucleotide - NCBI.” *Rhizobium Lemnae Strain JK4-Gm 16S Ribosomal RNA Gene, Partial Sequence*, U.S. National Library of Medicine, 2017, www.ncbi.nlm.nih.gov/nuccore/MF784377.1.

“Rhodanobacter Denitrificans.” *Staphylococcus* , Microbewiki, 2012, microbewiki.kenyon.edu/index.php/Rhodanobacter_denitrificans.

“Rhizobium Lemnae Strain JK4-Gm 16S Ribosomal RNA Gene, Partial Sequenc - Nucleotide - NCBI.” *Rhizobium Lemnae Strain JK4-Gm 16S Ribosomal RNA Gene, Partial Sequence*, U.S. National Library of Medicine, 2017, www.ncbi.nlm.nih.gov/nuccore/MF784377.1.

Schmutz, Jeremy, et al. *Genome Sequence of the Palaeopolyploid Soybean*. Nature , 2010, www.nature.com/articles/nature08670.pdf.

“Soybean.” *Encyclopedia of Life*, 2012, eol.org/pages/641527/overview/.

“The Nitrogen Cycle.” *PBS LearningMedia*, 2018, www.pbslearningmedia.org/resource/lsps07.sci.life.eco.nitrogen/the-nitrogen-cycle/.

Wagner, Stephen. “Biological Nitrogen Fixation.” *Nature News*, Nature Publishing Group, 2011, www.nature.com/scitable/knowledge/library/biological-nitrogen-fixation-23570419.

Zimin, Aleksey V, et al. “A Whole-Genome Assembly of the Domestic Cow, Bos Taurus.” *Genome Biology*, vol. 10, no. 4, 2009, doi:10.1186/gb-2009-10-4-r42.

Appendix: 1a **Sequence ID Worksheet**

CSI Sequence **Name:_____**

Sequence Identification Worksheet:

1. What color is your DNA sample?

2. What is your DNA sample sequence?

3. According to the Gel Electrophoresis results what organism might your sample be from? Compare the your unknown sample to known samples 1-4 which sample pattern matches your unknown?

4. How might you determine what your organism is exactly?

5. What are the first ten nucleotides in your organism's full sequence?

6. What does the BLAST search engine tell you your organism is?
   What is the % Ident for the sequence identity?

7. What are three characteristics of your organism?
   (example: How does it get energy? What kind of organism is it?)

CSI Sequence – ANSWER KEY

Sequence Identification Handout:

1. What letter and color is your DNA sample?
   Sample A: Yellow (N fixing bacteria)
   Sample B: Green (Denitrifying bacteria)
   Sample C: Blue (Plant – Soybean)
   Sample D: Purple (Herbivorous mammal – cow)

2. What is your DNA sample sequence?
   Sample A: CTACGGGAGG
   Sample B: CACCCGCACA
   Sample C: CTTATTTAAT
   Sample D: AGCCTCTTTC

3. According to the Gel Electrophoresis results what organism might your sample be from?
   Sample A: Bacteria
   Sample B: Bacteria
   Sample C: Plant
   Sample D: Animal

4. How might you determine what your organism is exactly?

   DNA Sequencing – this will allow you to determine what exactly your organism is via its genetic code.

5. What are the first ten nucleotides in your organisms full sequence?
   Sample A: CTACGGGAGG
   Sample B: CACCCGCACA
   Sample C: CTTATTTAAT
   Sample D: AGCCTCTTTC

6. What does the blast search engine tell you your organism is? What is the % Ident for the sequence identity?
   Sample A: *Rhizobium lemnae* % Ident: 100%
   Sample B: *Rhodanobacter denitrificans* % Ident: 100%
   Sample C: *Glycine max* % Ident: 100%
   Sample D: *Bos taurus* % Ident: 98%

7. What are three characteristics of your organism? (example: What kind of organism is it?, What is it used for? Etc.)

   Sample A: N fixing bacteria, symbiotic within plants, prokaryotes, microorganism (small),

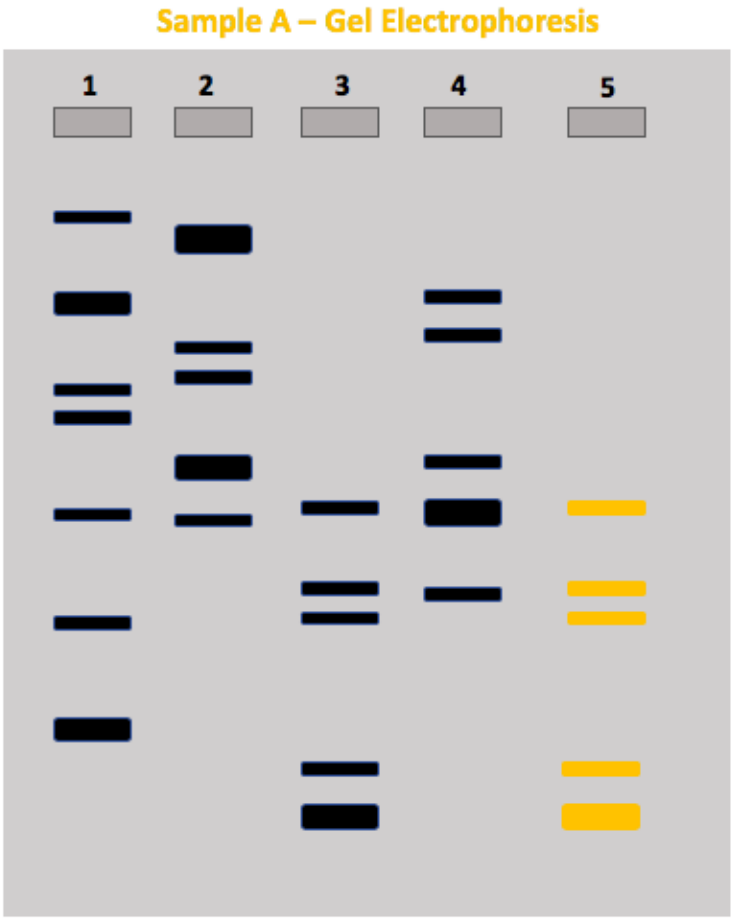
   Sample B: Bacteria, removes nitrogen, requires anoxia/no oxygen, microorganism (small),

   Sample C: Plant, used for animal feed, used in food products, used in personal care products, requires nutrients and sunlight.

   Sample D: Animal, eats plants (herbivore), mammal, ruminant (four stomachs), used for dairy products and beef

Appendix: 2a: **SAMPLE A – Gel Electrophoresis Results**

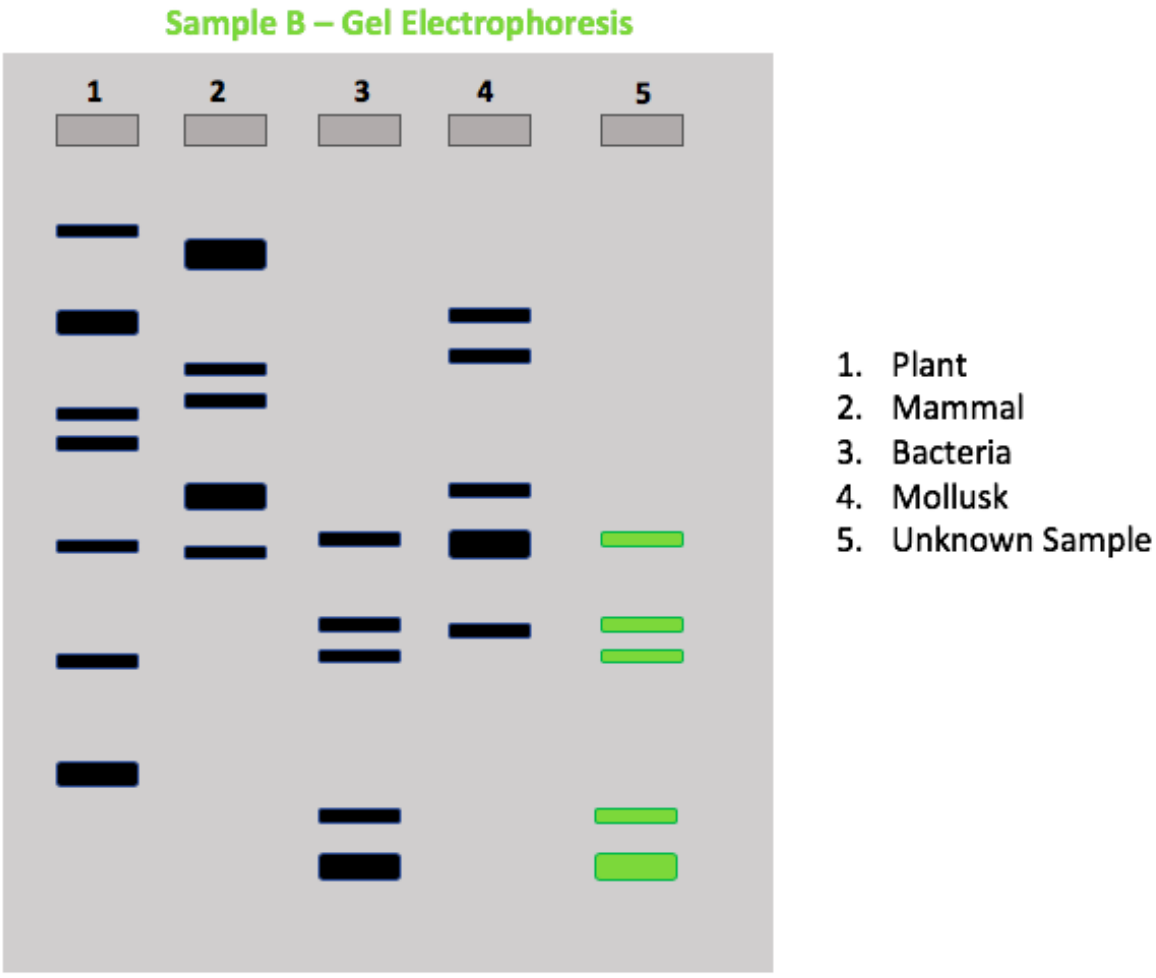
Gel Electrophoresis for Sample Identification:



1. Plant
2. Mammal
3. Bacteria
4. Mollusk
5. Unknown Sample

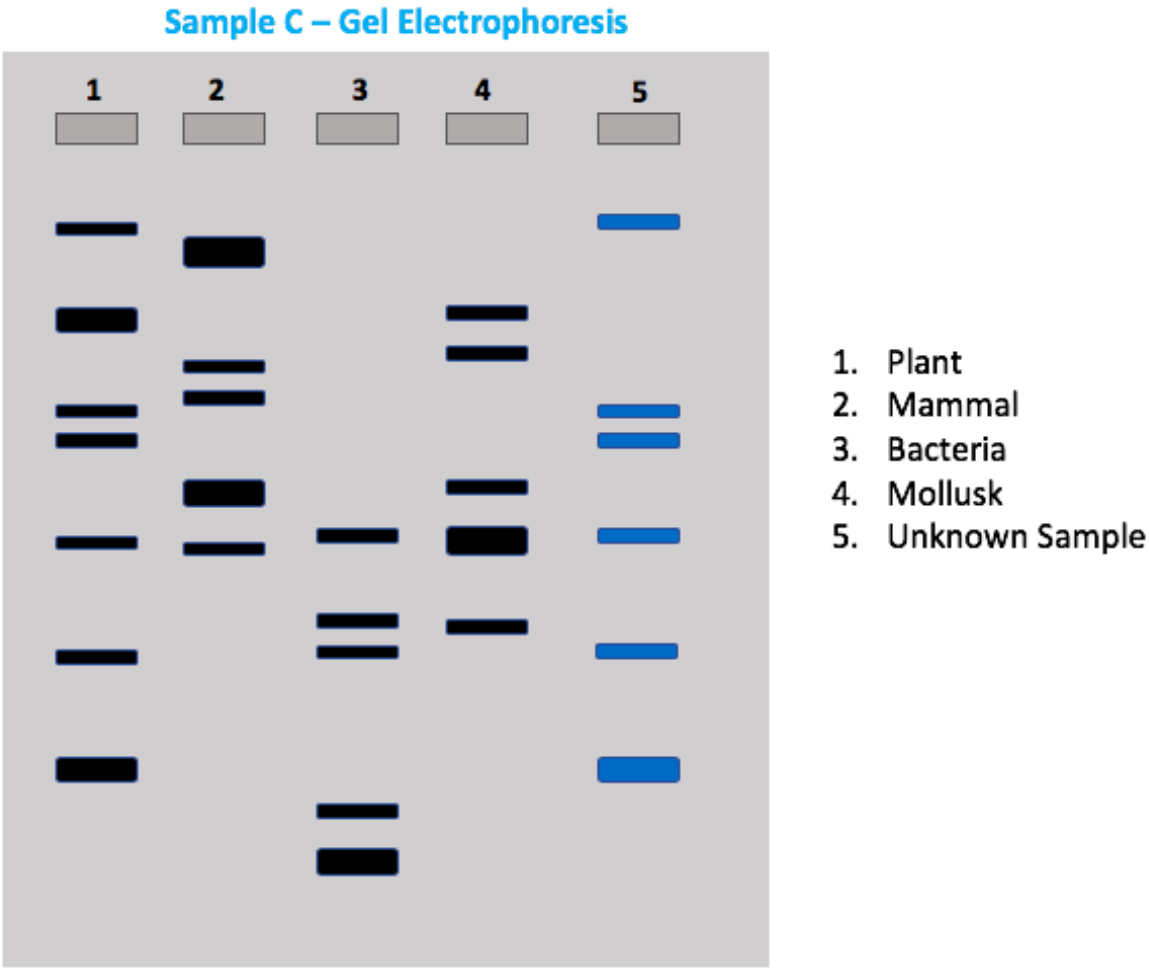
Appendix: 2b **SAMPLE B – Gel Electrophoresis Results**

Gel Electrophoresis for Sample Identification:



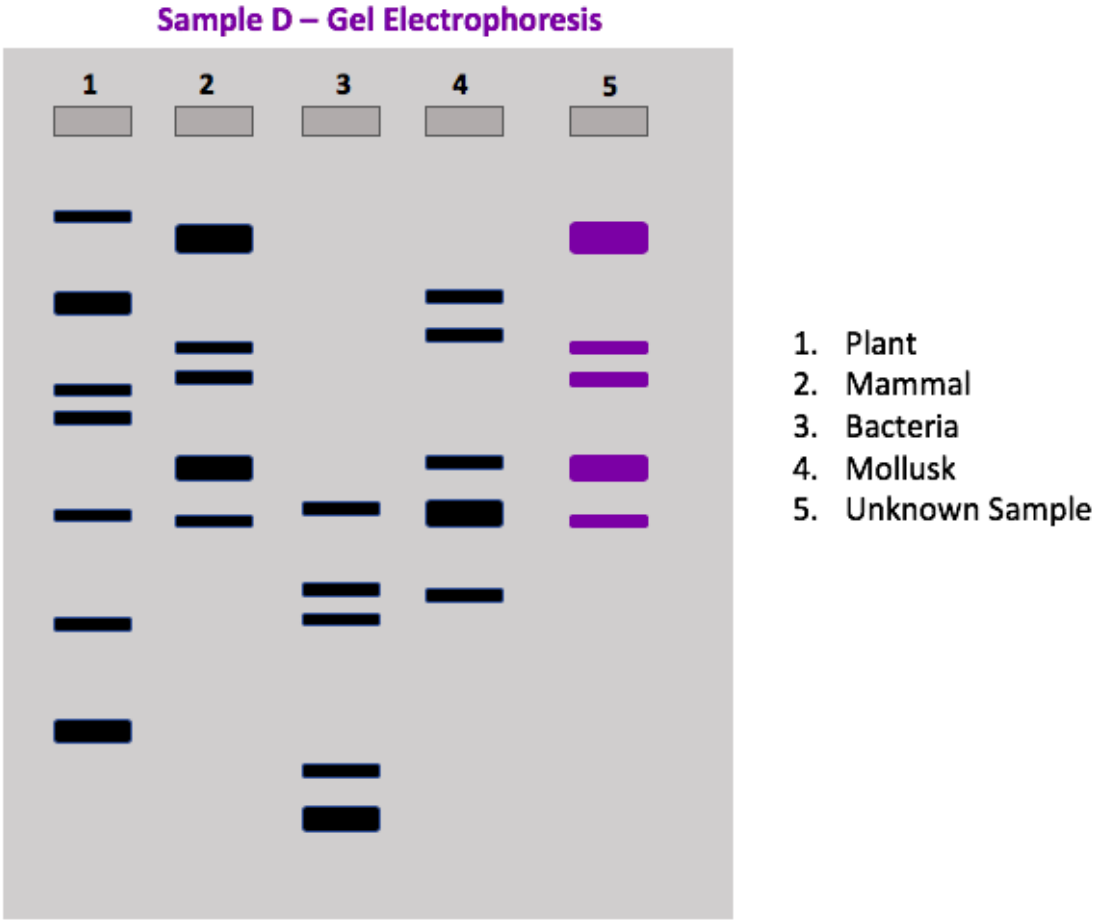
Appendix: 2c **SAMPLE C – Gel Electrophoresis Results**

Gel Electrophoresis for Sample Identification:



Appendix: 2d **SAMPLE D – Gel Electrophoresis Results**

Gel Electrophoresis for Sample Identification:



Appendix: 3a **NCBI Search Sequences**

CSI Sequence

Sample Sequences for NCBI Search:

Copy and paste sequence into NCBI search – **Read Instruction Sheet first!**

Sample A Sequence:

1 ctacgggagg cagcagtggg gaatattgga caatgggcgc aagcctgatc cagccatgcc
61 gcgtgagtga tgaaggcctt agggttgtaa agctctttca ccggagaaga taatgacggt
121 atccggagaa gaagccccgg ctaacttcgt gccagcagcc gcggtaatac gaaggggct
181 agcgttgttc ggaattactg ggcgtaaagc gcacgtaggc tgacatttaa gtcaggggtg
241 aaatcccggg gctcaacctc ggaactgcct ttgatactgg gtgtcttgag tgtggaagag
301 gtcagtggaa ttgcgagtgt agaggtgaaa ttcgtagata ttcgcaggaa caccagtggc
361 gaaggcggct gactggtcca caactgacgc tgaggtgcga aagcgtgggg agcaaacagg
421 attagatacc ctggtagtcc acgccgtaaa cgatgaatgt tagccgtcgg caagtttact
481 tgtcggtggc gcagctaacg cattaaacat tccgcctggg gagtacggtc gcaagattaa
541 aactcaaagg aattgacggg ggcccgcaca agcggtggag catgtggttt aattcgaagc
601 aacgcgcaga accttaccag cccttgacat gcccggctca ccacagagat gtggttttcc
661 cttcggggac cgggacacag gtgctgcatg gctgtcgtca gctcgtgtcg tgagatgttg
721 ggttaagtcc cgcaacgagc gcaaccctcg cccttagttg ccagcatttg gttgggcact
781 ctaaggggac tgccggtgat aagccgagag gaaggtgggg atgacgtcaa gtcctcatgg
841 cccttacggg ctgggctaca cacgtgctac aatggtggtg acagtgggca gcgagcacgc
901 gagtgtgagc taatctccaa aagccatctc agttcggatt gcactctgca actcgagtgc
961 atgaagttgg aatcgctagt aatcgcggat cagcacgccg cggtgaatac gttcccgggc
1021 cttgtacaca ccgcccgt

Source: National Library of Medicine on NCBI, 2017

Sample B Sequence:

1 cacccgcaca tgagccagac cgacggccac tacgacggcc gctacatctt catcaacgac
61 aaggcgcaca cccgcgtggc gcgcatccgc tgcgacgtga tgcgcaccga ccgcatcacc
121 gaggtaccca acgtgcaggc gatccacggc ctgcgcgtgc agcgcgcgcc gcgcaccggc
181 tacgtgttcg ctaacggcga gttcctggtt ccggcgccga acgacggccg cgacctggag
241 gacccgaaga aataccagac catgttcagc gcgctcgacg gcgacaccat ggaggtcgcc
301 tggcaggtgc tggtggacgg caacctggac aacaccgagg ccgactacgc cggcaagtac
361 gcgatctcca cctgctacaa cagcgagggc ggcatggacc tgcccagcac gatgcaggcc
421 gagcgcgact gggcggtggt gttcgacatc gcgcgcatcg aggcggcggt gaagaagggc
481 gacttcaaga ccatcggcga ttcgaaggtg ccggtgctgg acggccgcca cggctcctcg
541 ttcacgctat acatcccgat cccgaagaac ccgcacggga tcaacgccag ccccgacggc
601 aagtacgtgg tggccaacgg caagctctcg cccactgtca ccgtgatgga gtggagccgc
661 atcgacgact ggttcgccgg caagctcaag gacccgcgcg acaccgtggt ggccgagccc
721 gaactgggcc tgggcccgtt gcacaccgcc tacgacgggc gcggcaacgg ctacaccacg
781 ctattcatcg acagccagat cgccaagtgg aacatcgccg acgcgatcgc cgcccacgac
841 ggcaag

Source: Green, S. et al. 2010

Sample C Sequence:

CTTATTTAATAATTTTTTCCCCTAAAATATCCTTCTAGTTAACGTGTCTTATTTTTTCTTCTTGACATCCGAAAGATTAT
GATTATATGTACCTTATAAGAGAAATGCTATTGATTTAAATTGATTGGAAACTATCCAATTATGAGAGAAAATTATT
AAATAAGATAATTAATTATAATTTTGTGTAATTTTTAAGAAATTTAATACAATAAAAAAGAGTGTTGAAAAGACTGT
GATACAGTGTTGTTAGCATTTTTCGTATAAATAACGCGAGAGAGGTTAGAGTGAACGAGAGAAAGTGATGGATTA
CGATTTAGCAGCGTTGAAGCTTTTCTGTTCGCAGCTGAAGCAGGCTCGAGAAGTAGCAATGCCTTGTCAAACTCAA
AGTCAAAGTCAAAGCAGCTTCACTCTCGGCGGCCTCCTTTTCCAGCGCCCTTGGTTGCAGGTTTCCACTCCCATTTC
CATTCCCTAAACGCCGCCGTTTTCTCTCTCTCATTTCCATTTTCATTTTCGCCTAATCTCTCTCTCTAAGGGCGTTCTC
GTCTCCGCCTCCGACGCCGGCTCTCTGCTTCTCGACGACGGAACCGACCTCGTCGAGCTCTCCCTCAATAGCGAGTT
TCGCCACCGTCCATGGAAACTCGGTAATCAATGCTTCTTCTCTCTTTGTTCAATTCAATGTATTCTATATGTCTTTAAC
GAACCAAAAGATGGAAATTCTCCTCAACAAAAGACAATAATAGGAGGTGATTTATAATGATAATGCTTTGTAATTT
GTAGAATTGCAATTCTATATTGATCTTTTGGGTAAATACCTATTTTGGTCCTGAAAGTGTGAGGCGGTGTTAAATTG
ATCCCCAAAGATGAAAAATTTAAATTTAGTTCCTGGATGTGTTAAAAGTGCCATAGATTGGCCCTGCTTACAATTTA
ATGACCCTCAACTTTGTAACTTAATGTTAAATAGGTCCTACAAAAATATAACTTATTTTAAGTTAATCCTTGAACATT
AAAATGGTCTCAAGGATTTAATGACAGGACCTCACACTTTTTATACATTCAAAGACTAAATTTAAATTTTTCGTCTTT
TAAGGTCCAATTTGTCATCTAAGGTCCAATTTGTCATCGTCTTACATTTTCAAGGACCAAAATGAATATTTACCTTCC
ATTTTTTATGCATTTTTATACATGTATCTTACACGAATTGTATTCTATATAACTTTTTAGAATAATCGCATGGTACTGC
AACCGTACTGAATGACGTCGTGTGATTCATGCAGTGAACTTCACCTTTGTTGTTGTTATTGTAATCATACAGGTCTG
CCTTATCGGATACGGATCGTATCTGATGCATGTATTGTATCTGTGTATCATAGATTTATTTACTTTGTCAGGGATGT
ATGTAATGGTGGTTGGAGGGTATGCTGCACGTCCAGGGAAGCTTCCCATGATCAAGGTATTGTACACGTTTTTATC
ATTTTGAGTTAGAGGCACTTCTAAAATAAATGAATTAATCTCAAAATTAGCATTCATACTTGCATATATTTTCTCTCA
CACAAAATAATGGAGTTTGTTTTTGGTGATGTTCAAGTAAATGTATAAAATCTGATTTTCAATTTCCGTCTATCAAG
AGGAGCATTTGTATATGGCTATATGCACTCAGGTTAGGTGGATGTATGTAGTGGGAAGAAATATTTATTAGGTGTC
TTGTTTACTGGTAGAGTATAGGGAACACCACGGCTACATGCCTATGGTAGTTTTTCGACTGTTTGTGTTTGGATAAC
TGTTCTGAATTTACATTTGGGTGTTGAAAGTACTTTTGGAGGAAACTCTTTTTGTGTGTTTGGATGGGCGTTAATAA
AATTGGTTTTGGTGAAAATTGATTTTGAAGTGACGTGACTTATGTTTAGATATTTTATTATAAAATCAAGTTAGGAG
TCAATTTCAGTACAAAATTTTGGATCTAACGCAGAAGCTACTCAAAGTTGTTTCGACCCAGAATCAATTCTGAATTC
TTCTCCAATGTGAAACCAAACATGTGAAAATGTATCCAAAATCAATTATAAACTCAGAATCAATTTCTCAACACCAA
ACCAAACATGCACTCTATCTTTTATTCTAAACATACTTTTGAAAAAGGTGCAACAATTTTCCAAACTTACATGAAATA
AGCTGAATTTGTTCTTCCTATGGTTTGCTTAGCCTATAATTTGTGTGACCTGAAACAGAATAAATTTTTCATGGAAA
GGAACTGGTTTTCTTTCTTTTATTGGAACATTGCTAGAAAGAGAAAATGCTAGACAACACTTTTGGCACTCCTTTTA
TCACCCTTTATTATTGTGTGAAATCTAATGCTGATCTCACCAATAAGGAATAGGACAATTCAAAAGTTAGGTGCAGT
ATCATGAGTGATTTTTGTGTTATTCTCTGATTAAAATTGGAGTTTACATCAGTGGCTCATGCAATAAAGGTTGGTAT
TAAAGGCCAAAACGAATTACTGAAGTGAAATTTCAATTTTGACTGAAGGCACCTAAACTTCCAATGTTGAATGCAA
ATGCAATAATAGAACGCATCATAATATTGCTCTCACTAGTCACTAGTGGAAATAATGAAGTGATTGTGGGGCAAAA
TAGTTTAATTTGAAATTTTGAACTGATAATTGTAACAGGTTCACAAGATTGTTGATCTTTCATCGTCTCCTAACCGAG
AGGCTATGTGGTATCTTGAAGTTATGGAAGCATACAAACTGTTCTATCAGCCTCTTGTTGAAGAATTTACATGATG
GGAGTTCTTGATTAAACAGAGAATGGGTATTGAGGTTGGCATCTGACATATGGATTTTCCCTTCTTTTGACAAGTTT
AAGTTAGCACACAGAATTTTCTAAGATTGAGTTGGTATTCACATATCTATGAGAAATCTGTTCTTTGTTATAAAAAA
ATTCTATGTGCTAGGACTTAGATGAATTTTGATGTATATCATTTCGATGGCAAGAAATATAACAAGGCTAAAATACT
TAGCACGAGTCAAGGATGTCAAAATGGTTATCGAAATAATAAATTTTGATGAGGCTAATGTTTCGATGAGCAAATT
AAGCACATG

Source: Schmutz, J. et al. 2010

Sample D Sequence:

AGCCTCTTTCTCTACAAGTCACAGCGCTGGAGCCTAGAACCCAGTTCTTCCTGGCTTCTCAGACTGTGGACCGGCCT
CCGCAGCGCGCGCCGCAGGTGGCCCACGGCCCGGATCGCTCTCCACCCAAGGGCCCCGCGCTCCCTCCACACCGG
CCATGTCGTCCTACACCCGCGTGGCGCTGGTCACCGGGGCCAACAAGGGGATCGGCTTTGCCATCGCGCGCGACC
TGTGCCGGGAGTTCCCGGGGGACGTGGTGCTCACGGCGCGGGACAAGGCGCGGGGTCGGGCGGCCGTGCAGCA
GCTGCAGGCGGAGGGCCTGAGCCCCCGTTTCCACCAGCTGGACATCGACGACCTGCAGAGCATCCGCGCCTTGCG
CGACTTCCTGCGCAAGGAGTACGGGGGGCTCAACGTGCTGGTCAACAACGCGGGCATCGCCTTCAAGAGTAGGT
GAAGGGGGTGGTGGGAGGTCTCCGGCTCTCCGAGGGCGCCCCGCGACGAATTGCACGCGCCCCCCGCCCCCACC
AGTCCACTGGAGTGATCGGAAAGGCAGCGCCCGCCCCTCGGGATGCGCTCAGCTCTCGCCTCCCACGGCTGGTTC
TCGCTTCCCACGGAGTCTTGGGAGCGGTTCGTTTCGCTCGAGAGCAGCCGCGGTGCCGGGCACAGCGTGCCCCGC
CTGCCCGCCGGCCTGTGCGGGCCTGTGAGCCACAGTGCTCCACCCGGGACCCTCCCAGAACCATCAGCTGCTGAG
TTTTGCAAGAGAGTCGGTCTCCGAGGAGGGGACCACGCCCTGCTAAGCTCATCAGCTTTCTAATGCCAGAAACATT
TGCTCTGCGAAGGGTCTCCAAGCTCTCTTTGGGCTTTCGCTTTCCCTCAAAGGGCGCATTCATAGACACCAGGAGG
CATCTGCTCGGGACAGGAAGTGAGTCTCAGAGACACGGATTAGCGCGCATATGCACGCTAGTGGACTTGGGTCTC
TTTAACGCAGGGCTTTGATAGGCTCTGGCAAAAAGGAAGCGAGTGTGGTTCCCACCCCCAGGGAAGCTCGGCTTG
TTATCAAACATAGACGGTTTGTTCTGGGTTTTTCTTTTTTAAGGGTGGTTTTAAGTAGCATCTTTCTTTGGAGAGGG
GCAGGGAGATGGAGATTACTTGCTCCATGACTCGAATAAGAAAGTTATTGATCTGCCTTGTGCTGCGGGTAGAAA
TAAAAGGCCGGCCTCACTTACAACCTAATGGGAAAGAGAACTCTAAGAGGCCGAGCGGAGGAAAACGAAGCATT
GTGAAGCCTGTCTTCATTTCCCTCGGCCTTGGGCGTTGGACCGCCCGGGGCACCCGTTGGGGCAACTCAGTCTGCC
GGGTTTTCTGCGCGCTCAGGCGGGGAGTTGGCGGACACCCCCTCGGAGCCCGCAGCTGAGGGATGGGTGGACAA
CACATGGGGGGCATACTGACCTCATCTGGCAACCCTCGCCCCTCAGCGTTTGGACCCTCGACCAGCAGTACCTGCG
ACCCTGGGAACTGCGGGAATCGCAGAACGGCCAGCCTCAGACGGACACCTGCTGAACTGTAATCTGCATTTTAGC
AGAATCGCCAGGTGCCTCCTCCTATGCCCTTTCAAGATGGGTTTGACCACTGCTGCTCCAAGTGTAGCTCACAAACC
GGTAGTTGGAGCGTCACCTGGGAACTGAATTAGAAATGCAAATTATGAAGCCCCACCCCGGAACTCCTTAATCCCA
AACTCTGGGGGCAGGGCCCAGAAAGCTGTGTTCTAACAGGCCCTGCGGATGCTTCAGATAGAGGCTGGACGTCT
GAGAACCATCAGCCTCCTGCTTTTAAACATCTTTGCCTGGGCTGGATCTCCGTTGTTGTGCCGGCTTTTCTCTAGCT
GTGGCAAGTCGGGACGGCTGTCTGGTTGTATGCAGGCTTCTCACTGCAGTGGCTTCTCTTGTCGCAGAGCAAGGG
CTCTAGGGTGCTCGGGCTTCAGTAGTTGTGGCACAGGGTCTCAGTAGTTGCGGCTCCAGGGCCCTAGAGCACAGG
CTCAATGGTTGTGGCCCACGGGCTTAGCTGTTCTGCAGCACGTGGGATCTTCCCGGATCAGGGATGGAACACATG
TCCCCTGCATTGGCAGGCAGATTCCTTACCACTGAGCCACCAGGGGAGCCCTTGAAATTGTTTTAGTTACCTTGCTT
TGCTTTTAATCTGAGGTAAGATCCACTTTGTCTTGAGAATTTCAGGGGGGAAATGTGTGACTTTGTTTCCTCCTCTT
TTTGAAAAGCTGATGACCCAACACCCTTTGACATTCAAGCTGAGATGACGCTGAAGACAAACTTCTTTGCCACGAG
AAACGTCTGCACCGAATTACTGCCTATAGTCAAACCTCATGGTGAGTCCAGCTTTGCTGACCGTCAGGTTGTCTCCC
TTAGCAAGGAGTGAGCCACAGGGACCCGTCCCCCA

Source: Zimin, A. et al. 2009

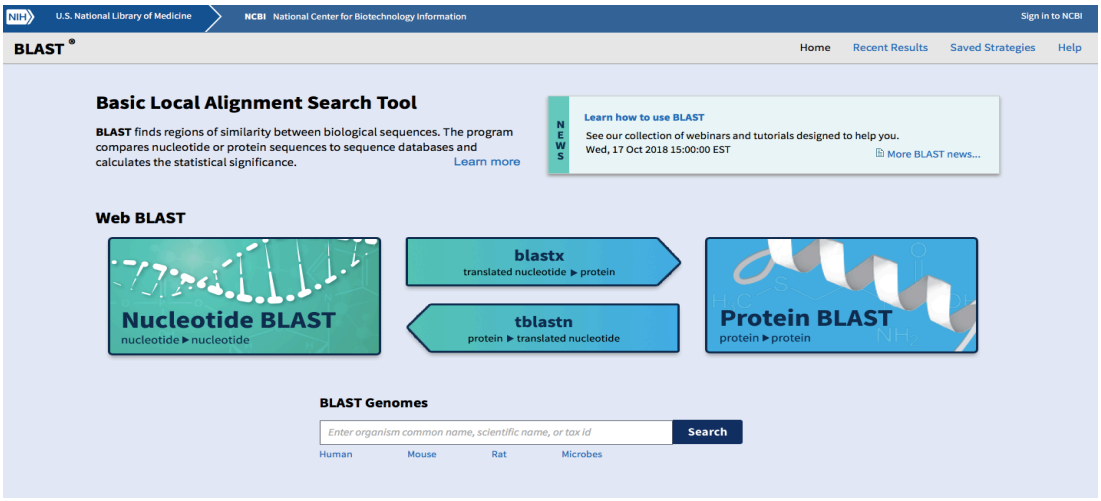
Appendix: 3b **NCBI Website Instructions**

Instructions - How to use the NCBI Website:

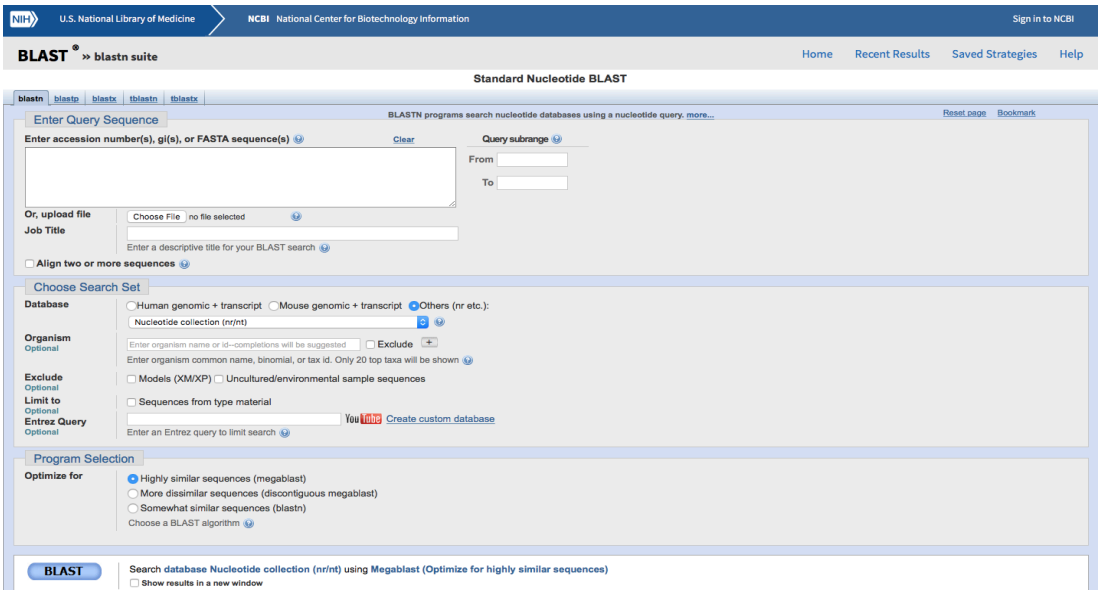
The NCBI website will be used in order to identify unknown sequences (Example with E.coli):

1. Go to this URL for the NCBI Website: https://blast.ncbi.nlm.nih.gov/Blast.cgi

You will see the following page:



Click on the section that says Nucleotide BLAST and it will take you to this page:

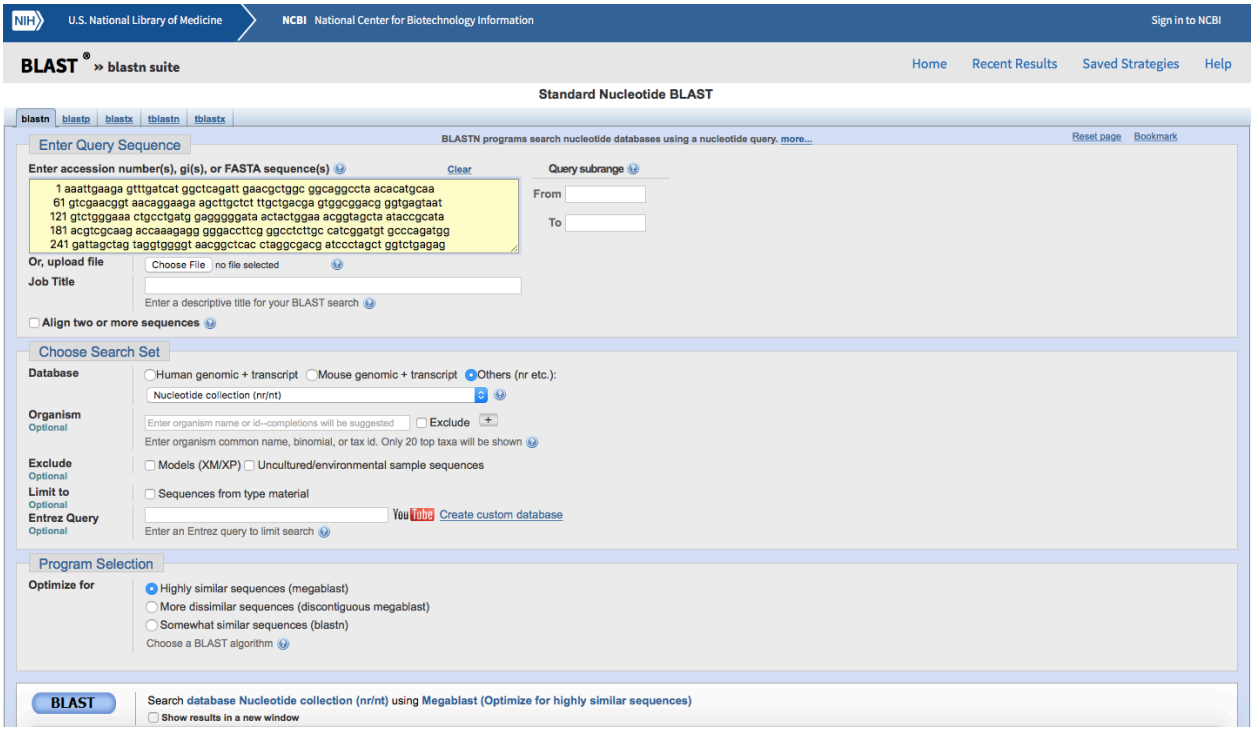


- This allows you to search a sequence for its potential identity
- The other options are for searching with other information such as with proteins

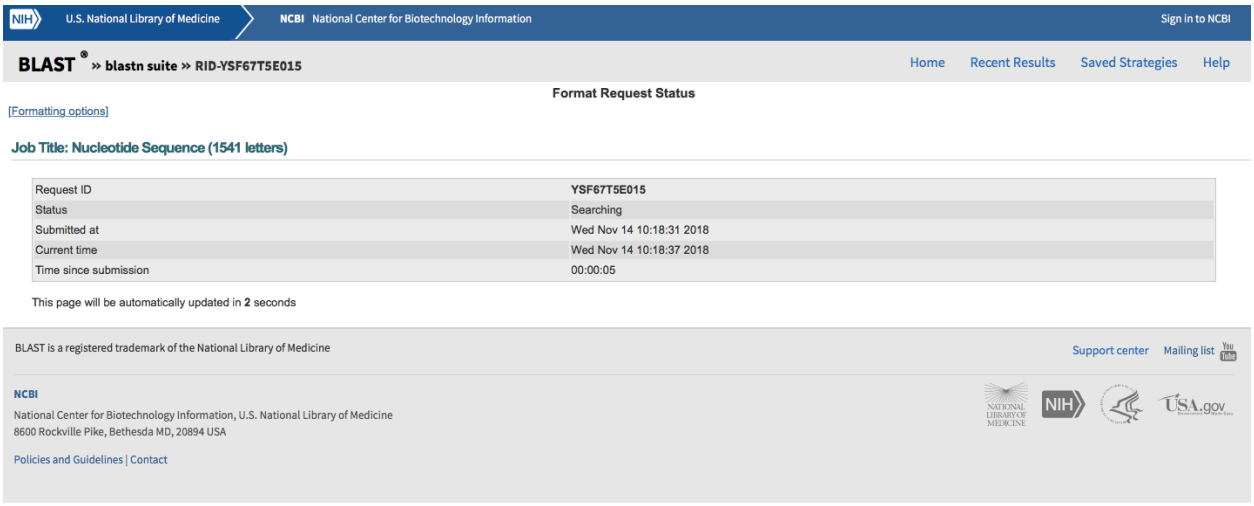
- BLAST stands for Basic Local Alignment Search Tool
    - The BLAST search works by using the sequence that you put in and looks for similarities to sequences within the NCBI database. It outputs a list of potential matches and tells you what the similarity of your unknown sequence is to the ones in the database.
    - If you have a highly conserved match between your unknown and a database sequence, you can identify what your unknown sequence most likely came from.

Once on the BLAST page you can copy and paste the unknown sequence into the text box that says above it "Enter accession number(s), gi(s), or FASTA sequences"

- You can upload a file to this box or enter in sequences or numbers assigned to known sequences in order to identify unknowns

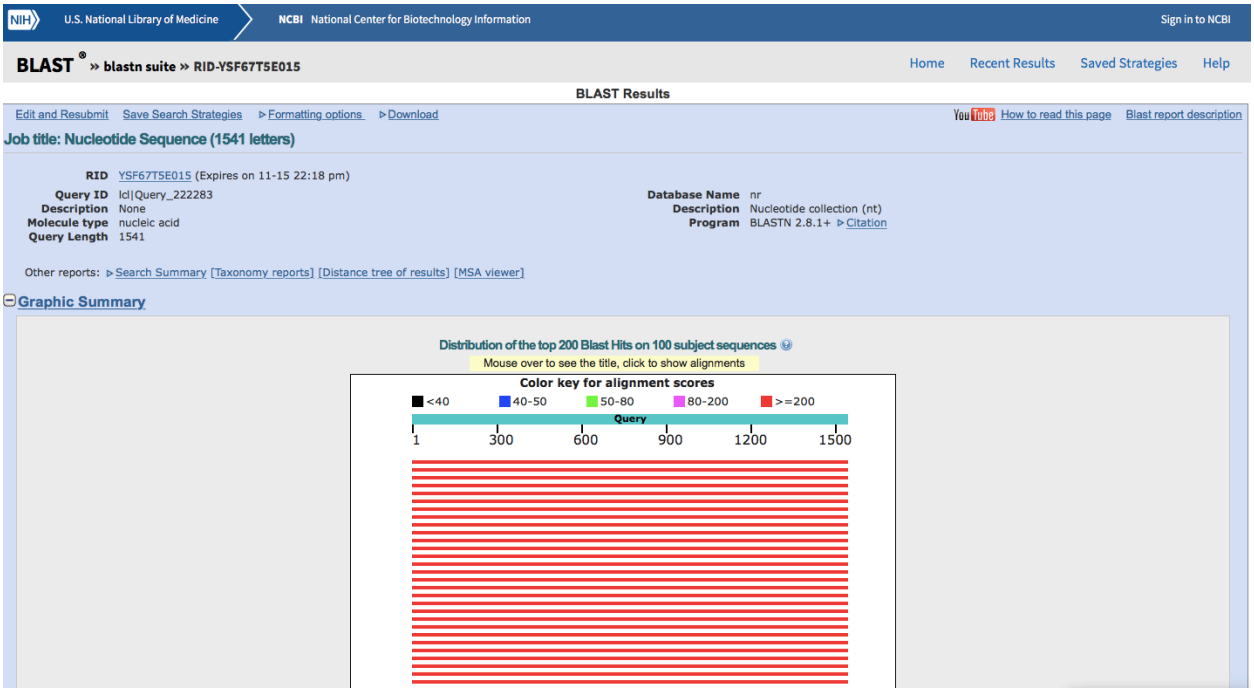


After you have copied and pasted the sequence for your unknown you press the blue, oval "BLAST" button at the bottom of the page and it will take a moment to identify the sequence:



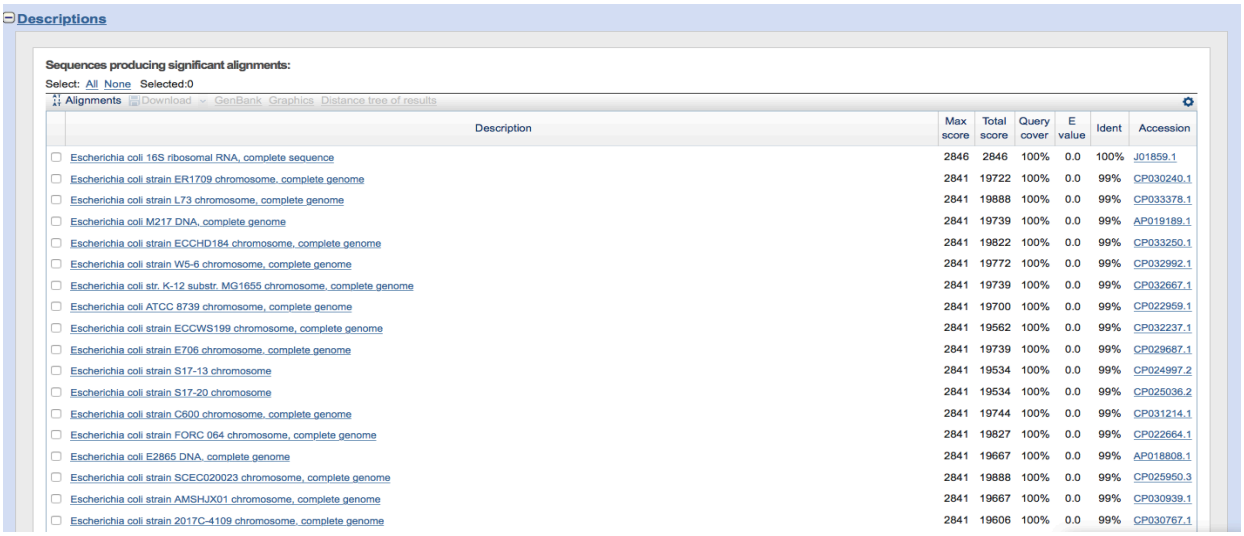
After a few minutes the site will give you an output with the proposed sequence identity. Be patient, some searches can take several minutes (e.g. Sequence C takes ~7 minutes). On the BLAST Results page:

*The first half of the page will show the actual sequence and its alignment scores (shown in color - red in the example) this tells you how much of the sequence matched and by how much based on the alignment score.



*Below the alignment score figure (Graphic Summary), scroll down to the “Descriptions” section. This includes a list of potential identifications, under “Sequences producing significant alignments,” which are the sequences to which your unknown matched. It will list a description and then a max score, a total score, a query cover, an E value, an Ident, and accession. The only thing to be concerned with is the Description, which will tell you what organism the sequence matches to by its scientific name, and the “Ident,” which is the percent identity. Take the first organism in the list because it will have the highest percent identity (Ident). In the example we have used a sequence from E. coli and it is

identified as E.coli with a % identity of 100%. Therefore the students could confidently say that their unknown sequence is a sample of E. coli.



Descriptions

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
Escherichia coli 16S ribosomal RNA, complete sequence	2846	2846	100%	0.0	100%	J01859.1
Escherichia coli strain ER1709 chromosome, complete genome	2841	19722	100%	0.0	99%	CP030240.1
Escherichia coli strain L73 chromosome, complete genome	2841	19888	100%	0.0	99%	CP033378.1
Escherichia coli M217 DNA, complete genome	2841	19739	100%	0.0	99%	AP019189.1
Escherichia coli strain ECCHD184 chromosome, complete genome	2841	19822	100%	0.0	99%	CP033250.1
Escherichia coli strain W5-6 chromosome, complete genome	2841	19772	100%	0.0	99%	CP032992.1
Escherichia coli str. K-12 substr. MG1655 chromosome, complete genome	2841	19739	100%	0.0	99%	CP032667.1
Escherichia coli ATCC 8739 chromosome, complete genome	2841	19700	100%	0.0	99%	CP022959.1
Escherichia coli strain ECCWS199 chromosome, complete genome	2841	19562	100%	0.0	99%	CP032237.1
Escherichia coli strain E706 chromosome, complete genome	2841	19739	100%	0.0	99%	CP029687.1
Escherichia coli strain S17-13 chromosome	2841	19534	100%	0.0	99%	CP024997.2
Escherichia coli strain S17-20 chromosome	2841	19534	100%	0.0	99%	CP025036.2
Escherichia coli strain C600 chromosome, complete genome	2841	19744	100%	0.0	99%	CP031214.1
Escherichia coli strain FORC 064 chromosome, complete genome	2841	19827	100%	0.0	99%	CP022664.1
Escherichia coli E2865 DNA, complete genome	2841	19667	100%	0.0	99%	AP018808.1
Escherichia coli strain SCEC020023 chromosome, complete genome	2841	19888	100%	0.0	99%	CP025950.3
Escherichia coli strain AMSHJX01 chromosome, complete genome	2841	19667	100%	0.0	99%	CP030939.1
Escherichia coli strain 2017C-4109 chromosome, complete genome	2841	19606	100%	0.0	99%	CP030767.1

Appendix 4: **Sequences KEY**

CSI Sequences KEY:

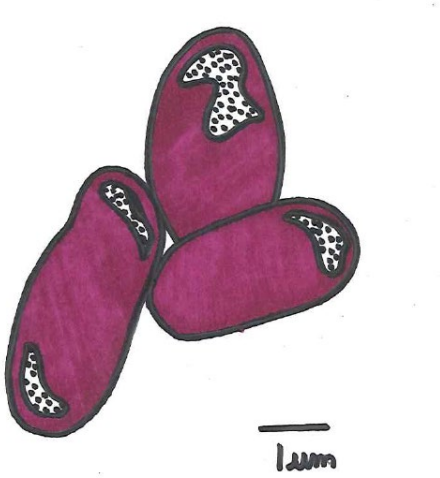
Organism Identification based on NCBI Nucleotide BLAST Search

Sample	Organism	Scientific name
A	Nitrogen-fixing bacteria	*Rhizobium lemnae*
B	Denitrifying bacteria	*Rhodanobacter denitrificans*
C	Soybean, plant	*Glycine max, soja*
D	Cow, mammal	*Bos taurus*

Appendix: 5a **Sample Information Sheet**

Sample A – Information Sheet

Rhizobium lemnae

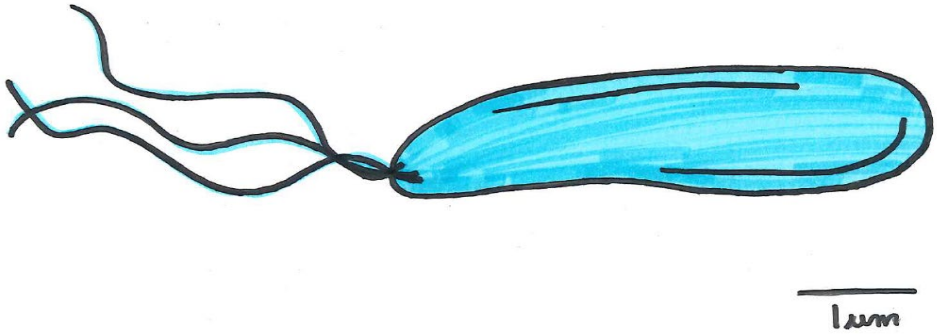


Domain: Bacteria
Phylum: Proteobacteria
Class: Alphaproteobacteria
Order: Rhizobiales
Family: Rhizobiaceae
Genus: Rhizobium
Species: *lemnae*

- Small, single-celled, Prokaryotes
- Nitrogen fixing bacteria
    - The organism can take nitrogen from the air (N_2 gas) and use it to make ammonium (NH_4^+).
    - Ammonium is a nutrient that is needed by many other organisms to live such as plants – all living things require Nitrogen, but in different forms.
- These bacteria often live within plant roots or nodules and can provide the plant with ammonium so that it can grow. The plant can make organic compounds that can be utilized by the bacteria. This is a mutualistic or symbiotic relationship that is common in legumes, alfalfa, and soybeans.
- Importance: Nitrogen fixing bacteria make nutrients available to other organisms such as plants which allows them to grow more than they would if they did not have the nutrients provided by the N-fixing bacteria. This is very important in agriculture in which there many plants in a small area where the soil may not have enough nutrients to support the crops. This is like a natural fertilizer for the plants.
- Fun Fact: The way that we make fertilizers that you can buy in the store is based off these organisms!

Sample B – Information Sheet

Rhodanobacter denitrificans



Domain:	Bacteria
Phylum:	Proteobacteria
Class:	Gammaproteobacteria
Order:	Xanthomonadales
Family:	Xanthomonadaceae
Genus:	Rhodanobacter
Species:	*denitrificans*

- Small, single-celled, prokaryotes
- Denitrifying Bacteria
    - Organism that utilize Nitrate (NO_3^-), a form of nitrogen, and reduce it to form N_2 gas. They do this in order to break down organic matter – similar to respiration which uses oxygen to break down organic matter these bacteria use nitrate instead because they live in anoxic environments – where there is no oxygen available.
- Found in soils as well as freshwater, estuarine, and ocean sediment.
- Importance: Denitrifying bacteria can remove nitrogen, in the form of nitrate, from an ecosystem. Nitrate can be found in soil and sediments, but can also be highly concentrated in wastewater and manure. This can be important because in some cases nutrients, such as nitrogen, can be harmful to an ecosystem in excess. However, in soil systems where nitrate is often required by plants, denitrifying bacteria can compete with plants for available nutrients.
- Fun Fact: Denitrification can deplete soil fertility and therefore decrease crop yields. If denitrification is reducing the available nitrogen in the soil farmers may need to add fertilizer to stimulate plant growth.

Sample C – Information Sheet

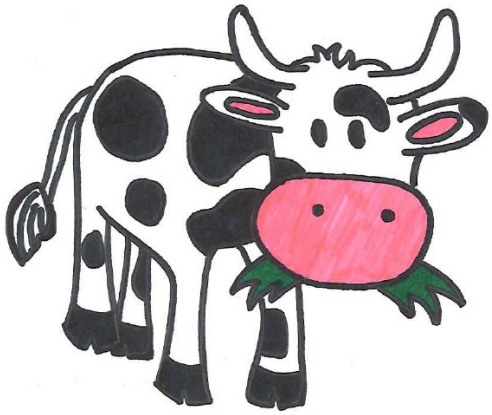
Glycine max

Kingdom: Plantae
Phylum: Tracheophyta
Class: Magnoliopsida
Order: Fabales
Family: Fabaceae
Genus: Glycine
Species: *max*

- Soybean plant
- Green plants that usually grow to about 3 feet tall which have brownish green seed pods.
- Photosynthesize and therefore they require sunlight and nutrients, such as nitrogen, in order to grow.
- *Importance:* Soybeans are a major crop as they are used for animal feed for livestock, the seeds are collected for use in food products and other commercial goods.
- *Fun Fact:* Soybeans are harvested for many different dishes, including the Japanese dish edamame, but also tofu, soy sauce, and cooking oil. It is also used in some cosmetic products as well.

Sample D – Information Sheet

Bos taurus



Kingdom:	Animalia
Phylum:	Chordata
Class:	Mammalia
Order:	Artiodactyla
Family:	Bovidae
Genus:	Bos
Species:	*taurus*

- Cow
- Quadruped mammals which can live to about 20 years of age and weight around 2,000 pounds.
- Cattle are livestock which are commonly raised for meat or dairy products.
- They are herbivores and require large amounts of plant material in order to get the nutrients they need. Cows will eat grass, alfalfa, soybean meal, and other plants. Due to their intake they also produce large amounts of manure which can have high levels of nitrate, a form of nitrogen that acts as a nutrient to plants, for this reason, manure can sometimes be used to make fertilizer for plants.
- Importance: Cattle provide milk and meat as food to people all over the world and they therefore have large impacts on people and agriculture.
- Fun Fact: Cows are ruminants meaning that they have four stomachs in order to digest their food because plant material is difficult to process.

CSI Sequence

Connections Worksheet:

1. List each of the organisms, including yours, that were identified.

2. What do all these organisms have in common?

3. How might these organisms relate to one another – how might they be connected?

4. As a group draw out how you think these organisms may form a cycle.

Appendix 6b: **Sequence Connections KEY**

CSI Sequence - ANSWER KEY

Connections Worksheet:

1. List each of the organisms, including yours, that were identified.

Rhizobium lemnae	**N-fixing bacteria**
Rhodanobacter denitrificans	**Denitrifying bacteria**
Glycine max	**Soybean**
Bos taurus	**Cow**

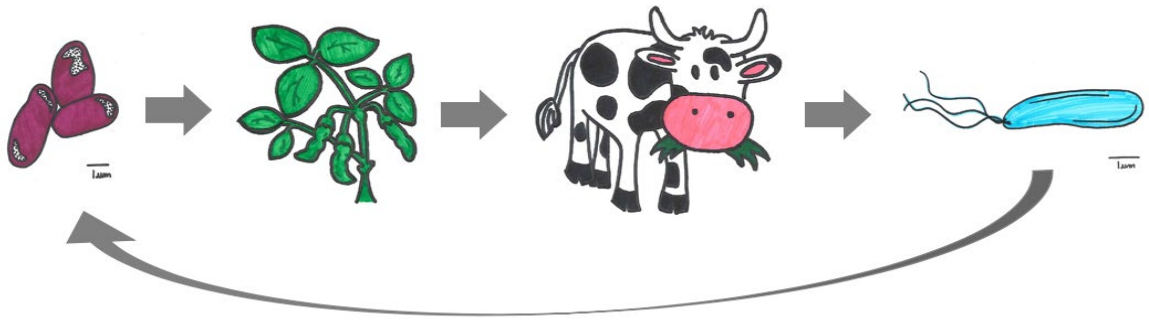
2. What do all these organisms have in common?

They all need nitrogen and they all impact the nitrogen cycle in some way.

3. How might these organisms relate to one another – how might they be connected?

Each organism provides what another one needs in order to live and grow. Some work together, like the plant and N fixing bacteria whereas the cow eats the soybean.

4. As a group draw out the relationships between the identified organisms such as in a cycle.



N Fixing bacteria provide nutrients, specifically nitrogen, to soybeans through symbiotic relationship. Cows eat soybeans and other plants and therefore get their nutrients, including nitrogen, by eating. Cow manure contains high levels of nitrate which can be utilized and removed by denitrifying bacteria. Denitrfying bacteria produce N_2 gas which is utilized by the N fixing bacteria therefore completing the cycle.