# What's lurking in the soil?

| Focus questions | How do researchers determine the identity of organisms from pieces of DNA? |
|---|---|
| Learning target | Students will search the NCBI database to identify a pathogen from a sequence of DNA. |
| Vocabulary | Bioinformatics, barcodes |

## HS-PS4 Waves and their Applications in Technologies for Information Transfer

| Performance expectation HS-PS4-2 | Classroom connection: Students use the NCBI database, a digital database, to identify an organism by DNA sequence analysis. |
|---|---|

## Science and engineering practices

| Asking questions and defining problems | Classroom connection: Students are given an unknown DNA sequence and must use bioinformatic tools to identify what organism it is from, explain why it might be found in the sample and recommend management, if necessary, for the organism. |
|---|---|

## Disciplinary core ideas

| PS4.A: Wave Properties<br><br>LS4.A Evidence of common ancestry and diversity | Classroom connection: Closely-related species show differences in specific portions of DNA; comparison using a database of digitized information can help identify the differences much faster than looking at sequences one by one. |
|---|---|

## Cross-cutting concepts

| Stability and Change<br><br>Patterns | Classroom connection: Influence of Engineering, Technology, and Science on Society and the Natural World Classroom connection: Patterns in sequences of DNA are identified through the use of the database. |
|---|---|

# Background

This lesson illustrates what scientists do after they have completed DNA extractions by introducing basic DNA sequence analysis. Importantly, this lesson goes beyond the visual appeal and celebrity status of the commonly-deployed strawberry DNA extraction by introducing students to DNA analyses through the use of bioinformatics.

**Bioinformatics** is a combination of computer science and biology that aims to organize, analyze, and interpret biological data. With the success of biotechnology in DNA sequencing, biologists need to use computer science to handle such large, rich databases from organisms being studied. DNA is both a unifying and defining characteristic of all living things. Even in its most raw form, the sequence of As, Ts, Cs, and Gs of DNA is a powerful tool of discovery. Bioinformatics can provide new insights in biology. For example, comparisons of DNA sequences across living organisms have helped biologists describe new species, discover new genes, and predict the health of populations. In agriculture, decoding DNA has boosted our food supply by identifying, tracking, and incorporating genes that provide protection from pathogens and pests. Likewise, cracking the DNA code has helped classify and better detect emerging pathogens that cause disease and substantial crop losses.

Disease accounts for up to 20% of the world's annual food losses, making pathogens a top threat to food crops. Many pathogens inhabit agricultural fields and attack food crops during the growing season. Soil contains a diversity of microbes, which may include fungi, oomycetes, animals, protists, and bacteria. The community of microbes can be examined by extracting DNA directly from soil and decaying plant matter (residue) found within a farmer's field. Importantly, not all microbes in soil are pathogens; some are beneficial to growing plants (i.e., nitrogen-fixing bacteria) or help to decompose dead plants (saprophytic bacteria). Following DNA extractions, there are DNA sequencing techniques that can be used to survey some of the more common groups of microbes, including bacteria and fungi. The sequencing techniques examine regions of the genome that are generally different among closely-related species. These regions are called **barcodes,** defined similarly to the barcodes used in supermarkets to identify unique products with a scanner at checkout.
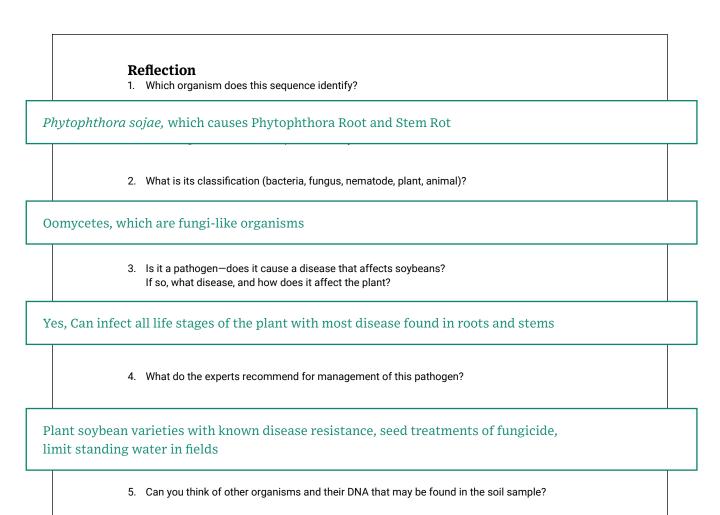
The DNA barcodes are different between eukaryotes (fungi) and prokaryotes (bacteria). In the scenario below, DNA sequences of the barcoding region were used to survey the fungal and bacterial diversity of a field that has given a farmer trouble when soybeans are grown. Specifically, soybean plants continue to produce low seed counts (also termed 'low yield') with noticeable symptoms for disease on the plants (stems and leaves). Multiple applications of a recommended fungicide treatment have not slowed the spread of disease, and more soybean plants are beginning to show signs of disease.

Generally, the sequence of determining an organism is: 1) soil sample collection, 2) DNA extraction, 3) amplification via PCR (targeted or barcode markers), 4) sequencing (barcode markers) or genotyping (targeted markers), 5) analysis. Essentially, this lesson is an introduction to the popular, public database managed by NCBI that looks at sequence analysis, the last operation listed above. The DNA sequence given below can be compared to the database by using the BLAST function. The results from the BLAST will give alignments (i.e., side-by side-comparisons) of the sequence you submit (query) and hits (subject) found in the NCBI database. By clicking on the top hits, the DNA sequences submitted can be identified by species, gene, or genetic marker. Each DNA sequence in the database has its own page that gives additional information, such as the scientists that uploaded the sequences to the database. There are other parameters displayed following a BLAST, which could be further explored (see NCBI presentation deck). Students could be asked to write up a short essay of one of the pathogens and how it could be managed to limit its impact on food crops.

# Teacher preparation

1. Review the included presentation to give background to students.
   You can find sequences at **nourishthefuture.org/go/sequences**.
2. Visit **ncbi.nlm.nih.gov**.
3. Click on 'BLAST' in the right column.
4. Click on 'Nucleotide blast'; paste the DNA sequence in the Enter Query Sequence rectangle;
   click BLAST below to submit; wait for result.
5. Copy handouts for students. Load the sequences into a LMS program for them to be able to cut
   and paste from the document.
6. Have students read the scenario:
   *Shadybrook Farm is having trouble with production of soybeans and has noticed increasing
   levels of disease, despite the use of fungicides. The farm manager elects for a soil and residue
   test from a diagnostic laboratory to identify potential disease-causing pathogens. The laboratory
   produced a report that contained the top five most common DNA sequences from the soil.
   Your job as a bioinformatician is to help the farm manager identify the species, using the given
   sequences of DNA, through the use of public databases and basic bioinformatic tools.*
7. Determine the way you would like to organize the class to solve the problem.
   - Different groups of students could be given different sequences, then collaborate
     with others to determine what might be causing the problem for the farmer.
   - Each group could get all samples to identify, then determine the problem.

# Student handout

## Sequence 1

**Reflection**

1. Which organism does this sequence identify?

*Phytophthora sojae,* which causes Phytophthora Root and Stem Rot

2. What is its classification (bacteria, fungus, nematode, plant, animal)?

Oomycetes, which are fungi-like organisms

3. Is it a pathogen—does it cause a disease that affects soybeans?
   If so, what disease, and how does it affect the plant?

Yes, Can infect all life stages of the plant with most disease found in roots and stems

4. What do the experts recommend for management of this pathogen?

Plant soybean varieties with known disease resistance, seed treatments of fungicide, limit standing water in fields

5. Can you think of other organisms and their DNA that may be found in the soil sample?

Soybeans, other crops (corn), nitrogen-fixing bacteria, saprophytes (i.e., decomposers)

# Student handout

## Sequence 2

**Reflection**

1. Which organism does this sequence identify?

*Soybean Cyst Nematode*

2. What is its classification (bacteria, fungus, nematode, plant, animal)?

Nematode

3. Is it a pathogen—does it cause a disease that affects soybeans?
   If so, what disease, and how does it affect the plant?

Yes; penetrates soybean roots and causes cysts on the plant root system.

4. What do the experts recommend for management of this pathogen?

Maintain plant health and reduce competition; keeping SCN-infested soils and residues OUT of non-infested fields; plant resistant soybeans; rotate to non-host crops; nematicides may be used, but they are not long-lasting.

5. Can you think of other organisms and their DNA that may be found in the soil sample?

Soybeans, other crops (corn), nitrogen-fixing bacteria, saprophytes (i.e., decomposers)

## Sequence 3

**Reflection**
1. Which organism does this sequence identify?

*Glycine max:* soybean

2. What is its classification (bacteria, fungus, nematode, plant, animal)?

Plant

3. Is it a pathogen—does it cause a disease that affects soybeans?
   If so, what disease, and how does it affect the plant?

No, it is what we want to grow in the soil

4. What do the experts recommend for management of this pathogen?

n/a

5. Can you think of other organisms and their DNA that may be found in the soil sample?

Other crops (corn), nitrogen-fixing bacteria, saprophytes (i.e., decomposers)

## Sequence 4

**Reflection**

1. Which organism does this sequence identify?

*Bos taurus:* domestic cattle

2. What is its classification (bacteria, fungus, nematode, plant, animal)?

Animal

3. Is it a pathogen—does it cause a disease that affects soybeans?
   If so, what disease, and how does it affect the plant?

No, it may have grazed there the year before, or manure from these cattle could have been spread there.

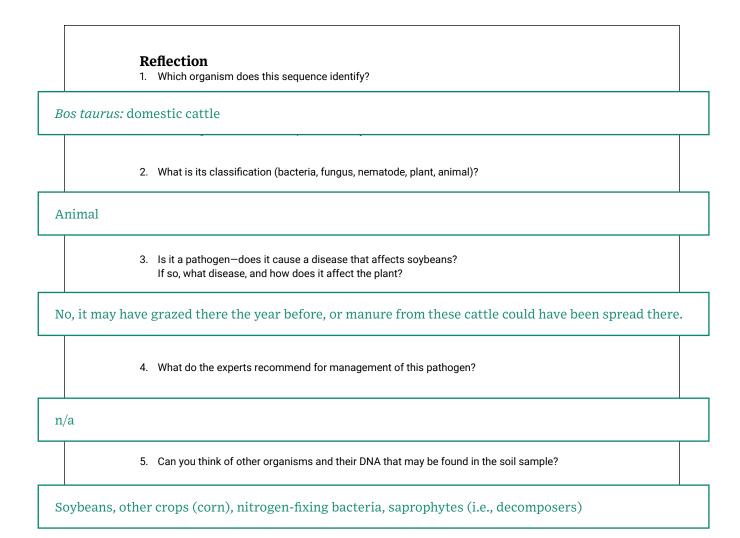4. What do the experts recommend for management of this pathogen?

n/a

5. Can you think of other organisms and their DNA that may be found in the soil sample?

Soybeans, other crops (corn), nitrogen-fixing bacteria, saprophytes (i.e., decomposers)
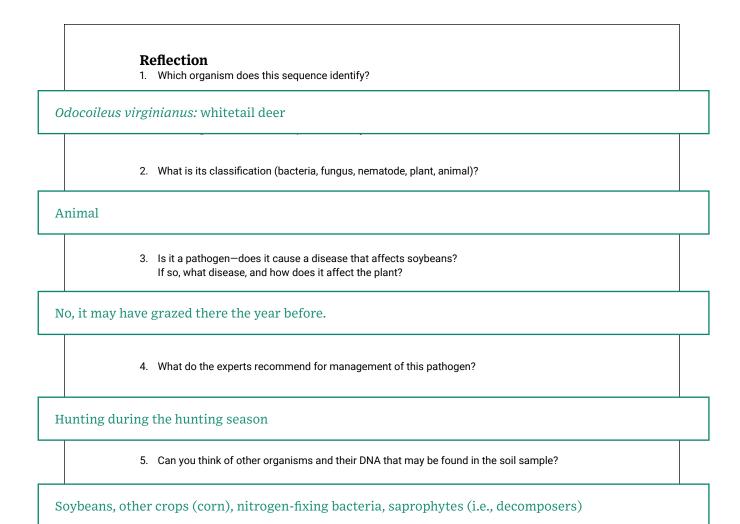
# Student handout

## Sequence 5

**Reflection**

1. Which organism does this sequence identify?

*Odocoileus virginianus:* whitetail deer

2. What is its classification (bacteria, fungus, nematode, plant, animal)?

Animal

3. Is it a pathogen—does it cause a disease that affects soybeans?
   If so, what disease, and how does it affect the plant?

No, it may have grazed there the year before.

4. What do the experts recommend for management of this pathogen?

Hunting during the hunting season

5. Can you think of other organisms and their DNA that may be found in the soil sample?

Soybeans, other crops (corn), nitrogen-fixing bacteria, saprophytes (i.e., decomposers)

# Differentiation

- **Local community:** Students may visit a local biotech lab or look for a virtual tour of a biotech lab where materials undergo PCR and sequencing.
- **Students with special needs (language/reading/auditory/visual):** Students will need to have access to a digital platform in order to copy the sequences to be used in the search. Type sizes can be increased to make it easier for visually-impaired students.
- **Extra support:** Students might work in groups where, for example, one member develops research about the organisms that are identified while others are responsible for identifying the organisms.
- **Extensions:** Additional questions for students:
  a. Click on one of the results: What particular region of the genome do all of these DNA sequences belong to?
  b. Why does the laboratory focus on this particular region of the genome?
  c. What is the Sequence ID number for the exact match in NCBI BLAST?
  d. Who uploaded the sequence to NCBI? (Click on the sequence ID)
- Students may want to find DNA sequences for other organisms of interest. Have students read: `apsnet.org/edcenter/disandpath/oomycete/pdlessons/Pages/PhytophthoraSojae.aspx`
- Additional reference:
  - Higher animals use the COI gene on the mitochondria for barcoding. More information may be found here: `ibol.org/about/dna-barcoding/`
  - Scroll down 'The DNA Barcode' section.
  - Search by genus and species name in the public database to find other organisms you may want to include.

# Assessments

## Rubric for assessment

| Skill | Developing | Satisfactory | Exemplary |
|---|---|---|---|
| Determine advantages to using the digital storage and computing power of NCBI database. | Student sees little to no advantage to using the database. | Student is able to identify and articulate two advantages of using the NCBI database, and successfully identify the sequence(s) assigned. | Student is able to identify and articulate two advantages of using the NCBI database, successfully identify the sequence(s) assigned, and uses the database to find additional sequences of interest. |

## Rubric for self-assessment

| Skill | Yes | No | Unsure |
|---|---|---|---|
| I see the advantage of using the NCBI database to identify organisms. | | | |
| I am able to identify an organism by using the NCBI database. | | | |