WATER QUALITY

Water bioinformatics

Focus questions	How might we determine the organisms that impact water using biotechnology? How can we identify individual organisms in a complex mixture of DNA? What computer tools are needed to decode the source of a DNA sequence?
Learning target	Students will search the NCBI database to identify organisms from a sequence of DNA.
Vocabulary	Bioinformatics, water quality, National Center for Biotechnology Information (NCBI), Basic Local Alignment Search Tool (BLAST), eDNA sequencing, barcodes, eukaryotes, prokaryotes, harmful algal blooms (HABs), taxon, rbcL gene, COI gene

HS-LS1-1 From Molecules to Organisms: Structures and Processes
HS-PS4 Waves and their Applications in Technologies for Information Transfer

Performance expectation HS-LS1-1	**Classroom connection:** Students use DNA sequence analysis taken from different cells to identify an organism.
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

Constructing explanations and designing solutions	**Classroom connection:** Students use the identified organisms to determine soil quality or pathogen presence.
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

LS1.A Structure and Function	**Classroom connection:** Students discover the region in which the DNA sequence is located within the organism and how the regions differ between prokaryotes and eukaryotes and plants and animals.

PS4.A: Wave Properties **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

Structure and Function	**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. **Classroom connection:** Different cellular structures contain DNA sequences that help to identify organisms.
Stability and Change	**Classroom connection:** DNA sequences determine the production of proteins that are useful to the organism.
Patterns **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 using bioinformatics. **Bioinformatics** is a combination of computer science and biology that aims to organize, analyze, and interpret biological data. With the success of other biotechnology tools used in eDNA sequencing, biologists need help from computer science to handle such large, rich databases that contain 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.

Generally, the order of determining an organism through DNA analysis includes: 1) water sample collection, 2) DNA extraction, 3) amplification of barcode region with PCR, 4) sequencing, and 5) analysis using bioinformatics.

Water ecosystems contain many organisms, both macroscopic (visible with the naked eye) and microscopic. DNA analyses help scientists understand the structure of these complex environments. For example, DNA analyses can find:
1) classification group or **taxon** linked to a healthy water system,
2) potential threats to human health,
3) endangered species, or
4) invaders that compete for resources with native species.

The community of organisms within lakes and rivers can be examined by extracting DNA directly from a water sample. This special type of DNA sample is called an **environmental DNA (eDNA)** sample. The resulting eDNA sample contains DNA from all cells collected within the water sample, including whole organisms or decaying tissue. The eDNA sequencing technique looks at genes or non-coding regions of the genome that are universal to life, but each species has a unique sequence signature. These regions are called **barcodes**, analogous to the unique barcodes that label products scanned at supermarket checkouts. The DNA barcode regions are different between

eukaryotes and **prokaryotes**. Within these major groups there are subcategories of ideal barcodes. For example, identification using the **rbcL** gene which encodes the large subunit of ribulose bisphosphate carboxylase, commonly referred to as rubisco, a key enzyme in photosynthesis, is ideal for plants, whereas the **COI** gene which codes for cytochrome C protein which is critical in ATP synthesis, is best for animals.

This activity is an introduction to eDNA analysis using bioinformatics within the public database managed by the **National Center for Biotechnology Information (NCBI)**. The DNA sequences given below are compared to those submitted to the database by using the **Basic Local Alignment Search Tool** (**BLAST)** function. The results from the BLAST will give alignments (i.e., side-by-side comparisons) of the sequence submitted (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.

Materials

- Sequences of DNA (**ntf.plus/watersequences**)
- Computer

Prior knowledge

In order to successfully complete this activity, students should know:

- the role of DNA in protein synthesis
- how genes code for proteins and enzymes
- how to cut and paste long sequences of letters into a database and other basic computer searching skills
- water quality parameters on judging healthy ecosystems

Suggested timing

45 min–90 min

Students should be able to complete the identification of their sequence within a single class period. If students are asked to identify the impact of their organism on water quality in terms of threat or health, additional time may be needed.

Teacher preparation

1. Review the included presentation to give background to students.
   You can find sequences at **ntf.plus/watersequences**.
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. Determine the way you would like to organize the class to solve the problem. There are 10 sequences, each impacting water quality in different ways. See the table below for potential categories.
   - Different groups of students could be given different sequences then collaborate with others to determine if there are potential threats.
   - Each group could get all samples to identify then determine threats.

Procedure

1. Show the presentation. Explain that the database is looking for differences between the organisms by comparing the regions in the rcbL gene for plants and the COI gene for animals.
2. Have students read the scenario:
   Ohioland Water Institute and Buckeye Genetic Laboratories have an ongoing water-monitoring project in Lake County. Last week, a summer field technician collected 100mL of water from Lake Erie for eDNA analysis. Buckeye Genetic Laboratories performed eDNA sequencing and released the report containing the top 10 sequences recovered from the water sample. Your job as a bioinformatician is to help identify the source of DNA by searching for similar sequences within the NCBI database using BLAST.
3. Assign groups sequence(s) to investigate. Ask them to determine the impact the given organism or organisms has/have on water quality.

Seq	Full description	Accession num	Common name
1	*Dreissena polymorpha* isolate Drpo4697 cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondrial	EF414493.1	Zebra mussel
2	*Lemna minor* ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, partial cds; chloroplast	KC584882.1	Duckweed
3	*Corydalus cornutus* cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial	EU839725.1	Eastern dobsonfly
4	*Sander vitreus* isolate ST675 mitochondrion, complete genome	OR552091.1	Walleye
5	*Escherichia sp.* strain Esraa 4 16S ribosomal RNA gene, partial sequence	MT647245.1	Intestinal bacteria
6	*Trichobilharzia stagnicolae* isolate atsr_43 cytochrome c oxidase subunit I (COI) gene, partial cds; mitochondrial	MK433252.1	Parasitic flatworm of birds
7	*Fragilaria capucina* clone TCC547 ribulose-1,5-bisphosphate carboxylase/ oxygenase large subunit (rbcL) gene, partial cds; chloroplast	KC736594.1	Diatom
8	*Microcystis aeruginosa* strain NIES-843 16S ribosomal RNA, complete sequence	NR_074314.1	Freshwater cyanobacteria
9	*Cyprinus carpio* 'Guilin' isolate Guilin mitochondrion, complete genome	MK291479.1	Common carp
10	*Anabaena flos-aquae* 0tu33s15 partial rbcL gene for RuBisCo large subunit and rbcX gene, strain 0tu33s15	AJ632023.1	Toxin producing cyanobacteria

Student handout

Sequence 1

Reflection

1. Which organism does this sequence identify?

Dreissena polymorpha (zebra mussel)

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

Animal, mollusk

3. How does your identified organism impact freshwater systems?

Invasive, filter feeder

4. If the organism is a threat, are there ways to manage it to reduce its threat? Explain.

None have been found

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

Answers will vary. For example: other indicators of good water quality (dragonfly larvae, mayflies, caddisflies), other invasive species, other species that produce toxins that are harmful to humans, other parasites, livestock DNA, etc.

Student handout

Sequence 2

Reflection

1. Which organism does this sequence identify?

Lemna minor (duckweed)

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

Plant

3. How does your identified organism impact freshwater systems?

Common, but can overgrow freshwater systems quickly under the right conditions

4. If the organism is a threat, are there ways to manage it to reduce its threat? Explain.

Physical removal

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

Answers will vary. For example: other indicators of good water quality (dragonfly larvae, mayflies, caddisflies), other invasive species, other species that produce toxins that are harmful to humans, other parasites, livestock DNA, etc.

Student handout

Sequence 3

Reflection

1. Which organism does this sequence identify?

Corydalus cornutus (Eastern dobsonfly)

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

Animal, insect

3. How does your identified organism impact freshwater systems?

Indicator of excellent water quality

4. If the organism is a threat, are there ways to manage it to reduce its threat? Explain.

Not applicable

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

Answers will vary. For example: other indicators of good water quality (dragonfly larvae, mayflies, caddisflies), other invasive species, other species that produce toxins that are harmful to humans, other parasites, livestock DNA, etc.

Student handout

Sequence 4

Reflection

1. Which organism does this sequence identify?

Sander vitreus (walleye)

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

Animal, fish

3. How does your identified organism impact freshwater systems?

Freshwater perciform fish native to most of Canada and to the Northern United States

4. If the organism is a threat, are there ways to manage it to reduce its threat? Explain.

Not applicable

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

Answers will vary. For example: other indicators of good water quality (dragonfly larvae, mayflies, caddisflies), other invasive species, other species that produce toxins that are harmful to humans, other parasites, livestock DNA, etc.

Student handout

Sequence 5

Reflection

1. Which organism does this sequence identify?

Escherichia sp. (intestinal bacteria)

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

Bacteria

3. How does your identified organism impact freshwater systems?

Indicate that the water may be contaminated with human or animal wastes

4. If the organism is a threat, are there ways to manage it to reduce its threat? Explain.

Remove source of contamination

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

Answers will vary. For example: other indicators of good water quality (dragonfly larvae, mayflies, caddisflies), other invasive species, other species that produce toxins that are harmful to humans, other parasites, livestock DNA, etc.

Student handout

Sequence 6

Reflection

1. Which organism does this sequence identify?

Trichobilharzia stagnicolae (parasitic flatworm from birds)

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

Flatworm

3. How does your identified organism impact freshwater systems?

Causes swimmer's itch

4. If the organism is a threat, are there ways to manage it to reduce its threat? Explain.

Removal of bird species

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

Answers will vary. For example: other indicators of good water quality (dragonfly larvae, mayflies, caddisflies), other invasive species, other species that produce toxins that are harmful to humans, other parasites, livestock DNA, etc.

Student handout

Sequence 7

Reflection

1. Which organism does this sequence identify?

Fragilaria capucina (diatom)

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

A type of plankton or microalgae

3. How does your identified organism impact freshwater systems?

Highly resistant to pollution and can often be found thriving in eutrophic conditions that indicate high nutrient levels

4. If the organism is a threat, are there ways to manage it to reduce its threat? Explain.

Reduce nutrient loads in freshwater

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

Answers will vary. For example: other indicators of good water quality (dragonfly larvae, mayflies, caddisflies), other invasive species, other species that produce toxins that are harmful to humans, other parasites, livestock DNA, etc.

Student handout

Sequence 8

Reflection

1. Which organism does this sequence identify?

Microcystis aeruginosa (blue green algae)

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

Cyanobacteria

3. How does your identified organism impact freshwater systems?

May produce a toxin in high nutrient conditions that makes drinking water dangerous to drink for animals and humans

4. If the organism is a threat, are there ways to manage it to reduce its threat? Explain.

Ozone in water will kill the cyanobacteria; reduce nutrient loads in freshwater

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

Answers will vary. For example: other indicators of good water quality (dragonfly larvae, mayflies, caddisflies), other invasive species, other species that produce toxins that are harmful to humans, other parasites, livestock DNA, etc.

Student handout

Sequence 9

Reflection

1. Which organism does this sequence identify?

Cyprinus carpio (common or Asian carp)

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

Animal, fish

3. How does your identified organism impact freshwater systems?

Invasive; voracious feeders that choke out native fishes

4. If the organism is a threat, are there ways to manage it to reduce its threat? Explain.

Attempt to harvest; try to keep them out of freshwater systems
See: **youtu.be/-A5xoChBtZk**

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

Answers will vary. For example: other indicators of good water quality (dragonfly larvae, mayflies, caddisflies), other invasive species, other species that produce toxins that are harmful to humans, other parasites, livestock DNA, etc.

Student handout

Sequence 10

Reflection

1. Which organism does this sequence identify?

Anabaena flos-aquae (blue green algae)

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

Cyanobacteria

3. How does your identified organism impact freshwater systems?

May produce a toxin in high nutrient conditions that makes drinking water dangerous to drink for animals and humans

4. If the organism is a threat, are there ways to manage it to reduce its threat? Explain.

Ozone in water will kill the cyanobacteria; reduce nutrient loads in freshwater

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

Answers will vary. For example: other indicators of good water quality (dragonfly larvae, mayflies, caddisflies), other invasive species, other species that produce toxins that are harmful to humans, other parasites, livestock DNA, etc.

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:**
  - Higher animals use the COI gene on the mitochondria for barcoding. More information may be found here: **ibol.org/about/dna-barcoding/**
  - Students may want to find DNA sequences for other organisms that impact water quality.
  - Have students visit **epa.gov/cyanohabs/learn-about-cyanobacteria-and-cyanotoxins** to learn more about HABs or refer to texts of environmental science to determine threats to native species and dangers to human health.
  - Additional questions for students:
    - Click on one of the BLAST results: What particular region of the genome do all of these DNA sequences belong to?
    - Why does the laboratory focus on this particular region of the genome?
    - What is the Sequence ID number for the exact match in NCBI BLAST?
    - Who uploaded the sequence to NCBI? (Click on the sequence ID)
  - Students could be asked to write up a short research brief of one of the identified organisms and how it impacts water quality or ecosystem.

Assessments

Rubric for assessment

Skill	Developing	Satisfactory	Exemplary
Understanding barcodes	Student can describe a barcode, but not how it is comparable between organisms.	Student can describe a barcode and how it is comparable between organisms due to the genetic differences between them.	Student can describe a barcode and how it is comparable between organisms due to the genetic differences between them, and can suggest another gene that might be used to compare organisms.
Use of NCBI database to answer questions and define solutions.	Student can paste a sequence in the database to get a result.	Student can paste a sequence in the database to get a result and show evidence of the impact the organism has on water quality.	Student can paste a sequence in the database to get a result, show evidence of the impact the organism has on water quality, and suggest additional organisms that may be found in a sample from a different water source.

Rubric for self-assessment

Skill	Yes	No	Unsure
I can explain the use of barcodes to compare organisms.			
I can use the NCBI database to identify organisms based on DNA sequences.			
I can describe the impact of various water organisms on water quality.			