

BLAST the invader: Solving respiratory illness mysteries

How do medical scientists identify the causal agent of respiratory disease? What tools allow medical researchers to decode a nucleotide sequence and match it to a specific pathogen?

This lesson is an introduction to bioinformatics used in medical diagnostics to identify an isolated respiratory pathogen from a patient sample. Bioinformatics, which combines biology and computer science, plays a vital role in identifying the genetic signatures of bacteria, fungi, and viruses that cause human illnesses. By analyzing the nucleotide sequence of an unknown pathogen, scientists can determine the cause and help the medical team decide on the best treatment options for the detected pathogen.

Ohio standards

Anatomy and Physiology

- AP.T.3 Lymphatic and immune system

Biology

- B.C.2 Cellular processes

Student prior knowledge

Students should be familiar with DNA structure and nucleotides, and how sequences match to specific organisms.

Suggested timeline

One class period

Materials

- A computer with internet access
- Nucleotide sequence

Teacher preparation

1. Access sequences at ols.plus/blast-sequences and upload to your learning management system for ease of use by students.
2. Copy the student handout.

Procedure

1. Describe bioinformatics as a scientific field of study and briefly discuss the history of the NCBI database: ncbi.nlm.nih.gov/books/NBK148949.
2. Set up students for success by walking through a short exercise. Ask students to follow the instructions on their handout and use this sequence:

>Demo

ATGGCAAACCTAGATAAAAACCAATTAATTGAATCGTTGAAGGAAATGACCATCATGGAAATCGATG
AAATCATTAAAGGCTGTAGAAGAAGCTTTTGGAGTATCGGCAACACCTGTAGTA

3. All students should arrive at this page of results:

The screenshot shows the NCBI BLAST results page for query RID-YR6N2UD8014. The search parameters are: Job Title: Demo, RID: YR6N2UD8014, Program: BLASTN, Database: core_nt, Query ID: lcl|Query_3502883, Description: Demo, Molecule type: dna, Query Length: 120. The filter results section shows: Organism: only top 20 will appear, Percent Identity: [] to [], E value: [] to [], Query Coverage: [] to []. The table of sequences producing significant alignments is as follows:

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Mycoplasma pneumoniae strain 10-980 chromosome, complete genome	Mycoplasma pneumoniae	222	222	100%	8e-54	100.00%	816424	CP039790.1
Mycoplasma pneumoniae strain 87768.fasta chromosome, complete genome	Mycoplasma pneumoniae	222	222	100%	8e-54	100.00%	816502	CP178987.1
Mycoplasma pneumoniae strain 88591.fasta chromosome, complete genome	Mycoplasma pneumoniae	222	222	100%	8e-54	100.00%	857800	CP179061.1
Mycoplasma pneumoniae strain 88841.fasta chromosome, complete genome	Mycoplasma pneumoniae	222	222	100%	8e-54	100.00%	816470	CP179081.1
Mycoplasma pneumoniae strain 519 chromosome, complete genome	Mycoplasma pneumoniae	222	222	100%	8e-54	100.00%	818735	CP017339.1
Mycoplasma pneumoniae strain FL1 chromosome, complete genome	Mycoplasma pneumoniae	222	222	100%	8e-54	100.00%	816637	CP017333.1
Mycoplasma pneumoniae strain 87428.fasta chromosome	Mycoplasma pneumoniae	222	222	100%	8e-54	100.00%	816567	CP178947.1
Mycoplasma pneumoniae strain 88065.fasta chromosome	Mycoplasma pneumoniae	222	222	100%	8e-54	100.00%	857808	CP179021.1

4. Assign students to search for additional pathogens on their own. Assign a specific sequence to each group. Have them present on questions 6–8 on the student handout.

Answer key to sequences:

- **Seq 1:** AB245475 Respiratory syncytial virus N gene for nucleoprotein, complete cds, subgroup B, strain: RSV2177
- **Seq 2:** CY222254 Influenza B virus (B/Ohio/05/2017) nucleoprotein (NP) gene, complete cds
- **Seq 3:** NR_025951 Bordetella pertussis strain 18323 16S ribosomal RNA, partial sequence
- **Seq 4:** NR_028665 Streptococcus pneumoniae strain ATCC 33400 16S ribosomal RNA, partial sequence
- **Seq 5:** NR_041742 Legionella pneumophila subsp. pascullei strain U8W 16S ribosomal RNA, partial sequence
- **Seq 6:** NR_149341 Histoplasma capsulatum var. capsulatum CBS 136.72 ITS region; from TYPE material

Suggested wrap-up

Lead a discussion to have students reflect on the following:

- What is the value of this database?
- How important is it to correctly diagnose a pathogen?
- Name the careers that would be involved in this type of diagnosis?

Differentiation

- Pair students to help with learning how to use the database and to complete the research. Assign roles for students; one to do the BLAST research while the other completes the pathogen treatment and extent of disease research.
- Provide sequences digitally in various-sized fonts.

Extensions

Ask a bioinformatics researcher, pathologist, or immunologist to speak to the class.

Support information

Respiratory infections can present with similar symptoms, such as cough, fever, and congestion. However, they may be caused by vastly different organisms that require very different treatments. For example, an antibiotic is used for treating bacterial infections, but has no effect on viral infections. In these instances, prescribing antibiotics for viral infections is not only ineffective but can contribute to the development of antibiotic resistance in bacteria. Diagnostic labs with sequencing capabilities can help reduce the misuse of antibiotics by identifying the causal agent (pathogen) based on a genetic fingerprint. Specific regions of the pathogen's genome, called barcodes, act like fingerprints that are unique to each species of pathogen. These barcode regions allow scientists to use public databases to search for a sequence match that identifies the pathogen group or species of the sequence isolated from a patient.

Career connections

- **Bioinformatics engineer:** Specializes in the application of computational techniques and tools to the management, analysis, and interpretation of biological data. Includes developing and implementing computational methods to analyze large data sets and extracting meaningful information from them. They also work closely with biologists, biochemists, and medical researchers to help them understand the data generated by their experiments and to advance in modern biology and medicine
- **Pathologist:** Diagnoses diseases by examining tissue samples, cells, and bodily fluids, serving as critical “doctors’ doctors” behind the scenes. These specialized physicians lead laboratory teams to identify cancers, infections, and chronic conditions, guiding treatment plans through analysis. They typically specialize in anatomic pathology (tissues), clinical pathology (fluids), or both.
- **Immunologist:** Studies, analyzes, and/or treats disease processes that involve the immune system. These research scientists or practicing specialists analyze the immune responses to disease in order to help treat or reduce the effects of the disease.