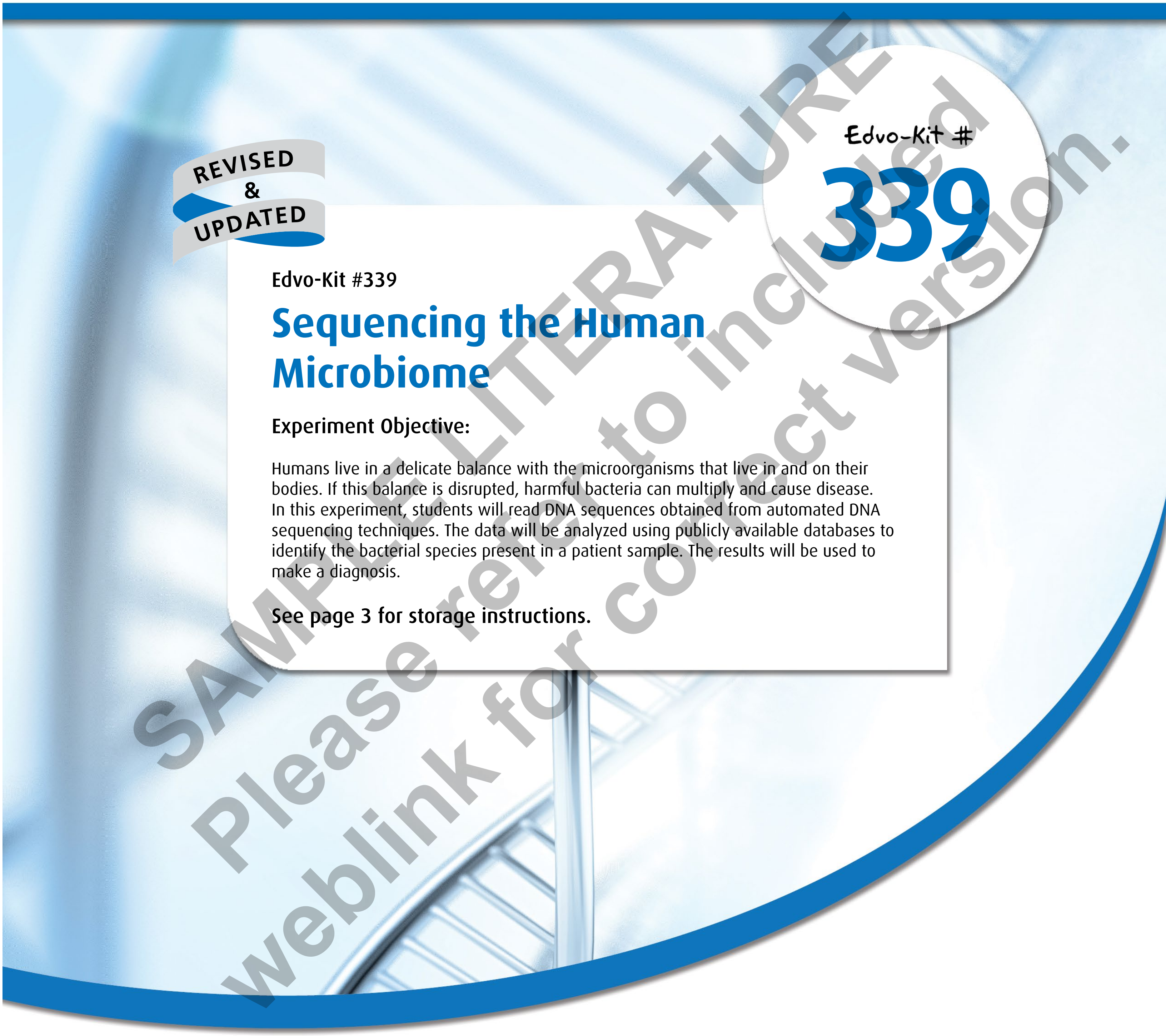
EDVOTEK®

THE **BIOTECHNOLOGY**
EDUCATION COMPANY®



REVISED & UPDATED

Edvo-Kit # 339

Edvo-Kit #339

Sequencing the Human Microbiome

Experiment Objective:

Humans live in a delicate balance with the microorganisms that live in and on their bodies. If this balance is disrupted, harmful bacteria can multiply and cause disease. In this experiment, students will read DNA sequences obtained from automated DNA sequencing techniques. The data will be analyzed using publicly available databases to identify the bacterial species present in a patient sample. The results will be used to make a diagnosis.

See page 3 for storage instructions.

Table of Contents

	Page
Experiment Components	2
Experiment Requirements	2
Background Information	3
Tutorial: Performing a BLASTN Search	8
Exercises	13
Study & Discussion Questions	15
Instructor's Guidelines	16
Answers to Exercises	17
Answers to Study Questions	19

Experiment Components

This experiment contains a total of twelve sections of automated DNA sequence printouts. Students can use any DNA sequence database to perform the activities in this lab. For purposes of simplification we have chosen to illustrate the database offered by the NCBI.

Requirements

- Computer with Internet access

Background Information

MICROBES: YOU'RE NEVER ALONE

Even when you're by yourself, you're never alone! Microbes are in and on your body – on your skin, in your mouth, in your intestines, everywhere. Current estimates suggest that for every single human cell, there are at least three bacterial cells in the body – and that ratio grows if we include fungi, protozoa, and viruses.

But this isn't a bad thing! These common microbes live in a symbiotic relationship with humans, in which both the person and the microorganisms benefit. For example, bacteria play an important role in the human gastrointestinal system. They help produce vitamins, digest our food, and even fight off invading pathogenic (disease-causing) bacteria, all while being nourished by our bodies. The normal microbiota, or the resident microorganisms, are essential for good health. Collectively, the vast community of microbes that colonize our bodies are known as the microbiome.

To understand more about the microbiome, the National Human Genome Research Institute launched the Human Microbiome Project (HMP) in 2008. This initiative characterized the microorganisms living on and in the human body by analyzing DNA, a cell's genetic material. Strands of DNA are built from four nucleotides – Adenine, Thymine, Guanosine, and Cytosine. The order of nucleotides creates genes, which are discrete units of genetic information that contain the instructions to build and maintain an organism. DNA sequencing uses biotechnology to determine the precise order of these nucleotides.

Each microorganism has differences in their DNA sequences which allows researchers to confirm the species present in a patient sample. Researchers sequenced the microbiome of 300 people over time to learn about the variety of species present, and how the proportions of microorganisms change, in response to stress, antibiotics, and other external factors. This information helps us to better understand human physiology and the interactions between our bodies and our flora and fauna.

The results published by the HMP confirmed that there are many families of bacteria that live in the gut (Figure 1). While most are from the phyla *Bacteroides* and *Firmicutes*, researchers have identified *Proteobacteria, Actinobacteria, Spirochaetes,* and *Cyanobacteria* living in the gastrointestinal system. The proportions of each type of microbe in the healthy adult body remains rather constant over time. However, a change in diet or the introduction of antibiotics can alter the species and proportion of microorganisms. Dangerous bacteria like *Clostridioides difficile* (often abbreviated as *C. diff*), which may already live in the intestine, can multiply after such events and cause diarrhea and an inflammatory bowel condition called colitis.

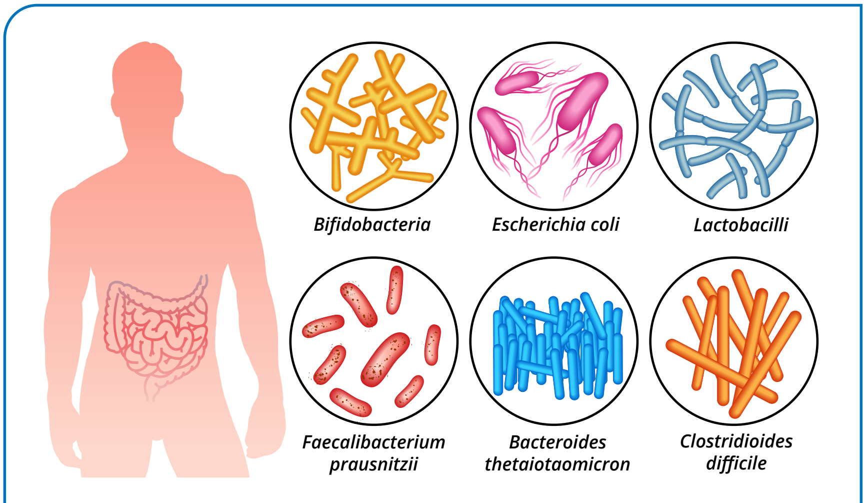

Bifidobacteria
Escherichia coli
Lactobacilli
Faecalibacterium prausnitzii
Bacteroides thetaiotaomicron
Clostridioides difficile

Figure 1: Members of the Gastrointestinal Microbiome.

For the DNA sequencing data to provide insight into these questions, scientists must address the computational challenges around data analysis and visualization. A challenge for researchers studying the microbiome involves finding creative and efficient ways to analyze and manage the vast amounts of data being generated. This has been addressed by the field of bioinformatics – a discipline that blends computer science, biology, and information technology.

DNA SEQUENCING AND DATABASE SEARCHES

The first step in a sequencing project is collecting the raw data – the precise order of the four nucleotides of the DNA in the microbiome sample. There are several approaches to generating sequence information and new methods are emerging each year. Two popular methods are chain termination sequencing and sequencing by synthesis.

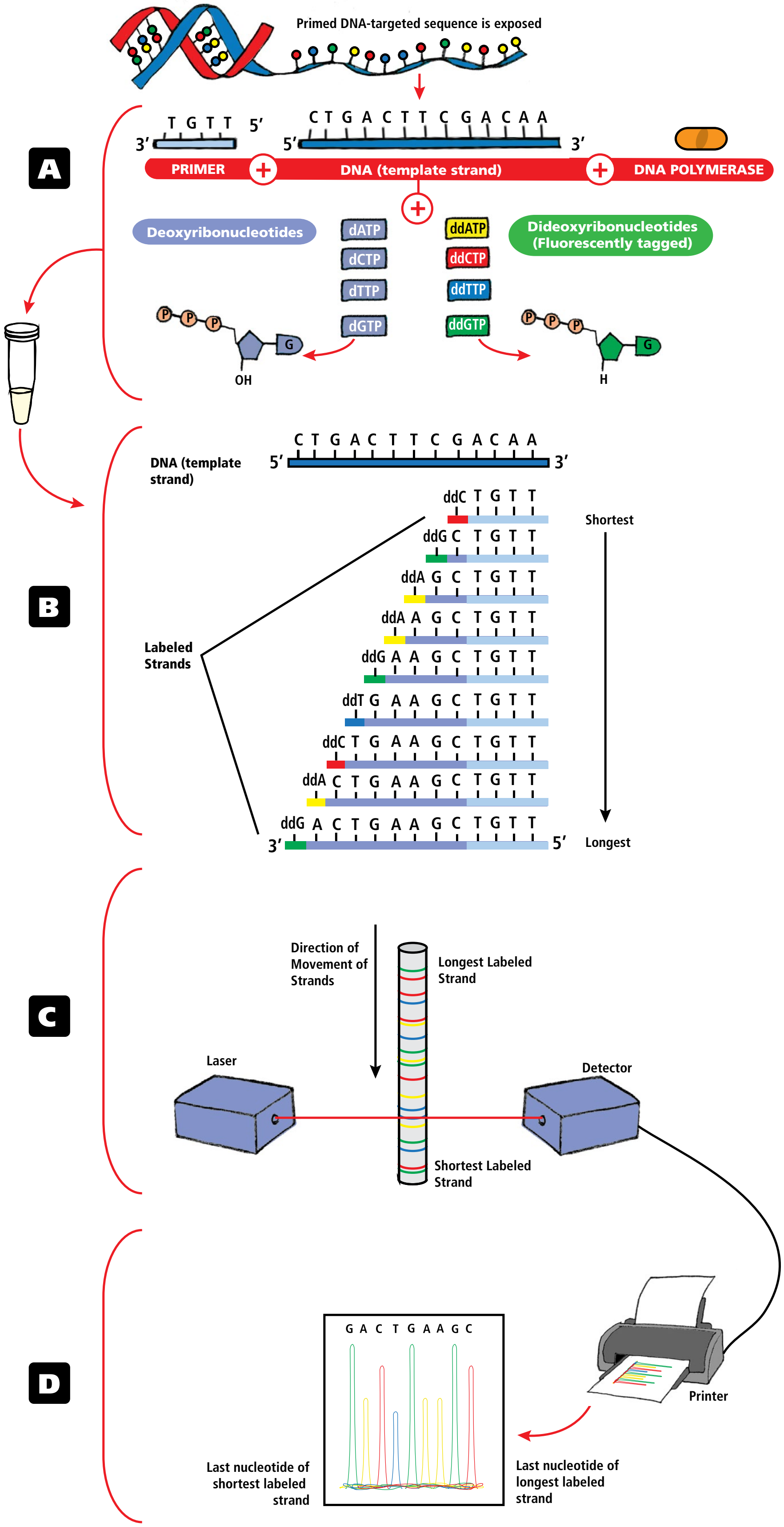
Frederick Sanger and colleagues developed chain termination sequencing in 1977. This technique, commonly known as Sanger sequencing, was the major method used to create the first genome sequence. It is still used today because it allows researchers to specifically target and sequence longer sequences of nucleotides (between 500 and 800 base pairs) from a specific location within a strand of DNA. Genes are targeted using a short, synthetic piece of DNA called a primer that base pairs with a specific DNA sequence. The sample to be sequenced is combined with the primer, the DNA-building enzyme DNA polymerase I (DNA Pol I), and a blend of nucleotides. This mixture includes a high concentration of deoxynucleotides (dNTPs) and a low concentration of dideoxynucleotides (ddNTPs) (Figure 2A). During the sequencing reaction, DNA Pol I reads the DNA template and adds nucleotides to the primer to build a complementary strand of DNA. Most times, the polymerase will add a dNTP to the growing nucleotide chain. However, when DNA Pol I adds a ddNTP to the DNA strand, it is impossible for the polymerase to add another nucleotide to the end of growing strand. This is because the ddNTPs lack the 3′ hydroxyl group that DNA Pol I uses to link nucleotides to the growing DNA chain (Fig 2B). This stops the reaction, creating a series of DNA fragments of differing sizes. The shortest fragments result from DNA strands that terminated near the primer, whereas longer fragments had more dNTPs linked to the growing DNA chain before the incorporation of the ddNTP (Figure 2B).

To read the sequence, the DNA fragments are separated by capillary electrophoresis (Figures 2B, 2C). Automated machines separate DNA fragments through a polyacrylamide gel formed in a thin capillary tube. Importantly, each ddNTP is labeled with a different fluorescent marker, allowing the sequence of a particular DNA strand to be “read” by a laser that is focused on the capillary (Figure 2C). As the DNA fragments pass through the gel matrix, the laser excites the fluorophores on the DNA. The four different ddNTP fluorescence colors are then automatically detected and the fluorescence intensity translated into a data “peak” that represents the order of nucleotides in the template DNA (Figure 2D). These high-throughput machines can provide fast, high-quality sequence for a fraction of the cost of traditional sequencing strategies.

Sequencing by synthesis also reads the sequence of nucleotides by building new DNA, but in a different way. First, the long strands of sample DNA are broken into random smaller pieces. These short double-stranded DNA fragments are anchored to a solid surface, and one strand is eliminated from the duplex. DNA polymerase copies the single-stranded DNA, creating distinct spots of identical DNA molecules attached to the scaffold. To sequence the sample, fluorescently labeled nucleotides are added to the reaction, one at a time. Using the bound DNA as a template, DNA Pol I adds the next nucleotide. These nucleotides are different than the ones used for Sanger sequencing because the fluorophore blocks the addition of the next nucleotide. After all four nucleotides are added, the fluorescent signals are recorded on a computer and translated into sequence information. The fluorophores are removed, and then the next round of dNTPs are added. In this way, we sequence DNA by building the complementary DNA one nucleotide at a time and taking a ‘snapshot’ of the sequence after each dNTP addition.

The advantage of sequencing by synthesis is it generates multiple DNA sequence reads in a single sequencing reaction. This means that sequencing by synthesis has a much higher throughput and a lower cost than Sanger sequencing. One disadvantage of this technique is that it produces shorter reads (50-150 bp), meaning that many reads must be performed to sequence a single gene.

Figure 2: Sanger DNA Sequencing.
(A) Setting up the sequencing reaction.
(B) Incorporation of the ddNTPs create different size DNA fragments.
(C) The labeled mixture is sequenced using capillary gel electrophoresis. A laser detects the fluorescent label on each of the ddNTPs.
(D) The information is analyzed using a computer.



INTERPRETING DNA SEQUENCE INFORMATION

After obtaining DNA sequencing data, molecular biologists filter out any human DNA sequences before comparing the results to sequences in existing public databases. These searches can reveal research already performed on the sequenced gene, including the three-dimensional structure of the gene product, diseases associated with the sequence, and in which tissues the gene is active. In cases where an organism has not been well-studied, finding similar sequences can provide clues to the sequence's function and its evolutionary relationship to other human genes. In the context of the microbiome, these searches identify and confirm the species of microorganisms present in the sample.

GenBank is one of the largest and most influential databases of DNA sequences. This free, open source database contains over a trillion nucleotide bases of publicly available sequence data. Each entry in GenBank contains a sequence, an accession number, and a wealth of supporting bibliographic and biological annotations such as author references and taxonomic data. The NCBI (National Center for Biotechnology Information) oversees and maintains the database as a whole but each entry is submitted directly by individual laboratories. Direct submission has allowed the database to keep pace with the rapid growth in sequence data production. However, it also means that the quality of the data is not always consistent, especially in the certainty of each nucleotide's identity and in the completeness of attached annotation.

Alongside the sequence, the NCBI features several useful bioinformatics tools that allow us to explore the data. One tool, the Basic Local Alignment Search Tool, or BLAST, compares a user's DNA sequence with those found in the GenBank database. BLAST works like a DNA search engine. The researcher enters the sequence into the search bar, sets the parameters, and hits the BLAST button. Within a few minutes, the sequence matches are displayed. In BLAST terminology, the user's input sequence is the query sequence, sequences in the database are the target sequences, and the sequence matches are the hits.

Microbiome researchers use BLAST to identify microorganisms present in a clinical sample based on their sequences. If the DNA sequence does not have an identical match, BLAST identifies closely related organisms with similar genes. Such similarity suggests shared ancestry between the genes, or homology. By looking at the sequences with similarities to the input sequence, or hits, researchers predict the molecular function of new genes. BLAST can also identify specific protein domains, like DNA-binding domains or ATP-binding motifs, based upon the DNA sequence.

BLAST takes a heuristic approach to the problem of searching through such a mammoth database of target sequences (Figure 3). This means that it takes shortcuts in order to find sequence matches in a reasonable time frame. These shortcuts assume that biologically similar sequences contain short stretches of DNA that match. BLAST attempts to find these matching segments by dividing the sequence into much shorter seeds, and then scanning the database for matches.

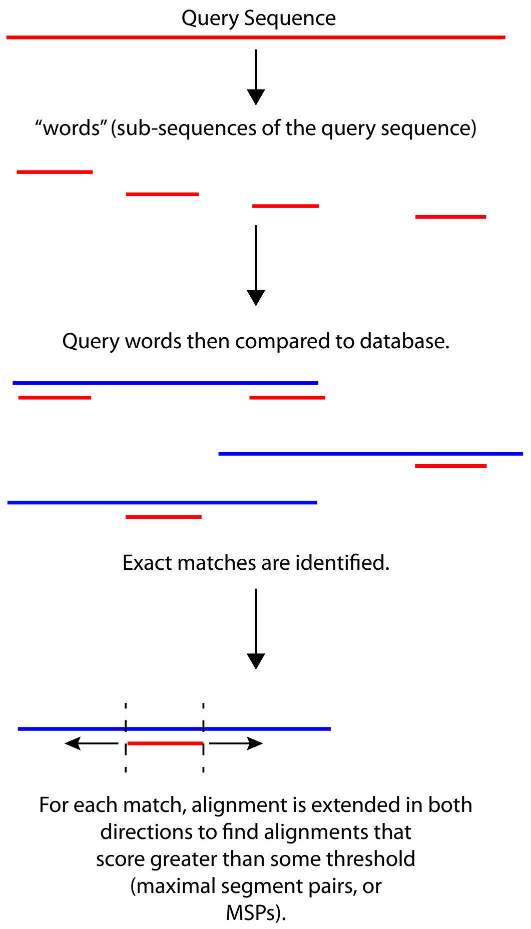

Query Sequence
"words" (sub-sequences of the query sequence)
Query words then compared to database.
Exact matches are identified.
For each match, alignment is extended in both directions to find alignments that score greater than some threshold (maximal segment pairs, or MSPs).

Figure 3: How BLAST Works

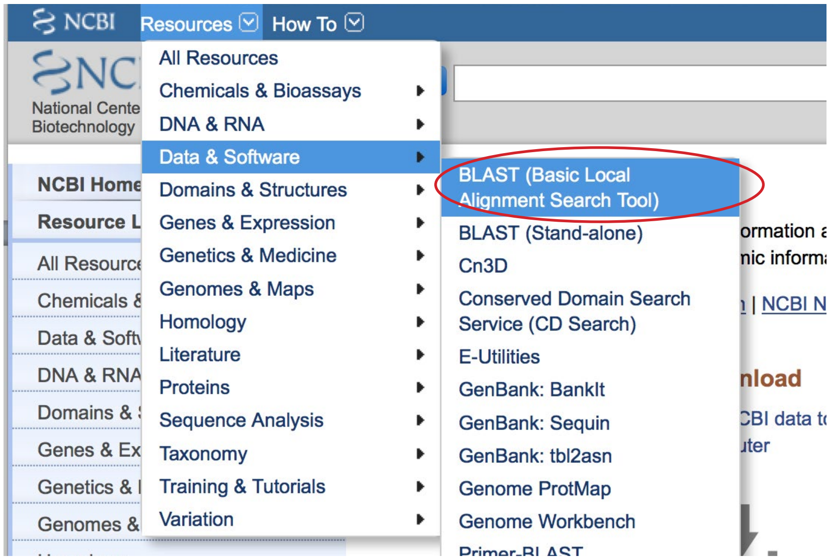
Once it has generated a list of high scoring matches, BLAST extends the seeds to see if they are contained in longer high scoring alignments. By searching the GenBank database this way BLAST can return results very quickly although it sacrifices some accuracy and precision.

The BLAST software is popular not only because of its speed but also because it computes the statistical significance of each match. In addition to the accession number, description, and genome link, BLAST provides a score, bit score, and e value for the matching sequences. The score, S, is a raw measure of the quality of alignment between the query and the hit. The bit score is the raw score (S) adjusted for the size of the database and the sequence length. The e value represents the probability due to chance that there is another alignment with a similarity score greater than the given S score. These three metrics are a good first indicator of similarity between sequences.

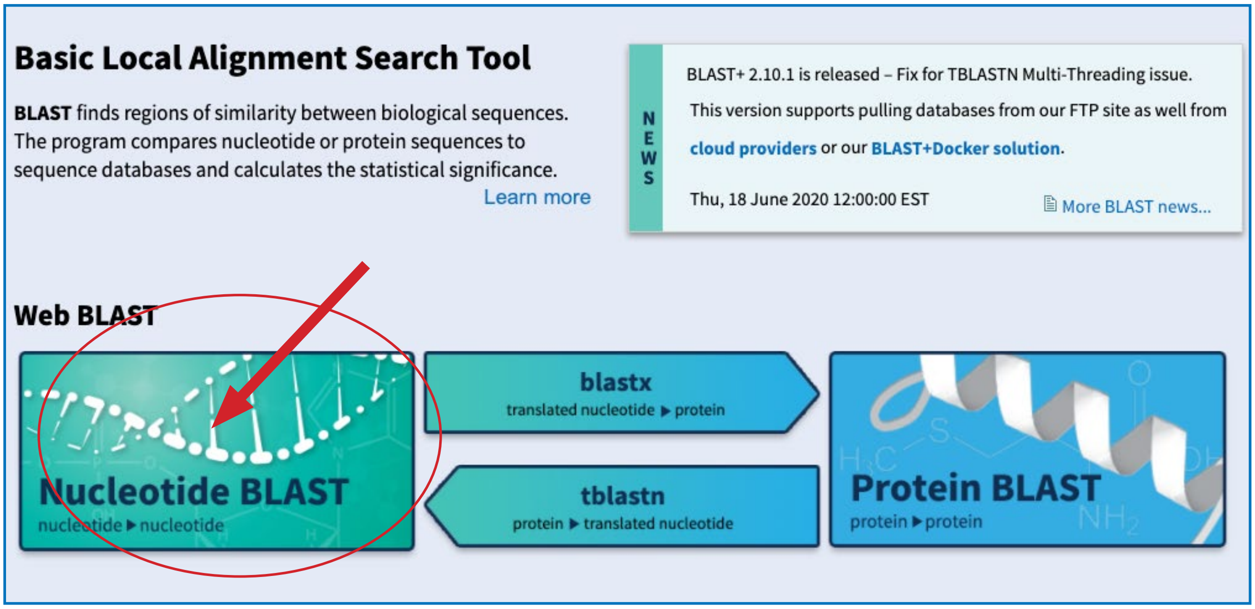
This exercise introduces students to genomics and bioinformatics by sequencing the microbiome of three patients who have been exhibiting several symptoms consistent with a *C. diff* infection. Students use BLASTN to search the NCBI database using DNA sequences from microbes found within patient samples. Students will identify the genus and species of each microorganism to determine which patient has been infected by *C. diff*.

Tutorial: Performing a BLASTN Search

1. Type: **www.ncbi.nlm.nih.gov** to log on to the NCBI web page.
2. On the top left of the screen click on the drop down menu “Resources”.
3. Click on “Data and Software”, then click on “BLAST” (Basic Local Alignment Search Tool.)



4. On the new BLAST Home screen select “nucleotide blast” which is the first opinion under the Basic BLAST list.



Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.
Learn more

NEWS

BLAST+ 2.10.1 is released – Fix for TBLASTN Multi-Threading issue.

This version supports pulling databases from our FTP site as well from **cloud providers** or our **BLAST+Docker solution.**

Thu, 18 June 2020 12:00:00 EST

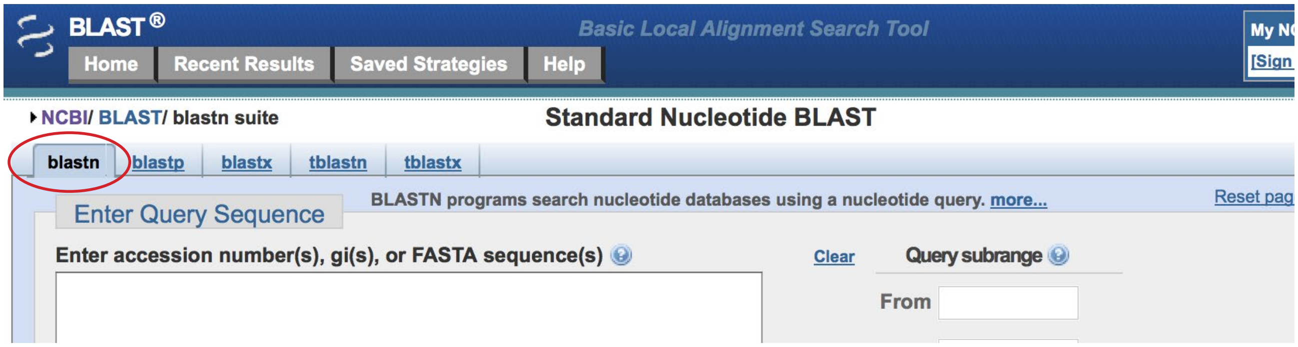
More BLAST news...

Web BLAST

continued...

Tutorial: Performing a BLASTN Search, continued

5. On the new screen make sure the tab selected is "blastn".



Standard Nucleotide BLAST

BLASTN programs search nucleotide databases using a nucleotide query.

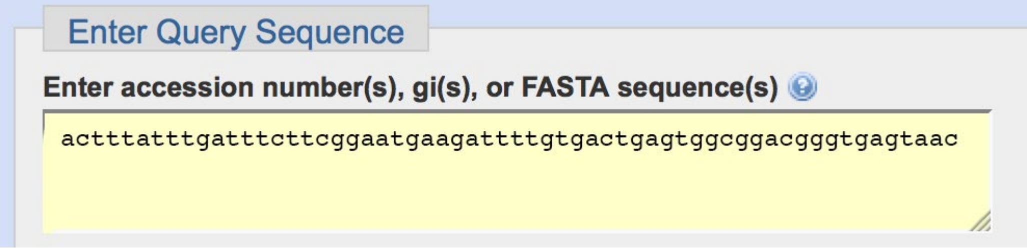
Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

Query subrange

From

6. Enter the nucleotide sequence into the large box in the "Enter Query Sequence" section; be careful to type the following sequence exactly: actttatttgatttcttcggaatgaagattttgtgactgagtggcggacgggtgagtaac

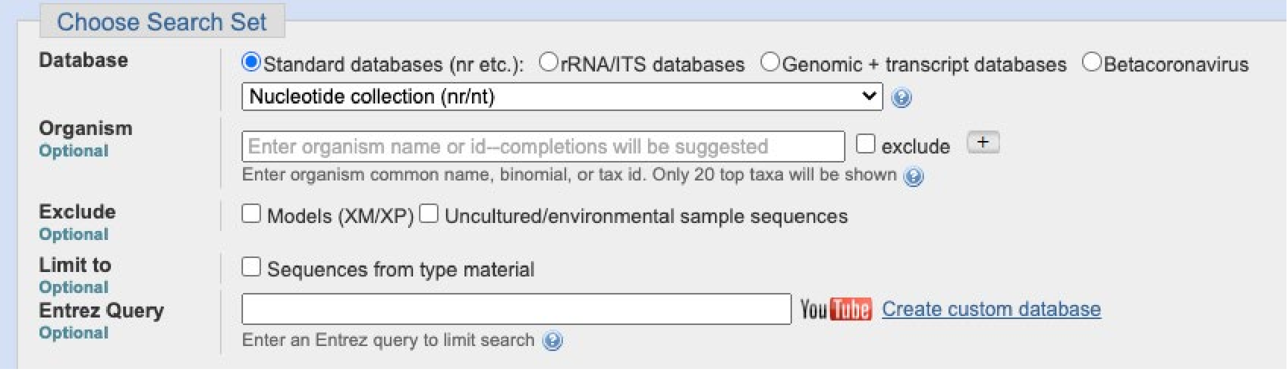


Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

```
actttatttgatttcttcggaatgaagattttgtgactgagtggcggacgggtgagtaac
```

7. Under "Choose Search Set" make sure that "Standard databases (nr etc) is selected and that "Nucleotide collection (nr/nt)" is highlighted in the dropdown menu. The remaining entries should be left blank.



Choose Search Set

Database ● Standard databases (nr etc.): ○ rRNA/ITS databases ○ Genomic + transcript databases ○ Betacoronavirus

Nucleotide collection (nr/nt)

Organism Optional: Enter organism name or id--completions will be suggested □ exclude +

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Exclude Optional: □ Models (XM/XP) □ Uncultured/environmental sample sequences

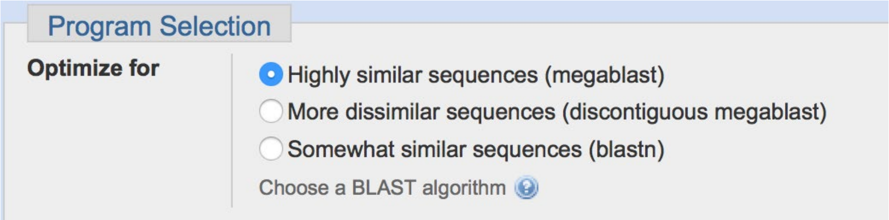
Limit to Optional: □ Sequences from type material

Entrez Query Optional: Create custom database

Enter an Entrez query to limit search

Tutorial: Performing a BLASTN Search, continued

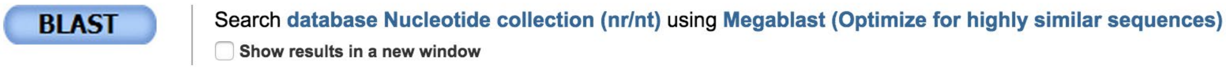
8. Under “Program Selection” select “High similar sequence (megablast)”



Program Selection

Optimize for
● Highly similar sequences (megablast)
○ More dissimilar sequences (discontiguous megablast)
○ Somewhat similar sequences (blastn)
Choose a BLAST algorithm

9. Click on the blue “BLAST” button to start your query.



BLAST

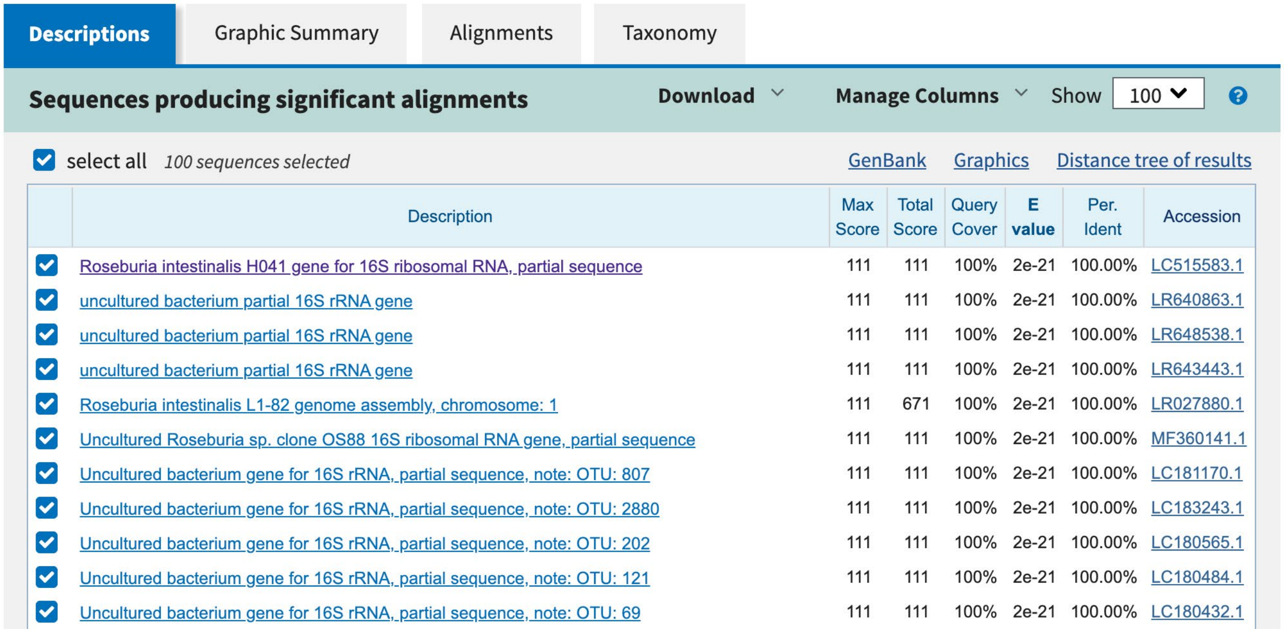
Search **database Nucleotide collection (nr/nt)** using **Megablast (Optimize for highly similar sequences)**
□ Show results in a new window

10. Once the “BLAST” button has been clicked you will be assigned an ID#. Record this number so you can check your results at a later time.

11. Examine the BLASTN search report. The search summary report at the top of the webpage shows an overview of the BLASTN search parameters.

12. The results are displayed in a series of tabs.

    a. Description section that shows all the sequences in the database with significant sequence homology to our sequence. By default, the results are sorted according to the E-value but you can click on the column header to sort the results by different categories. Notice that there can be several different entries with identical high scores.



Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download Manage Columns Show 100

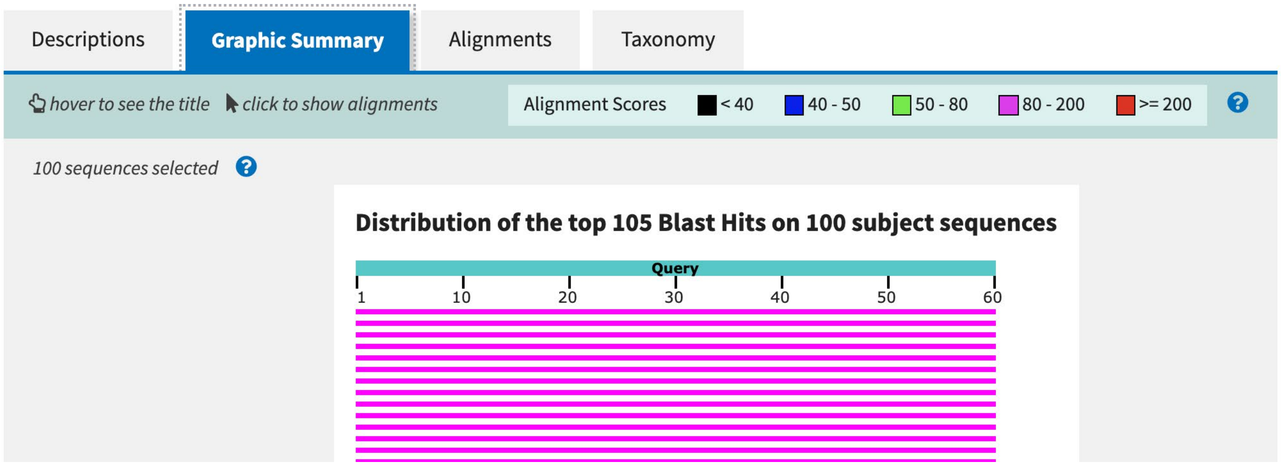
☑ select all *100 sequences selected* GenBank Graphics Distance tree of results

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
☑	Roseburia intestinalis H041 gene for 16S ribosomal RNA, partial sequence	111	111	100%	2e-21	100.00%	LC515583.1
☑	uncultured bacterium partial 16S rRNA gene	111	111	100%	2e-21	100.00%	LR640863.1
☑	uncultured bacterium partial 16S rRNA gene	111	111	100%	2e-21	100.00%	LR648538.1
☑	uncultured bacterium partial 16S rRNA gene	111	111	100%	2e-21	100.00%	LR643443.1
☑	Roseburia intestinalis L1-82 genome assembly, chromosome: 1	111	671	100%	2e-21	100.00%	LR027880.1
☑	Uncultured Roseburia sp. clone OS88 16S ribosomal RNA gene, partial sequence	111	111	100%	2e-21	100.00%	MF360141.1
☑	Uncultured bacterium gene for 16S rRNA, partial sequence, note: OTU: 807	111	111	100%	2e-21	100.00%	LC181170.1
☑	Uncultured bacterium gene for 16S rRNA, partial sequence, note: OTU: 2880	111	111	100%	2e-21	100.00%	LC183243.1
☑	Uncultured bacterium gene for 16S rRNA, partial sequence, note: OTU: 202	111	111	100%	2e-21	100.00%	LC180565.1
☑	Uncultured bacterium gene for 16S rRNA, partial sequence, note: OTU: 121	111	111	100%	2e-21	100.00%	LC180484.1
☑	Uncultured bacterium gene for 16S rRNA, partial sequence, note: OTU: 69	111	111	100%	2e-21	100.00%	LC180432.1

continued...

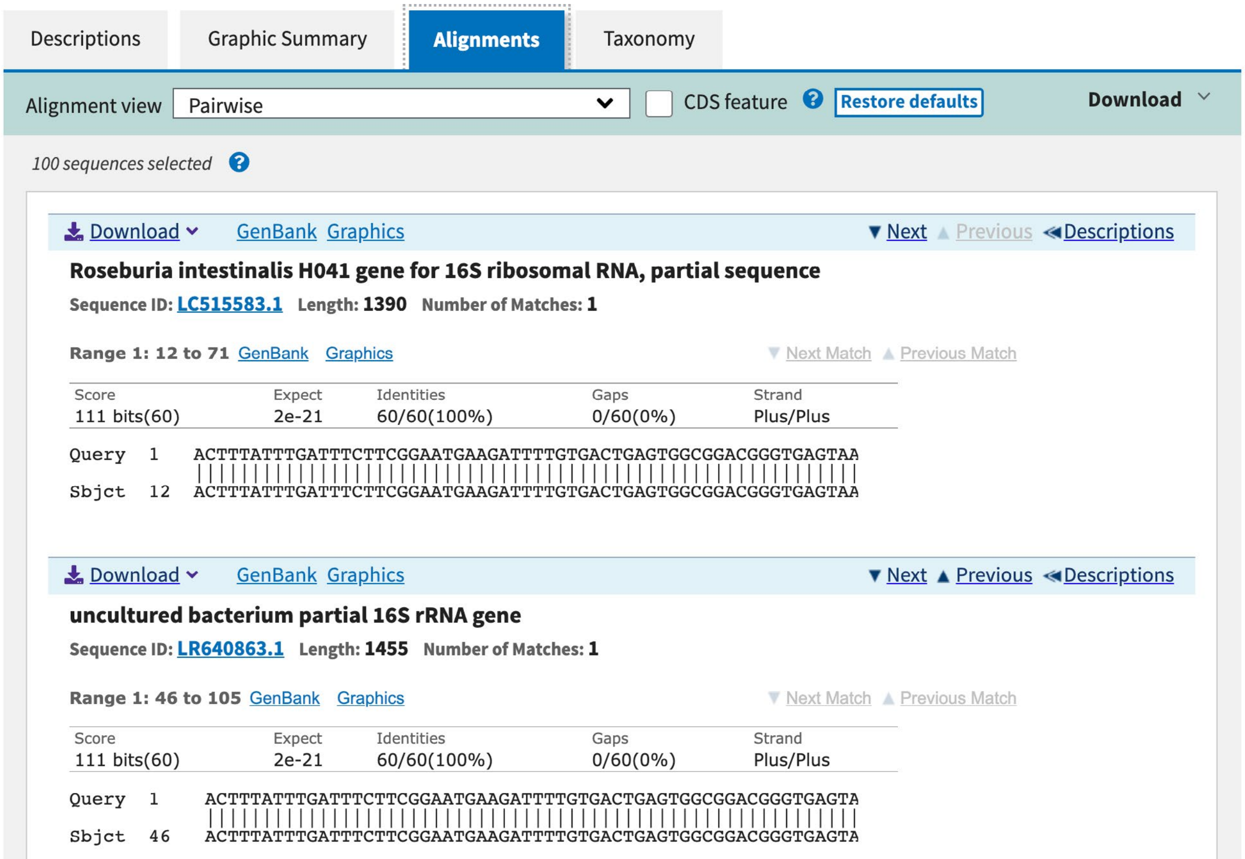
Tutorial: Performing a BLASTN Search, continued

b. Graphic Summary section that shows the alignment of database matches to the query sequence. The color of the boxes corresponds to the score of the alignment with red representing the highest alignment scores.



Distribution of the top 105 Blast Hits on 100 subject sequences

c. Alignment section that shows alignment blocks for each BLAST hit. Each alignment block begins with a summary that includes the score and expected value, sequence identity, the number of gaps in the alignment, and the orientation of the query sequence relative to the subject sequence.



Alignment view Pairwise

□ CDS feature

Roseburia intestinalis H041 gene for 16S ribosomal RNA, partial sequence

Sequence ID: LC515583.1 **Length:** 1390 **Number of Matches:** 1

Range 1: 12 to 71

Score	Expect	Identities	Gaps	Strand
111 bits(60)	2e-21	60/60(100%)	0/60(0%)	Plus/Plus

```
Query  1   ACTTTATTTGATTTCTTCGGAATGAAGATTTTGTGACTGAGTGGCGGACGGGTGAGTAA
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  12  ACTTTATTTGATTTCTTCGGAATGAAGATTTTGTGACTGAGTGGCGGACGGGTGAGTAA
```

uncultured bacterium partial 16S rRNA gene

Sequence ID: LR640863.1 **Length:** 1455 **Number of Matches:** 1

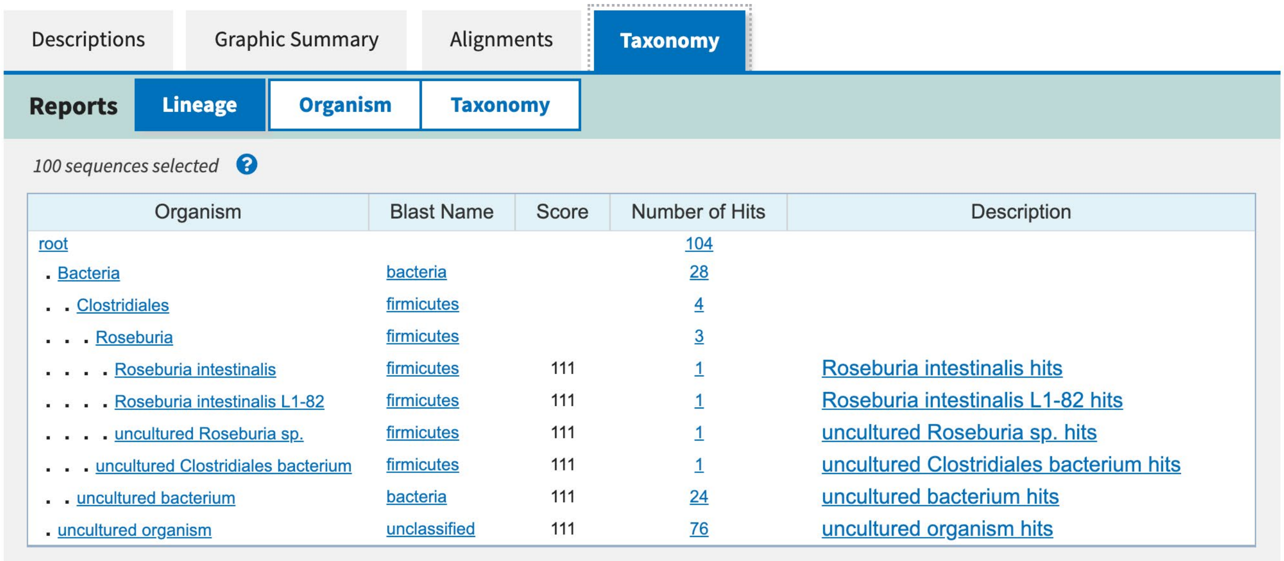
Range 1: 46 to 105

Score	Expect	Identities	Gaps	Strand
111 bits(60)	2e-21	60/60(100%)	0/60(0%)	Plus/Plus

```
Query  1    ACTTTATTTGATTTCTTCGGAATGAAGATTTTGTGACTGAGTGGCGGACGGGTGAGTA
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  46   ACTTTATTTGATTTCTTCGGAATGAAGATTTTGTGACTGAGTGGCGGACGGGTGAGTA
```

Tutorial: Performing a BLASTN Search, continued

d. Taxonomy section that shows the relationship between the target sequence and closely related sequences.



100 sequences selected

Organism	Blast Name	Score	Number of Hits	Description
root			104	
. Bacteria	bacteria		28	
. . Clostridiales	firmicutes		4	
. . . Roseburia	firmicutes		3	
. . . . Roseburia intestinalis	firmicutes	111	1	Roseburia intestinalis hits
. . . . Roseburia intestinalis L1-82	firmicutes	111	1	Roseburia intestinalis L1-82 hits
. . . . uncultured Roseburia sp.	firmicutes	111	1	uncultured Roseburia sp. hits
. . . uncultured Clostridiales bacterium	firmicutes	111	1	uncultured Clostridiales bacterium hits
. . uncultured bacterium	bacteria	111	24	uncultured bacterium hits
. uncultured organism	unclassified	111	76	uncultured organism hits

13. Select the top sequence in the description tab to explore in-depth. You can do this by clicking on a colored bar in the graphic section, clicking on the sequence name in the description section, or scrolling down to the alignment section. Then click on the sequence ID. This brings up additional information about the subject sequence, including the gene name, the genus and species of origin, and articles written about the gene. After performing this search, the top hit should be *Roseburia intestinalis* H041 gene for 16S ribosomal RNA, partial, Sequence ID: LC515583.1. If top hit does not match, try re-entering the sequence. Be sure to double check the search parameters before BLASTN searching.

Exercises

EXERCISE 1:

Three patients walk into your medical clinic with gastrointestinal distress. As a medical professional, you will interview each patient to identify symptoms and risk factors.

C. difficile, abbreviated as *C. diff*, is a Gram-positive bacterium that lives in the soil and in the animal digestive tract. It also can live in low numbers in the human gastrointestinal tract without causing disease. The bacteria is passed between individuals through the fecal-oral route, meaning that the bacteria can be passed by unwashed hands or contaminated food. *C. diff* is often transmitted in hospitals through contaminated surfaces or medical instruments.

In cases of active *C. diff* infection, the toxin produced by the bacteria attacks the epithelial layer of the intestines which produces an inflammatory reaction. This produces symptoms that include multiple loose or watery stools in a 24-hour period, severe abdominal pain and cramping, dehydration, rapid heart rate, fever, swollen abdomen, nausea, blood in the stool, kidney failure, and increased white blood cell count. Severe cases may cause intestinal inflammation and sepsis.

Risk factors for *C. diff* infections include age (> 65 years), exposure to antibiotics in the past 60 days, previous hospitalization, environmental transfer, and history of a previous *C. diff* infection. It is often spread in long-term care facilities and nursing homes. Chronic conditions like kidney disease, inflammatory bowel disease or liver disease increase the likelihood of *C. diff* infection.

READ through the patient histories and **IDENTIFY** the following:

1) Symptoms of *C. diff* infection
2) Risk factors of *C. diff* infection
3) Evidence that suggests a different cause for the symptoms.

Patient 1:

82 years old, recent use of laxative to relieve constipation. Blood in stool, nausea, vomiting, and cramping. Resident at a nursing home. Recent trip to a buffet restaurant that was closed for health code violations. Previous *C. diff* infection.

Patient 2:

60 years old, recent use of antibiotics when hospitalized for pneumonia. Patient has high fever, chills, shortness of breath, and severe abdominal pain. History of peanut allergy. Increased white blood count, intestinal inflammation, multiple loose stools over several days.

Patient 3:

32 years old, loose stools, headache, dehydration, rapid heart rate and breathing, elevated temperature. Patient was training for a marathon when symptoms began. In discussions, patient seems confused. Patient recently traveled internationally. Diagnosed with inflammatory bowel disease at 22.

Exercises, continued

EXERCISE 2:

Since all three patients have symptoms and risk factors for *C. diff* infection, the medical team collected patient samples and sent them to the laboratory for microbiome sequencing. We will be sequencing the 16s rRNA, which is a part of the bacterial ribosome responsible for initiating gene transcription. While the sequence is well-conserved, there are small sequence differences that can be used to distinguish between species. These sequence differences can be identified in BLAST.

Now that you have familiarity with the entry and submission process of BLAST, you will use the program to analyze four unique DNA sequences found in each patient sample. First, **READ** the DNA sequence information from the automated gel run sequence printout. Each patient has four sequence lanes to be analyzed, each of which represents the 16s rRNA sequence from a separate bacterial species. The sequence should be recorded from left to right in your notebook. Reversing the order of the nucleotides will give erroneous results.

Next, use BLAST to identify the bacterial species in the patient samples. To do this:

1) Identify each nucleotide sequence from the DNA Sequence readout.
2) Type all the bases in the query box of the BLAST program at the NCBI website. The bases can be from any region of the sequence, but they should be contiguous.
3) Examine the BLASTN search report, identify the bacterial species, and examine the gene ID for detailed information.

Once the microorganism has been identified, answer the following questions:

1) What is the name of this bacteria?
2) Compared to the GenBank entry, what strand have you read?
3) Can you find a paper that has been written about this microbe? Record the title and the authors.

Remember the following:

- The automated sequence differentiates the bases as follows: A is green, C is blue, G is black, and T is red.

- DNA is double stranded and contains a top (5′➔3′) and bottom (3′➔5′) strand (sometimes this corresponds to the coding and noncoding strands.) A DNA sequence is always entered in the 5′➔3′ direction.

- Sometimes it is difficult to read a nucleotide peak. This is particularly true at the beginning and end of a sequence read where peaks may overlap. Such ambiguous places are often labeled with a N rather than one of the four nucleotides.

- By clicking on the GenBank accession number you can access additional information such as the protein/ amino acid sequence, descriptions of the sequence/ gene, and the contributing scientists' names.

- Some high-scoring sequence matches may be predicted bacteria that do not have any research papers associated with them. Students may have to check several matches before they will find a sequence with an associated reference.

Study Questions

1. What is bioinformatics? How have advances in sequencing technology affected this field?

2. Name two sequencing methods and describe the tradeoff between the production rate and the length of the sequences produced.

3. What assumption does BLAST make? What are the advantages and disadvantages of making this assumption?

4. You have collected a medical history from three patients and sequenced the microorganisms present in clinical samples. Knowing the test results and the patient medical history, which patient has a *C. diff* infection? Write a full medical report for the patient detailing the risk factors, symptoms, and sequence analysis.

Instructor's Guide

PATIENT SEQUENCES

Patient 1

Sequence A: CTTTTACAATGAAGAGTTTGATCCTGGCTCAGGATGAACGCTAGCTACAGGCTTAACACATGCAAGTCGAGGGGCAGCATTTCAGTTTGCTTGCAAACTGGAGATGGCGACCGGCGCACGGGTGAGTAACACGTATCCAA

Sequence B: CAGCTTGCTGCTTTGCTGACGAGTGGCGGACGGGTGAGTAATGTCTGGGAAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATACCGCATAACGTCGCAAGCACAAAGAGGGGGACCTTAGGGCCTCTT

Sequence C: AGTGTGAAAAACTCCGGTGGTATAAGATGGACCCGCGTTGGATTAGCTAGTTGGTGAGGTAACGGCCCACCAAGGCGACGATCCATAGCCGACCTGAGAGGGTGACCGGCCACATTGGGACTGAGACACGGCCCAAACTC

Sequence D: GATCCTGGCTCAGGCGAACGCTGGCGGCGCGCCTAACACATGCAAGTCGAACGAGCGAGAGAGAGCTTGCTTTCTCAAGCGAGTGGCGAACGGGTGAGTAACGCGTGAGGAACCTGCCTCAAAGAGGGGGACAACAGTTG

Patient 2

Sequence A: GATGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGGGATCCATCAAGCTTGCTTGGTGGTGAGAGTGGCGAACGGGTGAGTAATGCGTGACCGACCTGCCCCATGCTCCGGAATAGCTCCTGGAAACGGGTGGT

Sequence B: GATCCTGGCTCAGGCGAACGCTGGCGGCGCGCCTAACACATGCAAGTCGAACGAGCGAGAGAGAGCTTGCTTTCTCAAGCGAGTGGCGAACGGGTGAGTAACGCGTGAGGAACCTGCCTCAAAGAGGGGGACAACAGTTG

Sequence C: GCGGCGGACGGGTGAGTAACGCGTGGGTAACCTACCCTGTACACACGGATAACATACCGAAAGGTATGCTAATACGGGATAATATATTTGAGAGGCATCTCTTGAATATCAAAGGTGAGCCAGTACAGGATGGACCCGCG

Sequence D: ACTTCGGTGATGACGTTGGGAACGCGAGCGGCGGATGGGTGAGTAACACGTGGGGAACCTGCCCCATAGTCTGGGATACCACTTGGAAACAGGTGCTAATACCGGATAAGAAAGCAGATCGCATGATCAGCTTATAAAAG

Patient 3

Sequence A: CTTTTACAATGAAGAGTTTGATCCTGGCTCAGGATGAACGCTAGCTACAGGCTTAACACATGCAAGTCGAGGGGCAGCATTTCAGTTTGCTTGCAAACTGGAGATGGCGACCGGCGCACGGGTGAGTAACACGTATCCAA

Sequence B: GATGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGGGATCCATCAAGCTTGCTTGGTGGTGAGAGTGGCGAACGGGTGAGTAATGCGTGACCGACCTGCCCCATGCTCCGGAATAGCTCCTGGAAACGGGTGGT

Sequence C: AGTGTGAAAAACTCCGGTGGTATAAGATGGACCCGCGTTGGATTAGCTAGTTGGTGAGGTAACGGCCCACCAAGGCGACGATCCATAGCCGACCTGAGAGGGTGACCGGCCACATTGGGACTGAGACACGGCCCAAACTC

Sequence D: ACTTCGGTGATGACGTTGGGAACGCGAGCGGCGGATGGGTGAGTAACACGTGGGGAACCTGCCCCATAGTCTGGGATACCACTTGGAAACAGGTGCTAATACCGGATAAGAAAGCAGATCGCATGATCAGCTTATAAAAG

TEACHING THIS LESSON ONLINE?

We've created an online patient interview e-learning module, available here: https://h5p.org/node/972719. Your students can complete the lesson and print the summary screen as part of their assessment.

RESULTS:

The identity of the bacteria found in the patient samples can be determined using the NCBI website as described previously in this experiment. Please note that details about Gene ID, strand, references, and author information will be different depending upon which hit a student focuses on. However, the genus and species itself should match the information presented below.

	Symptoms	Risk Factors	Results of Test	*C. diff* Infection?
Patient 1	Blood in stool, nausea, vomiting, and cramping	Resident at a nursing home, age over 65, previous *C. diff* infection	A. *Bacteroides thetaiotaomicron* B. *Escherichia coli* C. *Eubacterium rectale* D. *Faecalibacterium prausnitzii*	No
Patient 2	Fever, severe abdominal pain, increased white blood count, intestinal inflammation, loose stools	Recent use of antibiotics, recently hospitalized	A. *Bifidobacterium bifidum* B. *Faecalibacterium prausnitzii* C. *Clostridioides difficile* D. *Lactobacillus acidophilus*	Yes
Patient 3	Dehydration, loose stools, rapid heart rate, fever	Inflammatory bowel syndrome	A. *Bacteroides thetaiotaomicron* B. *Bifidobacterium bifidum* C. *Eubacterium rectale* D. *Lactobacillus acidophilus*	No

NOTE:

Chromatograms annotated with the correct DNA sequence are provided as a separate PDF file for reference. If you choose to print the files, please be sure to follow these guidelines for best results:

- Printing should be done on US Ledger/Tabloid Sized paper (11 by 17 inches).
- For best results, be sure that page sizing and handling is set to "Actual Size".
- The chromatograms must be printed in full color to allow for differentiation between nucleotides.

Please refer to the kit insert for the Answers to Study Questions