REAL SCIENCE:

Untangling Evolutionary Trees

A workshop for teachers by the Harvard Museum of Natural History and Srivastava Lab

July 2019

Table of Contents

Workshop Overview

In this series of workshop units, participants will learn hands-on molecular biology techniques for answering questions about how all living things are related-- including what we eat-- with the ultimate objective of bringing this curriculum to their classrooms. Teachers, and subsequently their students, will learn the techniques and analytical methods that scientists use to determine how organisms are related, and perform these techniques themselves in a guided, hands-on lab. We have also included some fun activities that will help clarify some of the techniques. The effectiveness of this workshop in enhancing learning is important to us, therefore, the workshop will also include pre- and post-workshop surveys and periodic learning check-ins.

The workshop contains three major **UNITS**, each corresponding to one day of the workshop. **Unit 1** begins with an in depth consideration of how we represent the relationships of organisms, why knowing this matters, and how DNA sequence can be used to determine these relationships. Participants will extract DNA from a variety of organisms. In **Unit 2**, participants will use molecular methods to obtain a specific region of the DNA, corresponding to the 18S gene, and send it to a company for determining its sequence. Next, in **Unit 3**, these sequences will be used by participants to build a tree to uncover the relationships of the organisms under study.

The three main units contain independent **modules**, which can be adjusted to facilitate an effective implementation of this workshop with students in classrooms. Sections demarcated as **In the classroom** provide ideas or recommendations for how the modules can be translated to the classroom. Graduate students volunteers from Harvard University will work with teachers in redesigning the curriculum to fit the classroom's needs, and they will also help the teachers run the experiments and activities. The end of this manual provides instructions on how to take the next step-- to bring this curriculum to the classroom!

Introduction to Phylogenetics

What is phylogenetics?

Phylogenetics is a field of research that investigates the relationships between organisms based on DNA or protein sequences.

Why do scientists use phylogenetics? What questions can they answer using these techniques?

Scientists use phylogenetic approaches when trying to determine the relationships of different organisms or species; often, these placements have far-reaching consequences for how we think of traits evolving over time, or determining when new molecular and morphological changes first arise among different groups of organisms. To determine how organisms are related, scientists will often compare DNA sequences that all of those organisms have in common, and determine how similar or different these sequences are. This method works because DNA sequences in organisms evolve or change over time just as the organisms' anatomical features might, with closely-related organisms having more similar DNA sequences than distantly-related species do. DNA-based approaches can be more accurate than using morphological comparisons alone, as similar traits could have evolved convergently in organisms that are in fact distantly related, despite looking similar. From the similarities or differences between DNA sequences, scientists can create a phylogeny (or tree) that best describes how those organisms are related to one another. There are many notable successful examples of applying phylogenetics to determine previously undescribed or inconclusive relationships between organisms, including the finding that whales are closely-related to hippopotamuses, deers, and cows.

How can phylogenetics fit into the classroom?

Although phylogenetic techniques have been developed by scientists in the lab, they can be easily employed in the classroom. Much of the equipment for these techniques is portable, the reagents are not toxic, and the majority of software programs used are free and available to the public. We have created a series of modules here to help participants learn both the molecular and computational approaches that are used by scientists in the lab to determine how organisms are related. Throughout this workshop, participants will use their critical thinking skills to make predictions about how certain species are related, then conduct hands-on labs to isolate and sequence DNA just as it would be done in a research lab. Using these sequences, participants will create their own phylogenetic trees using the same phylogenetic approaches utilized in the lab, bringing textbook figures to life one module at a time.

UNIT 1: INFERRING ORGANISMAL RELATIONSHIPS WITH MORPHOLOGY AND DNA

Module 1: Introduction to tree thinking
Module 2: DNA extraction

Overview

In this unit, participants will learn what DNA is, where it is located within a cell/organism, and how scientists can interpret the relationships between animals using DNA. During preparatory work prior to the lab portion of this module, participants will make predictions about the relatedness of samples (collected from the grocery store) that they will use in this workshop. In a laboratory experiment, participants will extract genomic DNA from a variety of tissue samples, specifically isolating DNA from the rest of the cellular material. The genomic DNA isolated in this module will be used in Unit 2.

Learning goals

By the end of this module, participants will be able to:
-- Describe the location of DNA in the cell
-- Build a conceptual understanding of DNA and what it does:
- What does DNA look like?
- Why is DNA important?
- What information is in DNA and how can we access it?

-- Build a tree to depict predictions about how the samples are related to each other

Practical skills:

-Learn how to use laboratory equipment (pipettes, etc.)
-Learn how to extract DNA from a variety of organisms

Suggested prior knowledge

All living organisms are made up of cells
Some basic biology vocabulary: cell, nucleus, organism

In the classroom

When getting samples to extract DNA from, we recommend either going to the grocery store and picking out some items that could be a "fun" reveal (is ground beef actually ground beef?) or going for a walk with your students and collecting samples from outside. We recommend avoiding lettuce (spinach is ok) and seasoned meat.

LAB SUPPLIES FOR MODULE 1

REAL SCIENCE Team will provide:

Equipment set-up (per student or team):
Gloves
Pipettes
Pipette tips
Eppendorf tubes (1/sample for each student)
Blue tube and blue pestle
Tube racks
Laminated micropipette sheets (also in Appendix)
Permanent markers

General lab equipment needed:
Centrifuge
Waste containers
Ice
Freezer
Water bath or heat block, heated to 65C

In the classroom, teachers provide:

Samples
Go crazy at the grocery store! Select about 10 samples that are commonly bought items, which students can easily make predictions about which species are related. Only a tiny amount of any one sample is needed. Some examples:

Spinach, mushrooms, ground turkey, ground beef, pork, chicken, seaweed, strawberries/berries, scallops, shrimp, squid, fish, clam, collard greens, shallots

Solutions:
Potassium Acetate Solution
Extraction Buffer
Isopropanol
100% ethanol
Molecular biology grade water

Worksheet: DNA and Phylogenetic Trees

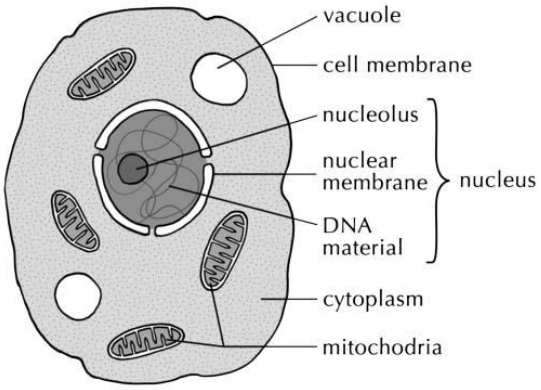
In this lab, you will extract DNA from plants and animals (maybe even a fungus) to get the information that scientists use to tell which species are closely related and which ones are not. Here're a few facts and questions to get you thinking about what we will be doing, and how that will tell us about the relationships of animals.

PART ONE: DNA

DNA (DeoxyriboNucleic Acid) is the material that carries genetic information. DNA is found inside each and every cell of our bodies (and those of other animals and plants too).

Question

Here's a drawing of a simple animal cell. What are some things that you could do to get the DNA out of the cell?



PART TWO: DNA IS INFORMATION

DNA is a chain of just four different molecules, which we call A, T, G, and C. These As, Ts, Gs, and Cs can be put together in different combinations. There are billions of these molecules in each cell in our bodies. The more closely related two people are, the more similar the strings of DNA look.

For example, a part of your DNA might look like this: **ATGCAG**

And the same part of DNA in a relative of yours might look like this: **ATGAAG**

But the DNA from someone outside your family might look like this: **AGCTTC**

If we line up these strings of DNA and color in the similarities, it is easier to tell which two strings are more alike, that is, more closely related:

ATGCAG
ATGAAG
AGCTTC

The arrangement of the As, Ts, Gs, and Cs in a string of DNA is called the DNA sequence. We can use DNA "sequence" to figure out how people or, more generally, organisms are related to each other.

Question

Given the following DNA sequences from three species, which two species are more closely related?

Species A: **TCAATGACTG**
Species B: **GCAATTGAAT**
Species C: **ACAATGTCTG**

PART THREE: BUILDING TREES FROM DNA

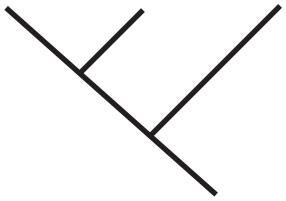
The DNA you will extract will be sent to a company in Cambridge, MA that will use sequencing machines to determine the sequence of one gene fragment – the 18S gene. This important gene helps our cells make proteins, and it does the same job in the cells of all of the species you will work on.

Even though the gene works the same way in all species, its sequence is a little bit different in each one! We can use this to build a "tree" – a drawing that depicts the relationships of these species.

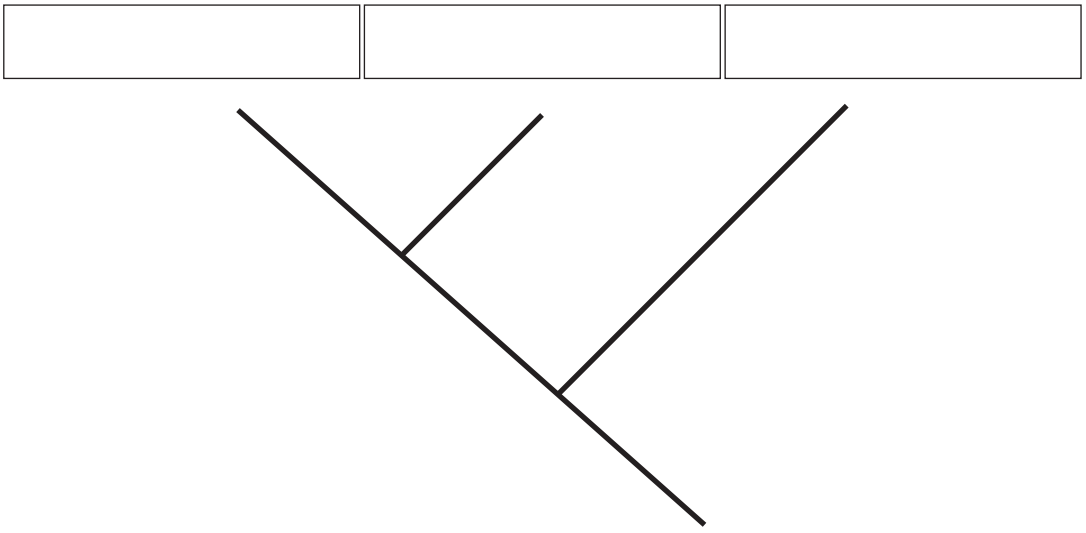
Question

Here's a tree – it shows you the relationship of the sequences from you, your relative, and a non-relative in the example above:

ATGCAG ATGAAG AGCTTC



Now consider how spinach, mushroom, and cows might be related to each other.
Fill in the boxes on this tree to show which two are more closely related:



Activity: Predicting relationships for our samples

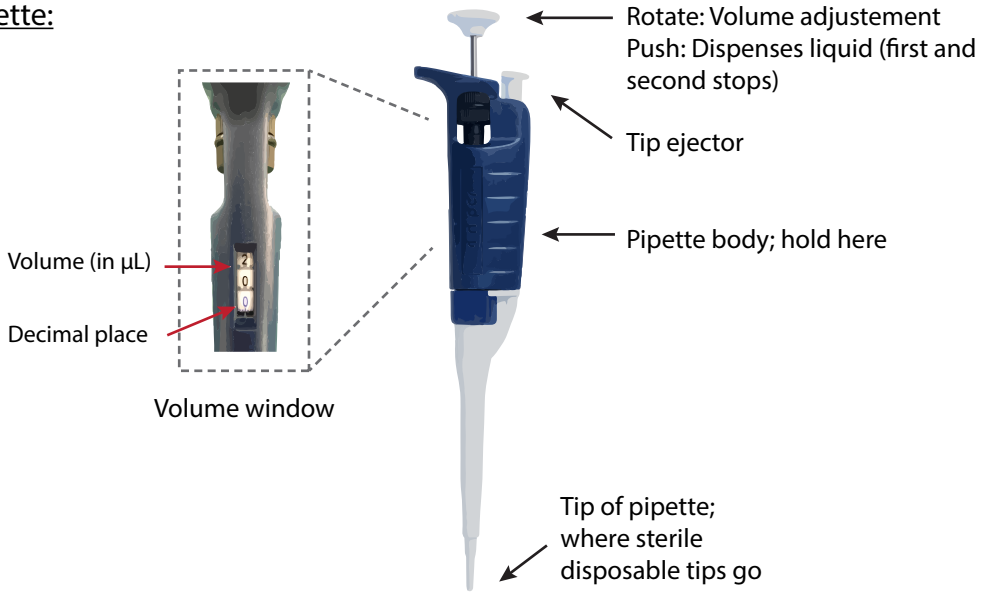
Now, extend what you have learned in the previous worksheet to all of the samples you will be using as a class, and draw your own tree below:

In the classroom

Using the samples that your students have chosen, have them create their own phylogenetic trees to predict the relatedness of the organisms that will be used in the lab. Have students work in lab groups to create their own trees, then decide on a consensus tree for the class prior to starting the lab. Have students record their predictions on this activity sheet, and keep the sheet for the duration of the workshop.

Activity: How to pipette

Parts of the pipette:



- [ ] (1) Pair up with a partner. One of you should have a p1000 pipette, and one a p200. Decide which one of you will measure out: 500µL or 50µL
- [ ] (2) Put a fresh tip on your pipette (place the tip of the pipette in the sterile tip in the box; push down gently to make sure it is on your pipette)
- [ ] (3) Once the sterile tip is on your pipette, push down the top button to the first stop and hold your hand there
- [ ] (4) Place the sterile tip below the level of the liquid, then let go of the top button slowly to suck up the liquid
- [ ] (5) Move the pipette carefully out of the liquid, and then over to your measurement card. Slowly press down the top button until you reach the first stop, then press down to the second stop.
- [ ] (6) Without releasing the top button, move your hand out of the liquid, and then let go
- [ ] (7) Eject the sterile tip into the waste bucket by pushing on the eject button

Important notes:

-- Do not lay down a pipette with fluid in the tip or hold it with the tip pointed upward, as this can cause damage to the pipette.

-- Once a pipette tip touches more than one liquid, it should be ejected and a new tip should be used.

In the classroom

Have your students look at the pipette and make predictions about what each part does. What do they feel when they depress the button? What moves when the buttons are pressed? What parts twist? What parts move when the top part is twisted?

Explain what the little windows mean; these correlate to the volumes that students are measuring. Have them move the pipette to different volumes to ensure they know where the decimal points are.

Protocol: DNA Extraction

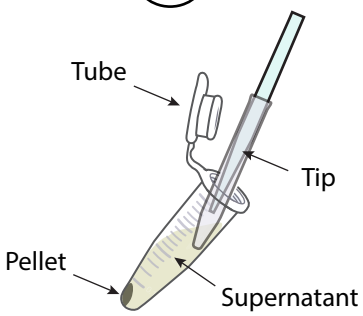
Find a classmate to pair-up with for this experiment. Before getting started, pick one tube containing a piece of tissue from one species-
make sure you and your partner have different samples!

Fill in the boxes below with the names of your samples:

Your sample:	Your lab partner's sample:

Follow the steps below.
Check the box off to the left as you go to keep track of which step you are on.

- ☐ (1) Add 100 microliters (µL) of Extraction Buffer to the tube
- ☐ (2) Use a blue pestle to mash the tissue in the tube
- ☐ (3) Close the cap of the tube, and place it in the water bath at 65 degrees centigrade (ºC) 
- ☐ (4) Wait for 8 minutes 
- ☐ (5) Add 14µL of Potassium Acetate solution to the tube and mix
- ☐ (6) Place on ice and wait for 8 minutes 
- ☐ (7) Spin the tube at maximum speed in the centrifuge for 15 minutes 
- ☐ (8)Take out the tube carefully, come back to your bench, and open the cap
- ☐ (9) Use a pipette to move the liquid **"supernatant"** to a new tube. Label this tube with your initials.
- ☐ (10) Add 47µL of Isopropanol to the new tube and mix.
- ☐ (11) Spin the tube in the centrifuge at maximum speed for 5 minutes 
- ☐ (12) Remove the liquid using a pipette, making sure not to touch the **"pellet"** of DNA



Protocol: DNA Extraction, Continued

- [ ] (13) Add 100µL of Ethanol to the pellet to "wash" it
- [ ] (14) Spin the tube in the centrifuge at maximum speed for 5 minutes
- [ ] (15) Remove the liquid using a pipette and leave the cap open to let the pellet dry 
- [ ] (16) Add 20µL of water to dossolve the pellet

UNIT 2: USING MOLECULAR BIOLOGY TECHNIQUES TO GET DNA SEQUENCES

Module 3: Using PCR to obtain DNA for a particular gene

Overview

In this module, students will learn about DNA complementarity and how to amplify a target sequence of DNA for further analysis. In a laboratory experiment, students will amplify a particular sequence of DNA from a variety of organisms (either provided samples, or their own samples generated in Module 2) using Polymerase Chain Reaction, or PCR. The product from this Module will be analyzed in Modules 4 and 5 with gel electrophoresis and subsequent DNA sequencing.

Learning goals

By the end of this module, students will be able to:

--Understand DNA complementarity
--Explain the action of DNA polymerase
-- Describe PCR (method and function)

Practical skills:

Learn how to use laboratory equipment, learn how to perform a PCR

Suggested prior knowledge

General structure of DNA
Preparation of genomic DNA

In the classroom

The pipetting introduction in Module 2 is essential to ensure the PCR protocol in Module 3 will work; since we are working with such small volumes, ensuring that your students are pipetting accurately and without contamination is key!

LAB SUPPLIES FOR MODULE 3

REAL SCIENCE Team will provide:

Equipment set-up:
Gloves
Pipettes (1 set/2 participants)
Pipette tips
Eppendorf tubes
PCR tubes (1/sample for each participant; small, clear plastic)
Tube racks
Centrifuge
Mini PCR machine

Solutions:
Molecular biology grade water
PCR reaction buffer
Nucleotide mix
Taq DNA polymerase
18S Forward primer
18S Reverse primer

In the classroom, teachers provide:

Ziploc bags
Shipping materials (envelope, tape, labels)
Freezer (for long-term storage of PCR products)
Ice

If your classroom opts to complete Unit 1, you will have samples from Module 2 that you can use for this portion of the workshop. Otherwise, samples can be provided for you.

Samples:
Genomic DNA (gDNA) from different organisms (i.e., previously prepared DNA from Module 2)

Worksheet: Learning about PCR

Polymerase chain reaction, or PCR, is a Nobel-prize winning discovery that allows us to make more copies of a specific piece of DNA. Each chromosome in our cells is a long string of As, Cs, Gs and Ts, and some regions of this string contain meaningful information. The most well-studied regions, called "protein-coding genes", contain information for making a protein, but some genes contain information to make an RNA that will become part of the ribosome, a large molecular machine that helps make protein. We will use PCR to make more copies of the 18S gene, which codes for a ribosomal RNA and is found in all eukaryotic cells, so it should be present in all DNA samples we are working with.

PCR harnesses the natural process of how DNA replicates to make more of itself when a cell divides. When a cell divides, it has to double the amount of DNA so that each daughter cell ends up with the same amount as the "mother" cell. Cells use an enzyme called DNA polymerase to copy the existing DNA. The two strands of the original DNA pull apart and the DNA polymerase uses them as "templates" to build a new string of DNA, matching As, Ts, Gs, and Cs on the original strand with Ts, As, Cs, and Gs respectively. PCR allows us to recreate this process in a plastic tube.

PART ONE: PCR INGREDIENTS

DNA (DeoxyriboNucleic Acid) is the material that carries genetic information. DNA is found inside each and every cell of our bodies (and those of other animals and plants too).

There are **five components of a PCR reaction** (see the figure on the next page):

(1) DNA template: The genomic DNA extracted in Module 2

(2) Primers: These are short strings of DNA (also called oligonucleotides) that are about 18-23 nucleotides long. They are designed to have similarity to the region of the DNA we seek to amplify. In our case, these primers will match with parts of the 18S gene in all of the DNA samples. They serve as "beacons" for the DNA polymerase - they call the enzyme to the correct region in the DNA to start doing its job of copying DNA. The specific primers in our PCR protocol are:

Forward Primer (18SF): ATGGTTGCAAAGCTGAAAC
Reverse Primer (18SR): GATCCTTCCGCAGGTTCACCTAC

(3) Deoxyribonucleotides , or dNTPs: a mix of many individual molecules of As, Ts, Gs, and Cs that are not strung together yet.

(4) DNA polymerase: an enzyme, or a protein machine, that can string together As, Ts, Gs, and Cs. DNA polymerase is found in nearly all organisms, but the particular type used in PCR comes from a bacterium called *Thermus aquaticus*, and is known as *Taq* polymerase. This particular DNA polymerase can survive high temperatures close to the boiling point of water, as it was isolated from deep-sea thermal vents!

(5) Buffer: A water-based mixture of salts that provide the right conditions for the DNA polymerase to work well.

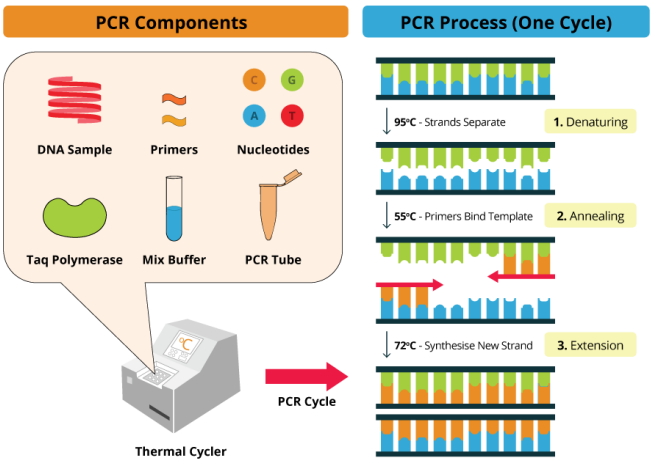
Question
Why is a special buffer, and not just plain water, used for the PCR?
[Hint: Think about where in the cell DNA polymerase normally works.]

PART TWO: PCR STEPS

There are **four major steps** in a PCR (see figure below):

(1) Denaturation: Denaturation is the word for the pulling apart of the two strands of DNA so that primers and polymerase can access the DNA template. This is achieved by heating the samples to a high temperature of 95C.

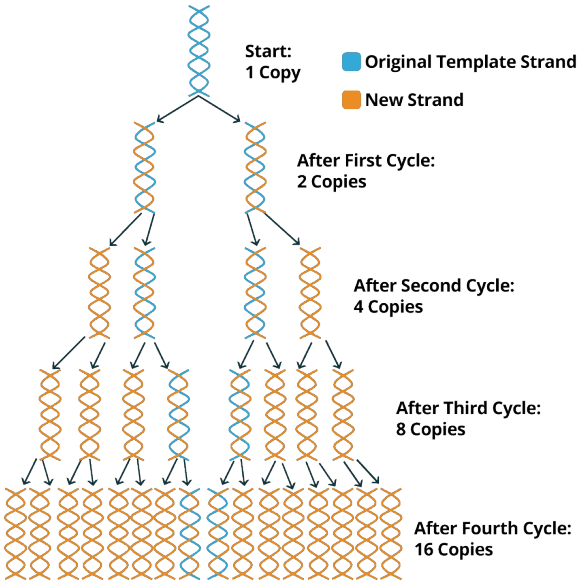
(2) Annealing: This step allows primers to pair up with their match in the DNA template. Too high a temperature would make the primers have too much energy to pair up at all, but too low a temperature would allow the primers bind to too many regions of the DNA template, even those that are not a perfect match.



Source: Bosterbio.com

(3) Extension: During this step, *Taq* polymerase will string together As, Ts, Gs, and Cs based on the template sequence. The temperature for this step is the one that the enzyme works best at.

(4) Repeat: The three steps above are repeated many times, each one called a cycle. The first cycle doubles the DNA, making double the number of templates available for the next cycle. With this doubling every cycle, the number of DNA molecules grows exponentially (see figure below).



Source: Bosterbio.com

Question

A PCR machine allows rapid changing of temperatures so that the samples can move quickly through the steps.

What would happen if there was a long delay in going from Step 1 (Denaturation) to Step 2 (Annealing)?

PART THREE: IMPORTANT CONSIDERATIONS FOR PCR

(1) Contamination: PCR is a very sensitive reaction as it can turn one molecule of DNA into millions over the course of a couple of hours. Even small amounts of contaminating DNA (e.g., from a pipette carrying DNA from your lab partner's sample) can obscure your results.

(2) Controls: Even the most experienced scientists make mistakes during experiments. Control experiments help ensure that the protocol was done correctly. A positive control is a sample that you know will work, based on previous experience. A negative control should give a negative result, but if it gives a positive result, it alerts you to possible contamination.

(3) Precision: Very small amounts of the *Taq* polymerase and template DNA are needed, so pay particular attention when pipetting these small volumes.

Question

Give an example of a positive control and a negative control that we could use for the PCR protocol during this workshop.

MODULE 3 Protocol: Amplification of the 18S gene using PCR

Follow the protocol below to set up your PCR reaction.
Each participant should set up one reaction using the DNA extracted in Module 2.

- [ ] (1) Obtain your DNA sample from Module 2.

- [ ] (2) Get a clean PCR tube and close the cap. Label it with your initials on the side of the tube.

- [ ] (3) Add the following things one at a time to the tube. Pipette up and down to mix and use a fresh pipette tip each time:

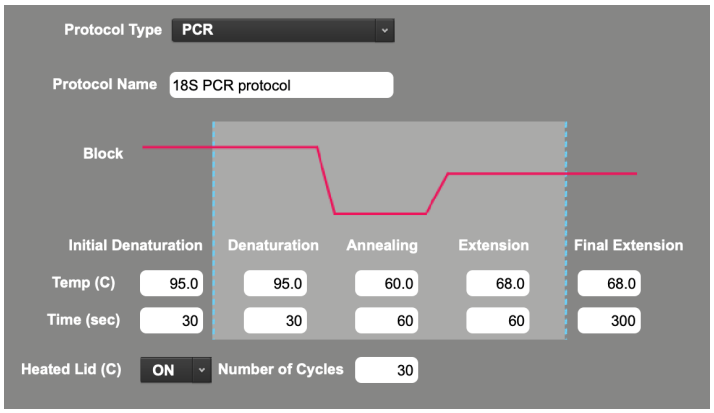
    - [ ] 40 microliters (µL) water
    - [ ] 5 µL PCR buffer
    - [ ] 1 µL dNTP mix
    - [ ] 1 µL 18S Forward primer
    - [ ] 1 µL 18S Reverse primer
    - [ ] 1 µL DNA polymerase
    - [ ] 1 µL of genomic DNA

- [ ] (4) Close the cap of the tube

- [ ] (5) Flick the tube gently to mix the contents, tap the tube on the bench lightly to collect the reagents at the bottom of the tube.

- [ ] (6) Take your PCR tube to the miniPCR machine and place it in one of the slots. Record here which position you put your sample:

- [ ] (7) We will program the machine as follows:



- [ ] (8) Once the reaction has finished, PCR samples can be stored in the freezer. During the workshop, they will be analyzed right away in Module 4.

UNIT 2: USING MOLECULAR BIOLOGY TECHNIQUES TO GET DNA SEQUENCES

Module 4: Using gel electrophoresis to check your PCR

Overview

In this module, students will learn the basics of gel electrophoresis and how it is used in the lab to measure the size and charge of DNA, as well as to determine if a PCR has run successfully. Participants will first learn how a gel electrophoresis apparatus works, and in a hands-on lab portion of the module, load and run their own gels. Additionally, we will utilize an active learning activity in which the classroom becomes a life-sized gel.

Learning goals

By the end of this module, students will be able to:
-- Understand the concepts behind gel electrophoresis:
- DNA can be separated by size and charge
- DNA can be visualized on a gel

-- Understand why gel electrophoresis can be used as a visual confirmation of how effective a PCR was

Practical skills:
Learn how to use a gel electrophoresis box

Suggested prior knowledge

DNA is a molecule that is negatively charged
PCR produces quantities of DNA that we can visualize on a gel

In the classroom

We will provide pre-prepared gels in your classroom on the day of this module.

In addition, we will provide detailed information for sending your samples off for sequencing. This sequencing will be free of cost for students!

LAB SUPPLIES FOR MODULE 4

REAL SCIENCE Team will provide:

Equipment set-up:
Gloves
Pipettes
Pipette tips
Eppendorf tubes for mixing loading dye with samples
Tube racks
Permanent markers

General lab supplies:
Gel rig (MiniPCR BlueGel) and power cord
Agarose gels with GelGreen (pre-made)
Waste container

Solutions:
Running buffer
Loading dye
DNA ladder

In the classroom, teachers provide:

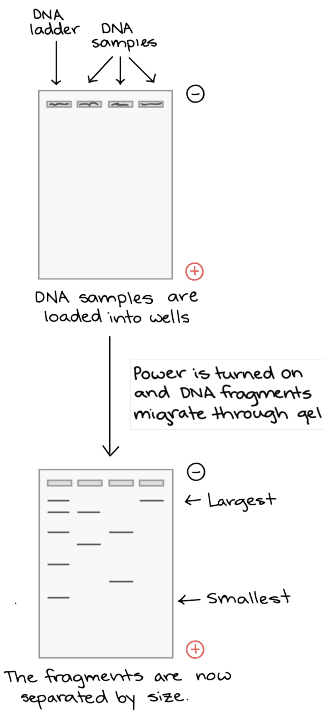
Envelopes for shipping samples
Ziploc bags for holding samples during shipping

If your classroom opts to complete Unit 2, you will have samples from Module 3 that you can use for this portion of the workshop. We suggest doing both Modules 3 and 4, as they build upon one another.

Samples:
PCR samples from Module 3

Worksheet: How DNA gels work

DNA is a negatively charged molecule. Gel electrophoresis relies on this property to separate DNA of different sizes. A small gel is made by mixing agarose (a sugar made from red seaweed) in a water-based solution and pouring it into a rectangular mold. The mold includes a comb-like structure with "teeth" that creates "wells" in the gel once it has solidified. DNA is loaded into these wells, the gel is submerged in a liquid solution, and a current is passed through the gel. Because of its negative charge, DNA will move towards the positively charged end of the gel. But the speed of the DNA will depend on its size - larger molecules of DNA tend to get stuck in between the sugar molecules in the gel but smaller molecules of DNA move more easily, and go further into the gel (see schematic from Khan Academy below). We will use this method to check whether our PCR worked.



Source: Khan Academy

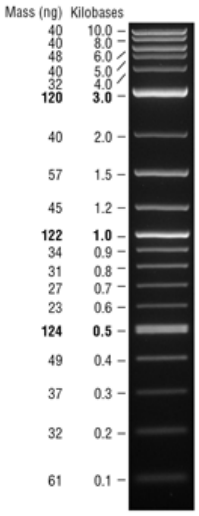
Here are some important aspects of running gels:

(1) Gel: The pre-made gels you will use contain a dye called GelGreen that binds DNA. This dye will glow under blue light and will show us where the DNA has traveled in the gel.

(2) Running buffer: The gell is submerged in a liquid called "running buffer" which allows a current to pass through the gel.

(3) Loading dye: The PCR reaction is a clear liquid and has the same density as water. To make it easier to see what we are doing we mix the PCR sample with loading dye. This colored dye allows us to see our sample as we pipette it into the gel. The loading dye also contains glycerol, which is heavier than water and helps the PCR sample settle into the bottom of the well instead of squirting out into the running buffer.

(4)DNA ladder: This is a mixture of DNA molecules of known size. When run on a gel, it creates a ladder-like pattern with the largest molecules of DNA at the top of the gel and the smallest molecules at the bottom. By seeing where the DNA from your samples lines up with the DNA molecules in the ladder, you can estimate the size of the DNA you are working with. The particular ladder we will use is called the Quick-Load® Purple 2-Log DNA Ladder and looks as shown to the right:

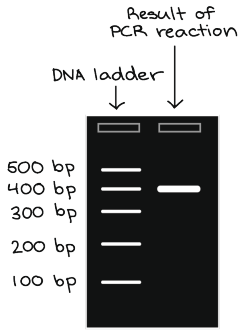


Source: NEB

(5) Checking a PCR: The main objective of Module 3 was to amplify a part of the 18S gene, the region bordered by the two primers we used. If the PCR was successful, we expect to see only the DNA of the correct size. This should appear as a band of about 1 kilobases (kb) or 1000 base pairs.

Question

Given the gel experiment below, what is the size of the DNA amplified in this PCR experiment?



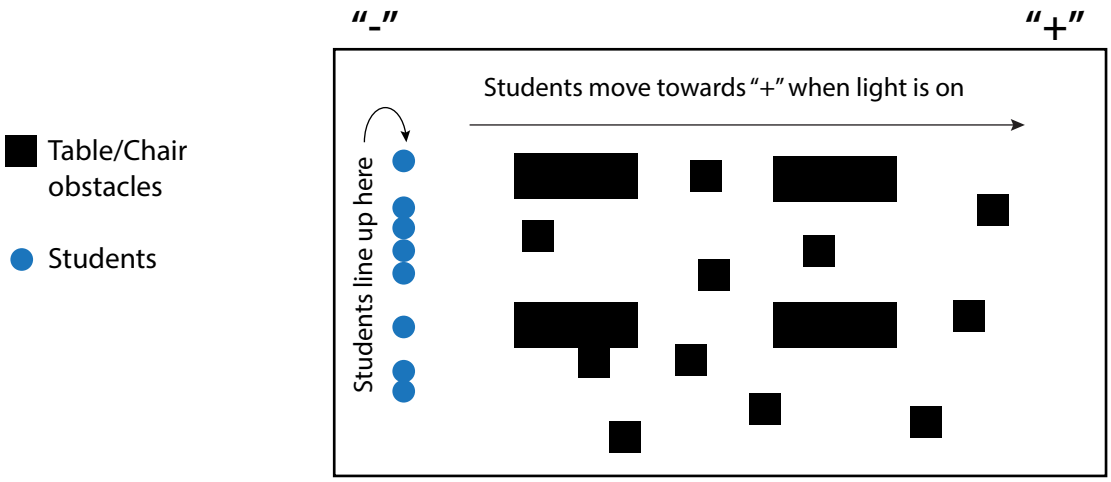
Source: Khan Academy

Activity: "The classroom is a gel!"

This is a hands-on activity that requires the entire classroom, and demonstrates how DNA moves from a negative (-) towards a positive (+) charged environment. It illustrates how larger bands of DNA take longer to progress through the gel than shorter bands. As part of this activity, the whole classroom is turned into a gel! The participants are the strands of DNA, the classroom chairs/desks/obstacles represent the polymers in the gel that slow the DNA down, and the lights in the classroom represent the electricity being turned on or off (if the gel is running or not).

Classroom Set-Up

Participants should line up at one end of the classroom, which serves as the "-" side of the classroom, and the opposite side is the "+". (It helps to write this on the board). Participants then link elbows to create various "sizes" of DNA bands. Create some 3-4 people-sized bands, and let some participants move through the "gel" without a partner as single person bands. When the lights are turned off, the gel is off and no one can move forward; once the lights are turned on, participants can walk towards the properly charged side of the room. Remember that a gel is not a liquid; the obstacles in the room represent the gel itself, which is a bunch of polymers.



Prediction questions:

(1) Who will move the fastest through the "gel"?
(2) Which direction on a gel does DNA run? Towards the "-" or "+" ?
(3) Will the obstacles in the gel make it easier or harder for some groups of students/bands of DNA?

Ready, set, go!

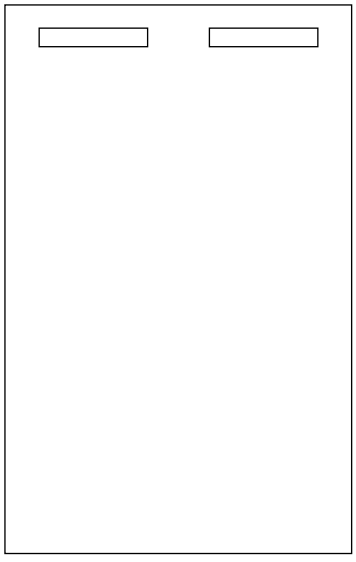
Protocol A: Running a DNA gel

- [ ] (1) Obtain your PCR samples.
- [ ] (2) Take a fresh eppendorf tube and label it "gel" and add your initials.
- [ ] (3) Take 5 microliters (µL) of your PCR sample, and pipette it into the tube labeled "gel"
- [ ] (4) Using a fresh pipette tip, add 1µL of DNA loading dye (purple liquid) to the "gel sample" tube
- [ ] (5) Close the cap to your "gel sample" tube
- [ ] (6) Flick the tube to mix all of the contents
- [ ] (7) Take your tube over to the pre-prepared gels
- [ ] (8) Using a fresh pipette tip, carefully load 6 µL of your sample into one well of the gel. Record which lane you loaded your gel in:

- [ ] (9) Once everyone has loaded their samples, we will load 5 µL of the ladder into the first lane of the gel.
- [ ] (10) Switch on the BlueGel machine and run the gel for 20 minutes.
- [ ] (11) Visualize your gel using the Blue Light. Do you see a band as expected for your sample? Record what your sample looks like here:

Sample name: Ladder

Drawing of sample:



Some tips and tricks for loading gels:

-- Once you have pipetted up your sample, inspect your pipette tip to ensure that you have the right amount of sample and no air bubbles.

-- Place your elbow on the table to steady your pipette hand, and if necessary use your other hand to steady the pipette as well (just make sure to not touch the tip!).

-- Find where your well is within the gel, and lower the tip until it is under the buffer and the point of the tip is resting gently in the well.

-- If you feel any resistance when resting the tip in the well, make sure that you have not punctured the agarose. If the agarose has been punctured, you can no longer use this well- move down to the next one and record the broken well on your lab protocol.

-- Press down the pipette to the first stop, slowly dispensing your sample into the well. You should see the well filling with your sample and loading dye.

Protocol B: Sending samples for sequencing

Once you have confirmed that your PCR has worked, i.e. it resulted in a single, clear band of the expected size, it is ready for sequencing!

We will be sending the sample to a local company, GeneWiz (genewiz.com), for sequencing. We will show you how to place an order for sequencing on their website, the most important aspect of this is keeping track of sample IDs.

1) We will generate a list of simplified tube names so that the technicians at GeneWiz can read them easily. Make a note of what your sample is and which ID it was assigned here:

2) Bring your PCR sample over to the station where samples are being prepared for shipping.

3) Pipette 10 µL of your PCR sample into a PCR strip tube that has the correct ID for your sample.

4)These PCR tube strips will be placed in a ziploc bag and dropped off for sequencing by GeneWiz.

UNIT 3: USING YOUR DATA TO RECONSTRUCT ORGANISMAL RELATIONSHIPS

Module 5: Phylogenetic analysis based on DNA sequence

Overview

In this module, participants will learn about how DNA sequences can be used to address questions in evolutionary biology. Participants will align DNA sequences from their own samples from Modules 3 and 4. Using publically available computational platforms, participants will use this sequence to confirm the identity of the species they chose. Next, participants will perform a phylogenetic analysis to infer the evolutionary relationship of their samples. At the end of this module, participants will revisit the prediction tree that was made in Unit 1, and determine if their results, based on empirical data, match their predictions. Participants will be asked to reflect on the similarities and differences between these two trees.

Learning goals

By the end of this module, students will be able to:
-- Perform DNA sequence alignment
-- Perform a phylogenetic analysis
-- Interpret evolutionary relationships based on phylogenetic analysis
-- Compare the inferred evolutionary relationships to the previously hypothesized relationships

Practical skills:
Learn how to use software to predict organism relationships

Suggested prior knowledge

18S DNA is contained in all eukaryotic organisms

In the classroom

If your submitted samples have come back with strange sequencing results, or you don't wish to send your samples for sequencing, we can provide the sequence files for this analysis.

LAB SUPPLIES FOR MODULE 4

REAL SCIENCE Team will provide:

Equipment set-up, Day-of:
Computers/Laptops with internet connection
Access the website:
http://www.phylogeny.fr/simple_phylogeny.cgi

Solutions:
None

Samples:
Sequenced DNA fragments in FASTA format

In the classroom, teachers provide:

Sequenced DNA fragments in FASTA format

If your classroom opts to complete Units 1 and 2, you will have sequences from Module 4 that you can use for this portion of the workshop. Otherwise, sequences can be provided for you.

MODULE 5 Protocol A: Organism identification by BLAST

Upon successful DNA extraction, PCR amplification, and gel electrophoresis to visualize our PCR product, we sent samples out for DNA sequencing. The sequencing company has determined the order of A's, C's, G's and T's that is unique to every sample we sent. Before we do a phylogenetic analysis, we will first make sure that the sequences we got back make sense using a program called BLAST.

BLAST uses a massive database of sequences of DNA from thousands of species to find a sequence that is like the one that you input. It can tell you how similar sequences are to one another and which species those sequences are from. Today, we'll be using BLAST to find which species contain a sequence most similar to your sample.

Step 1: Open the sequence file and find your sample's sequence

(1) On the desktop, find the file name "RealScience_DNA.fasta"

(2) Right-click on the file and choose the application "TextEdit" to open your sequence. The file is in a FASTA format which looks like this:

```
>AR_mushroom
ATTTGTACTGTGAAACTGCGAATGGCTCATTAAATCAGTTATAGTTTATTTGATGATACCTTGCTACATGGATAACTGTGGTAATTCTAGAGCTAATACATG
CCTTGTGCCGGCGATGCTTCATTCAAATATCTGCCCTATCAACTTTCGATGGTAGGATAGAGGCCTACCATGGTTTCAACGGGTAACGGGGAATAAGGGTTC
CAATATAGGGCTCTTTCGGGTCTTATAATTGGAATGAGTACAATTTAAATCCCTTAACGAGGAACAATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAA
>AR_beef
CTGCGAATGGCTCATTAAATCAGTTATGGTTCCTTTGGTCGCTCGCTCCTCTCCTACTTGGATAACTGTGGTAATTCTAGAGCTAATACATGCCGACGGGCG
TGACTCTAGATAACCTCGGGCCGATCGCACGCCCCCCGTGGCGGCGACGACCCATTCGAACGTCTGCCCTATCAACTTTCGATGGTAGTCGCTGTGCCTACC
CTCCCGACCCGGGGAGGTAGTGACGAAAAATAACAATACAGGACTCTTTCGAGGCCCTGTAATTGGAATGAGTCCACTTTAAATCCTTCCGCGAGGATCCAT
>AR_fish
AATCAGTTATGGTTCCTTTGATCGCTCTAACGTTACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGCCAACGAGCGCTGACCTCCGGGGATGCGTGC
ACGTCTCATTCGAATGTCTGCCCTATCAACTTTCGATGGTACTTTCTGTGCCTACCATGGTGACCACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGG
TTTCGAGGCCCTGTAATTGGAATGAGTACACTTTAAATCCTTTAACGAGGATCCATTGGAGGGCAAGTCTGGTGA
```

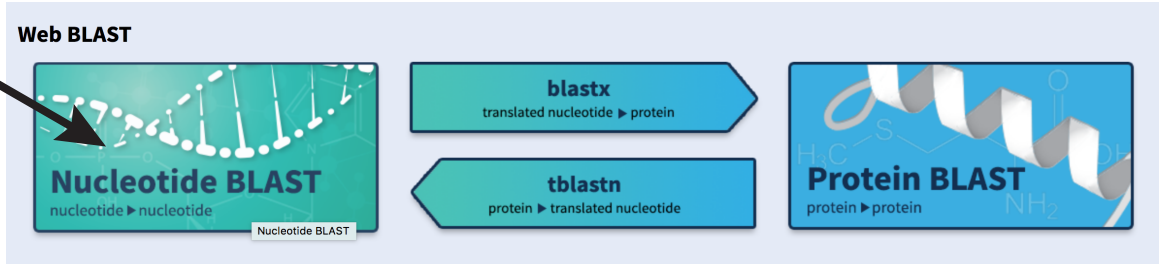
(3) FASTA files start with the '>' (greater than) symbol followed by a unique identifier (eg. >AR_mushroom) for each sequence. On the next line, the sequence corresponding to this identifier appears as a string of As, Ts, Gs, and Cs. The next '>' symbol indicates a new sample/sequence with a different identifier. Locate the two lines that correspond to your sample, based on the ID you noted at the end of Module 4. Copy and paste these lines into a new text file. It may look something like this:

```
>AR_mushroom
ATTTGTACTGTGAAACTGCGAATGGCTCATTAAATCAGTTATAGTTTATTTGATGATACCTTGCTACATGGATAACTGTGGTAATTCTAGAGCTAATACATG
CCTTGTGCCGGCGATGCTTCATTCAAATATCTGCCCTATCAACTTTCGATGGTAGGATAGAGGCCTACCATGGTTTCAACGGGTAACGGGGAATAAGGGTTC
CAATATAGGGCTCTTTCGGGTCTTATAATTGGAATGAGTACAATTTAAATCCCTTAACGAGGAACAATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAA
```

Step 2: Go the BLAST website, and load Nucleotide BLAST

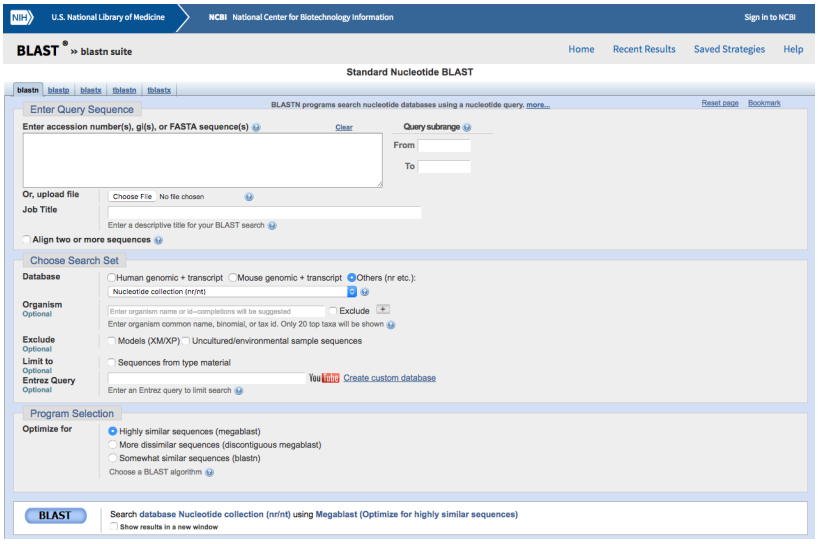
(1) Type the following address into your browser: https://blast.ncbi.nlm.nih.gov/Blast.cgi

(2) Click on "Nucleotide BLAST". This will tell the program that you are looking to match a DNA sequence with other DNA sequences in their database.



Step 3: Input your sequence and BLAST

Next, BLAST will ask you to fill in information so it can make the best prediction/conduct the best search it can to find a probable matching sequence.



(1) Copy and paste your sequence into the "Enter Query Sequence" box at the top of the page

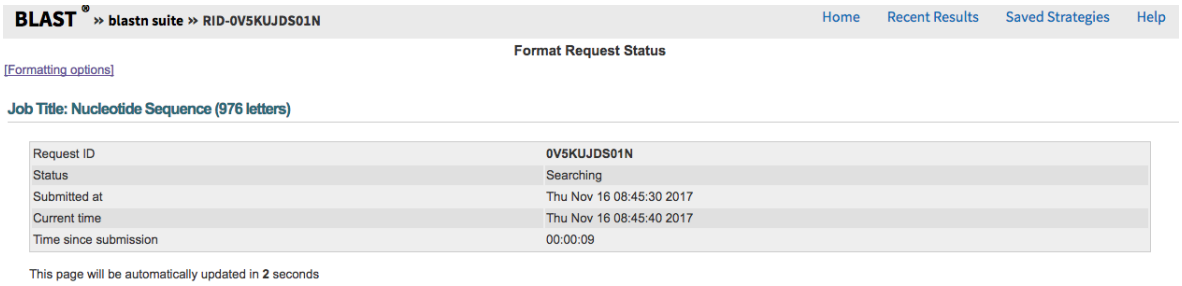
(2)In the "Choose Search Set" portion, use the default option for the Database: "Nucleotide collection (nr/nt)".

(3) The "Program Selection" input will determine how closely matched our sequence will be with the sequences in the database; we want a close match, so choose the "Highly Similar Sequences (megablast)" option.

(4) When you're ready, click  to start your analysis!

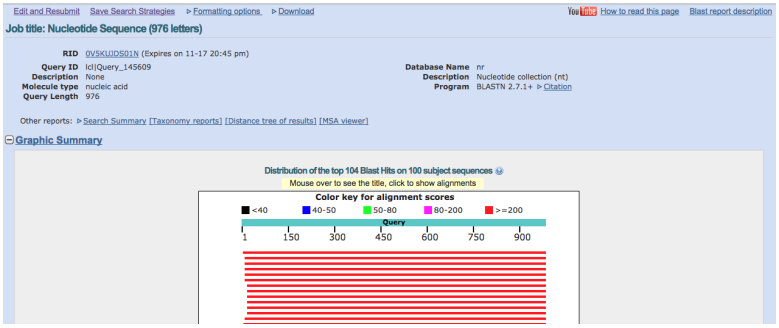
Step 4: Sequence homology analysis

A new window will pop up, showing that your search has been submitted. This window will stay open until your search has completed. This page will automatically update every 30 seconds, but the alignment can take up to ten minutes!

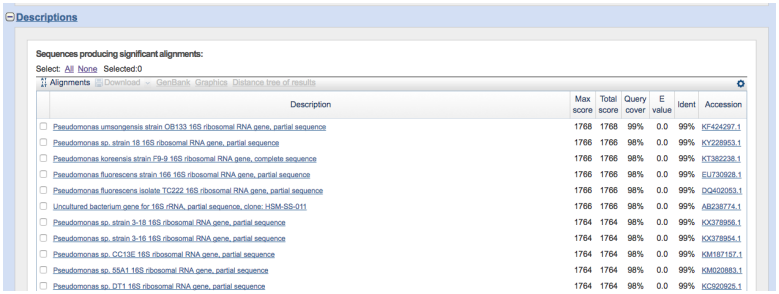


Once complete, your window will show the alignment and the best matches it has found for your sequence. It may look something like the image below, but this depends on your sequence:

The "**Graphic Summary**" shows how well the BLAST alignment worked. Each red bar indicates a match from the database to your sequence; if the bar extends along the entire sequence that your input (the "Query" at the top), it is likely a perfect or almost perfect match.

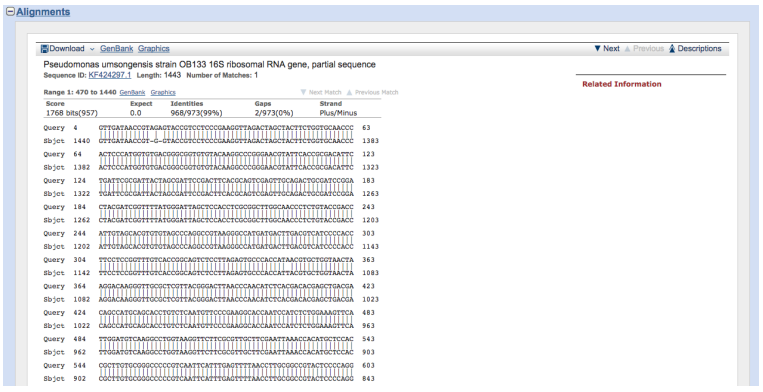


Scroll down further to find the "**Descriptions**" section. This contains a list of the sequences BLAST found as matches to your sample; the first few should be sequences from 18S genes in species closely related to the sample you chose for DNA extraction.



(1) Check the "**Query Cover**" and "**Ident**" columns here; the closer they are to 100%, the better the match. What are some of the species that the top ranking sequences come from? Are they all parts of the 18S gene?

(2) Click on the first link in the "**Descriptions**" tab. This will immediately take you to the nucleotide alignment in the "**Alignments**" tab of your BLAST search. For an alignment that is considered "great", you should find most the nucleotides in your sequence matching the nucleotides of the sequence BLAST has found (we call this the "BLAST hit"). Here's an example of a great alignment:



Query = sequence you input
Sbjct = sequence BLAST found

The lines between the A/C/T/Gs here indicate a match; where there is no line, the sequence you input has changed.

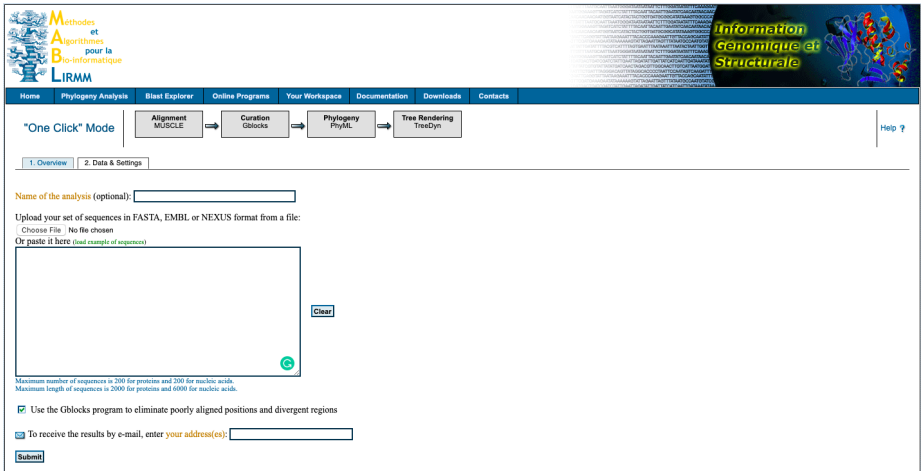
(3) If you have a great alignment, the name of the species in your BLAST hit is the species that you have isolated DNA from!

Write the name of this species here:

Protocol B: Phylogenetic analysis to generate a species tree

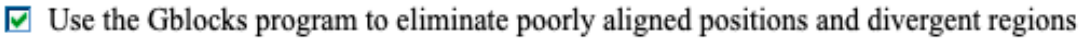
A phylogenetic analysis begins by lining up sequences for the same gene from many species. A program that can calculate similarity is then used to connect these sequences into a "tree" that depicts the relationships of the sequences. This tree then becomes a statement about the relationship of the species that these sequences come from.

(1) Using a web browser navigate to the following URL: http://www.phylogeny.fr/simple_phylogeny.cgi
The website should look similar to this:



(2) We will be operating in **"One Click"** Mode. Next, you can upload the FASTA file by clicking the **'Choose File'** button, then selecting the **'RealScience_DNA.fasta'** file.

Make sure the option is checked (small check mark) for using Gblocks:



You may enter your email address if you wish to have the results sent to you.

Finally, click the **'Submit'** button to perform the analyses.

(3) Once the analysis is completed, a page titled **'Tree Rendering results'** will be loaded. You have now completed an entire phylogenetic analysis with the click of a button!

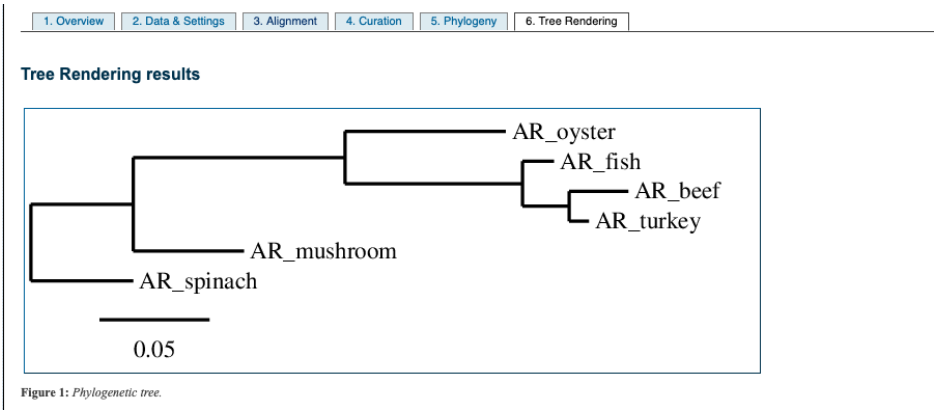


Figure 1: *Phylogenetic tree.*

(4) **But how did we get to this phylogenetic tree?** Let's walk through the steps leading up to the tree.

At the top of the web page, below **'One Click'** Mode, there is a series of numbered tabs (*1. Overview, 2. Data & Settings, 3. Alignment, 4. Curation, 5. Phylogeny,* and *6. Tree Rendering*).

(A) The first tab, labeled **'1. Overview'** shows an overview of the input and analyses performed on our nucleotide sequences (FASTA file).

(B) The second tab, labeled **'2. Data & Settings'** shows us our input FASTA file sequences.

(C) The third tab, labeled **'3. Alignment'** shows which nucleotides are conserved across samples (high similarity in cyan, medium similarity in blue, and low similarity in gray):

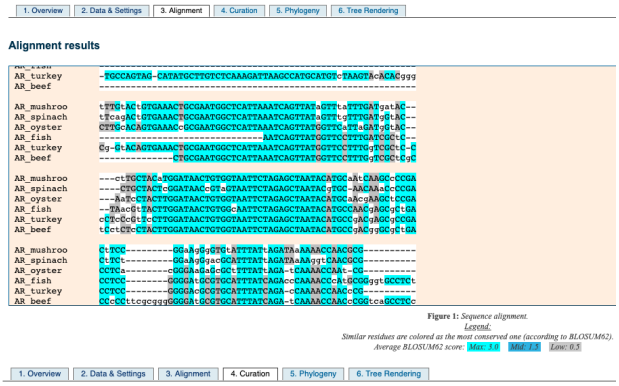


Figure 1: *Sequence alignment.*
Legend:
Similar residues are colored as the most conserved one (according to BLOSUM62).
Average BLOSUM62 score: Max: 11 Mid: 1.5 Low: 0.5

(D) The fourth tab, labeled **'4. Curation'** shows which nucleotides were used for the subsequent maximum likelihood analysis. These selected 'blocks' are underlined with a blue bar. In this example, nucleotide positions 130-240 are shown.

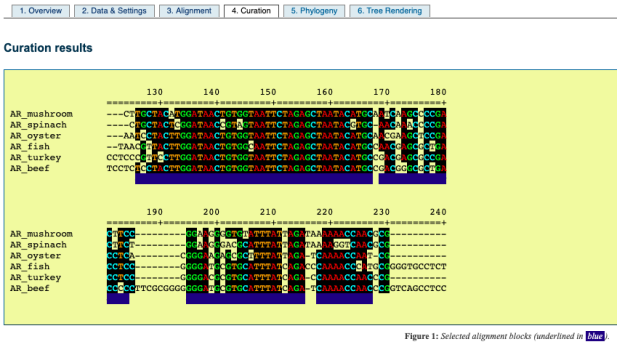


Figure 1: *Selected alignment blocks (underlined in blue).*

(E) The fifth tab, labeled **'5. Phylogeny'** shows a simplified version of the phylogenetic tree.

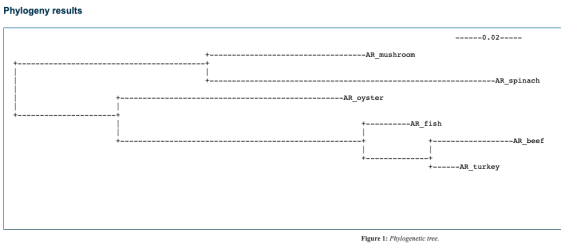
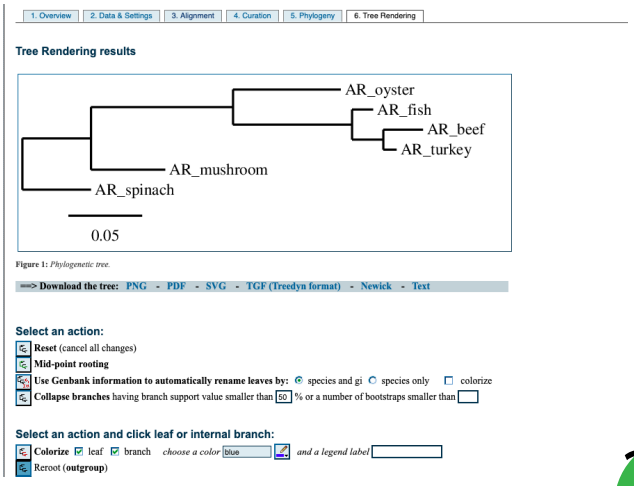


Figure 1: *Phylogenetic tree.*

(F) The sixth tab, labeled **'6. Tree Rendering'** allows us to interact with the phylogenetic tree. Play around with some of the options and ask us questions about what you see.



How do these relationships compare to your hypothesis from Unit 1?

Next Steps:
How to bring this workshop to the classroom

The units in this program have been developed so that they can serve as stand-alone lessons, or a series of lessons that build upon content from one module to the next. A graduate student volunteer from the Real Science team will work with you to figure out the best way to bring all or some of these modules to students in your classrooms. Volunteers will bring the materials to your classroom and help you run the protocols with your students. When you are ready to start planning this, or if you want to have a preliminary conversation to discuss the feasibility of doing this, send an email to:
Mansi Srivastava at mansi@oeb.harvard.edu and
Jennifer Cross Peterson at jenniferpeterson@hmsc.harvard.edu.

The list below indicates which modules are essential and which are optional for bringing this curriculum to your classroom.

Module 1: Introduction to Tree Thinking (essential)
The objective of this program is to help students understand evolutionary biology, in particular, phylogenetics via hands-on experiences with the kinds of experiments or analyses that researchers use. Your students will need broad introduction to why phylogenetic trees are important, and the concepts and materials used in this workshop can serve as a starting point for a lesson plan.

Module 2: DNA extraction (optional)
This module serves as a great tool to teach students about cells and DNA. If time constraints or interest only allows you to bring Module 3 to the classroom, we can provide pre-made DNA from the species studied in this workshop.

Module 3: Using PCR to obtain DNA for a particular gene (optional)
Module 4: Gel electrophoresis to check the PCR (optional)
Module 3 and 4 go together as the objective of running a gel is to check whether the PCR worked or not. These modules together will help students learn about molecular biology concepts and methods. If time constraints or interest don't allow you to bring these modules to your classroom, we can provide sequences from the samples from this workshop to your classroom to help them build phylogenetic trees in Module 5.

Module 5: Phylogenetic analysis based on DNA sequence (essential)
The goal of this module is to give students a hands-on experience in building phylogenetic trees using DNA on their own. Using publicly available programs on the internet to study DNA sequences will also highlight how powerful and accessible the tools used by researchers are.

Glossary and Acknowledgements

Glossary

DNA: Deoxyribonucleic Acid; the material that carries genetic information. DNA is found inside each and every cell of our bodies (and those of other animals and plants too).

Sequence: the arrangement of nucleotides (A's, C's, T's, and G's) in a string of DNA. We can use DNA "sequence" to figure out how people or, more generally, organisms are related to each other.

Phylogeny/"Tree": a drawing that depicts the relationships between species

Pipette Tip: a disposable plastic piece that allows us to re-use pipettes

Pipette: a device used to measure small liquids

Eppendorf Tube: a small tube with a cap

Centrifuge: a machine that spins tubes at high speed

Supernatant: the liquid that sits on top of a pellet

Pellet: the solid mass that you see at the bottom of a tube

Phylogenomics: the field of studying organismal relationships using genomic techniques; here, it is the study of defining how organisms are related by finding the similarities and differences in DNA sequences.

18S: the gene or "string" of DNA that is sequenced in this lab; this important gene helps our cells make proteins, and it does the same job in the cells of all of the species you will work on.

microliters/µL: a unit of liquid measurement. One milliliter is equal to 1000 µL.

Acknowledgements

This workshop is part of a 5-year program supported by the NSF (Award#1652104). We also thank New England Biolabs for generously providing PCR reagents free of cost and GeneWiz for providing discounted pricing for sequencing. Support from the HMNH education office, in particular the work of Jennifer Cross Petersen, Wendy Derjue-Holzer and Amy Gunzelman, was crucial for the conception and implementation of this workshop. The experimental and analytical aspects of this workshop and accompanying materials were designed and tested by graduate students Alyson Ramirez, Julian Kimura, and Ryan Hulett.

Additional Information and Suggested Reading

Additional Resources/Information

Khan Academy overview of phylogenetics
https://www.khanacademy.org/science/high-school-biology/hs-evolution/hs-phylogeny/a/phylogenetic-trees

"What did T-rex taste like?"; interactive reconstruction of evolutionary relationships
https://ucmp.berkeley.edu/education/explorations/tours/Trex/guide/index.html

Fun articles about phylogenomics:

Surprises From Placental Mammal Phylogeny 2: Skunks Are Not Weasels
https://blogs.scientificamerican.com/tetrapod-zoology/skunks-are-not-weasels/?print=true

How closely related are humans to apes and other animals? How do scientists measure that? Are humans related to plants at all?
https://www.scientificamerican.com/article/how-closely-related-are-h/?print=true

Surprises from Placental Mammal Phylogeny 1: Pangolins Are Close Kin of Carnivorans
https://blogs.scientificamerican.com/tetrapod-zoology/surprises-from-placental-mammal-phylogeny-1-pangolins-are-close-kin-of-carnivorans/