



Undergraduate teaching of scientific concepts using DNA barcoding of Trichoptera

JADE SHEVCHENKO¹, ACHMAD BAKHTIAR M YUNI¹ & PAUL B. FRANSDEN^{2,3,*}

¹Department of Ecology, Evolution, and Natural Resources, Rutgers University-New Brunswick, New Brunswick, New Jersey 08901, U.S.A. Email: jadeshev@gmail.com and achmadbakhtiar@gmail.com

²Department of Plant and Wildlife Sciences, Brigham Young University, Provo, Utah 84602, U.S.A. Email: paul_frandsen@byu.edu

³Data Science Lab, Smithsonian Institution, Washington, D.C. 20002, U.S.A.

*to whom correspondence should be addressed: paul_frandsen@byu.edu

Abstract

Student instruction is often segregated into specific subjects, giving the impression that different areas of study are largely unrelated. Teaching in “silos” inhibits the formation of connections among subjects. An academic curriculum that integrates DNA barcoding can expose students to a wide variety of scientific concepts, technical skills, problem solving, and fieldwork. In an approach that emphasizes conceptual interconnections, students are instructed in methods of caddisfly collection, specimen vouchering, DNA extraction, PCR amplification and sequencing of the COI barcode fragment, data curation in the Barcode of Life Data Systems (BOLD), and phylogenetic analysis. Accompanying development of these technical skills are learning objectives which include recognizing that DNA comes from real organisms and consists of molecules that are a physical part of those organisms, rather than being an obtuse symbolic abstraction. Students learn to logically pull together concepts from multiple scientific fields, use them in a practical setting, and experience science more broadly as a unified and interconnected discipline.

Key words: teaching, Trichoptera, collection, PCR, DNA sequencing

Introduction

DNA sequencing is a multi-step process that integrates valuable technical skills and biological concepts from different scientific fields. Often there is a divide between organismal and molecular biologists, or those who work in the field and those who operate in the lab. The lab specialist may have limited knowledge of the once-living specimen in the plastic tube, while the field expert may be unaware of the molecular and chemical characteristics of the animal. Lessons learned from DNA sequencing take a student from field collection to generating and analyzing DNA sequences. By continually cross-referencing the conceptual aspects of laboratory techniques, students gain problem-solving abilities. Creative and critical thinking in the laboratory requires an understanding of chemical and physical properties of the substances that are being manipulated. Effective teaching can turn abstract concepts that rely on symbolic representations into an understanding that DNA is a real thing that has a shape, occupies space, and responds predictably to chemical and physical manipulation. Undergraduates can also learn technical skills, such as pipetting and making solutions, which are commonly practiced in many laboratories.

Through instruction in DNA barcoding of caddisflies, students are introduced to the complexities of species identification, made more difficult by the vast diversity of insect life. DNA barcoding uses a standardized 658-nucleotide fragment of mitochondrial DNA from the cytochrome oxidase I gene (COI) to identify species (Hebert *et al.* 2003). Caddisflies were chosen because the Trichoptera Barcode of Life

initiative now has sequences from 58,000 individuals, representing over 5,000 species, and is large enough to identify most North American caddisflies to the species level. Students are introduced to the Barcode of Life Data Systems (BOLD; <http://www.boldsystems.org>), and to the promise and limitations of DNA taxonomy. Here, 10 one-hour lessons associated with teaching DNA barcoding are discussed. Each phase of the barcoding process is detailed with the lesson's rationale, teaching approach—including technical skills, and learning objectives.

Lesson 1: Collecting in the field

Rationale: Students who think that all science happens in a lab need to understand that DNA exists in living organisms. Many students have never been camping. Others have not experienced the vast diversity of insect life. Barcoding is a means to estimate and describe diversity. It is important to take students into the field.

Methods: Caddisflies can be collected using the standard methods and devices used in aquatic macroinvertebrate surveys. In general, larval, pupal and adult specimens should be collected to illustrate local and regional species diversity. Larvae can be collected from streams using an aquatic D-net, held in a strong current while upstream substrate is overturned and disturbed (Fig. 1). The D-net can also be passed through submerged vegetation, either growing in standing water or attached to rocks in currents. Specimens from sandy deposits are best collected with small hand-sieves and strainers. These methods yield a diverse number of species when the water is clean. Larvae are best preserved in ethanol. Adult caddisfly specimens can be collected during the day around vegetations close to streams by using a standard sweep net. Adults are best collected at night by light trapping, using a UV light source, powered by a car battery and coupled with either a white sheet or a pan (Fig. 2). A jar with a killing agent such as ethyl acetate can be used to manually collect caddisflies from the light trap. Captured adult specimens can either be pinned or preserved in 95% ethanol. Both preservation methods work well for DNA analysis (Holzenthal *et al.* 2015).



FIGURE 1. Collecting aquatic larval specimens with a D-net (center).

Learning Objectives:

1. Compare specimens captured using D-net vs. light-trapping. There are opportunities here for discussing ecological and developmental strategies of subdivided habitats and aquatic vs. terrestrial ecosystems.
2. Compare the advantages and disadvantages of using ethanol for storing as wet samples or storing as pinned/

dried specimens. Both work well. The presence of water is detrimental because enzymes that break down DNA are active in solution. Both methods remove water.

3. Understand that there are many populations interacting with one another within an ecosystem.
4. Understand that insect life is diverse, and clean water is full of bugs.
5. Recognize that field work creates a bonding experience where students and professors can communicate, interact, and discuss science. Seeing one-another as individuals in a beautiful setting is an important interpersonal part of the educational experience.



FIGURE 2. Collecting adult caddisflies at night with a UV light trap in front of a white sheet.

Lesson 2: Insect diversity

Rationale: The complexity of species identification is often underestimated. Most non-specialists underestimate the diversity of life, and assume that experts can simply look at a caddisfly larva, consult a field guide, and identify it. That this is false will be obvious to students as they attempt, and struggle with the task.

Methods: In the lab, ask students to try identifying the specimens they have collected. Simply learning that this is not possible is a valuable lesson in itself. Find a box of insects with many species that look similar (as in Fig. 3). Ask them: “How many species are in the box?” They might say 5 in the example shown in (Fig. 3). However, there are at least 35 species of fritillary in this box. Explain that this is not an unusual example: life is subdivided into many distinct species that do not always look phenotypically distinct from a distance. Lectures on species concepts and speciation would be presented at this point.

Learning Objectives:

1. Explain that it is not unusual for life to be subdivided into many distinct species that do not always look phenotypically distinct from a distance.
2. Describe species concepts and speciation.



FIGURE 3. Teaching box meant to pose the question: “How many species are in the box?”.

Lesson 3: Vouchering and photographing specimens

Rationale: Most science involves careful note-taking and record-keeping. Museum work is particularly meticulous with record-keeping.

Methods: Each specimen is photographed, and labeled with a specimen ID (Fig. 4) that includes a short and unique code that is permanently associated with the individual specimen. Locality labels are given to each individual. The specimen is permanently stored in a museum as a voucher specimen, with a voucher label for cross checking any future discrepancies or inquiries. Explain that a voucher specimen allows future workers to check the exact specimen for morphological details, and permits future checks on the accuracy of identification.

Learning Objectives:

1. Learn how to handle camera equipment and various methods of photography, including photostacking methods.
2. Develop the organizational skills of labeling, storing, and keeping track of specimens.
3. Maintain a legible, detailed, and organized lab notebook.
4. Recognize the important morphological structures used in identifying and classifying specimens by looking at insects through the microscope.



FIGURE 4. Voucher photograph with the specimen ID.

Lesson 4: Extracting DNA

Rationale: Students should feel competent in the lab. It is important to show them that the lab is not a scary and intimidating place. If they follow directions, they can succeed. It is similar to following a recipe in the kitchen. Extracting DNA provides a platform to talk about cell biology, and the organelles in the cell that house DNA, as well as the enzymes that degrade DNA. There is an opportunity to teach about enzyme cofactors, such as magnesium, as well as chelating agents, such as EDTA.

Methods: Students need to practice pipetting at an accurate scale. They should be able to pipette 10 μ l into a weigh boat, and it should weigh 10 μ g. Have them practice until they can pipette consistently. For DNA extraction, the caddisfly is placed in a glass dish, with ethanol, under a dissecting microscope. The leg of the caddisfly is grabbed at the base of the coxa with fine forceps, and pulled distally to retrieve the abundance of mitochondrial DNA located in the muscle attached to the coxa, and extending into the thorax. The leg is then put into a 96-well plate, in a specified position, or in individually labelled PCR tubes. The legs at the bottom of each well of the plate (or tube) are treated with 25 μ L of 25 mM NaOH and 0.2 mM of EDTA and incubated in a PCR machine at 95°C for 90 minutes according to the “Hotshot” method (Truett *et al.* 2000). Next, a 25 μ l solution of 40 mM of tris-HCl is added to neutralize the solution and prevent further degradation of the DNA. The resultant liquid is now ready for DNA amplification (PCR), and can be stored in the freezer. This method is fast, inexpensive and easy, but is not ideal for the long-term storage of DNA. Search the web with the terms "eukaryotic cell", "mitochondrion", and "nucleus" for instructional images for class.

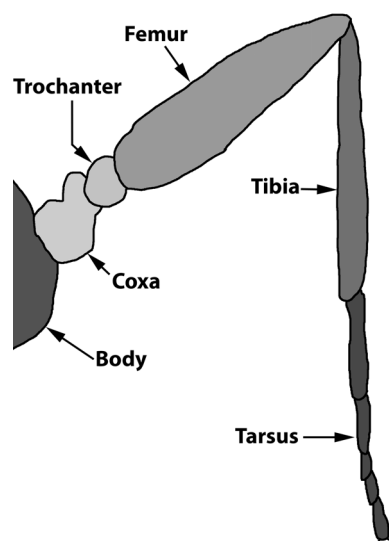


FIGURE 5. Insect leg (Wikimedia Commons 2015).

Learning Objectives:

1. Learn about the metric system including common measurements, units and their conversions.
2. Learn about the basic techniques of handling a micropipette.
3. Learn how to produce salt solutions of precise molarity and how to calculate dilutions.
4. Recognize the basic anatomy of an insect leg which has muscles in the coxa that extend into the thorax (Fig. 5).
5. Be aware that DNA in animal cells is located in the nucleus and the mitochondria.
6. Understand muscles are full of mitochondria because they are metabolically active.
7. Explain why the nucleus contains the nuclear DNA, while the mitochondria contain their own genome. The result is from an ancient symbiotic phagocytic engulfment of a prokaryotic organism which has now lost most of its genome.
8. Understand the function of lysosomes. They are present within the cell and contain enzymes that can degrade DNA. These enzymes are released into the solution during the extraction process.
9. Understand the role of EDTA which is a common additive to solutions in DNA lab and is used to protect DNA from the nucleases released from the lysosome. EDTA chelates magnesium, which is a cofactor for many enzymes, including nucleases.

Lesson 5: PCR amplification of the COI barcode fragment

Background: A 658-nucleotide fragment of the COI gene, located in the mitochondria, is the standard chosen by the Barcode of Life initiative (Hebert *et al.* 2003) and other sequence databases due to the presence of relatively conserved universal primer sites, and sufficient variation to distinguish among populations of animal species (Fig. 6, right panel). Because there are hundreds of copies of each mitochondrial gene in muscle cells, the COI is relatively easy to amplify with PCR. Random errors in the replication of DNA (mutations) are observed as substitutions (mutations that have survived the filter of selection) in the sequences. The segments of the DNA that code for crucial physical characteristics are less likely to change and thus some segments and some codon positions are more conserved over time. In contrast, those specific sites that do not have much impact on advantageous and important behavioral or morphological traits (such as 3rd codon positions) will have a higher substitution rate. An ideal target region has both conserved primer regions, but also contains enough variation within the gene to be able to align the data, and differentiate among species.

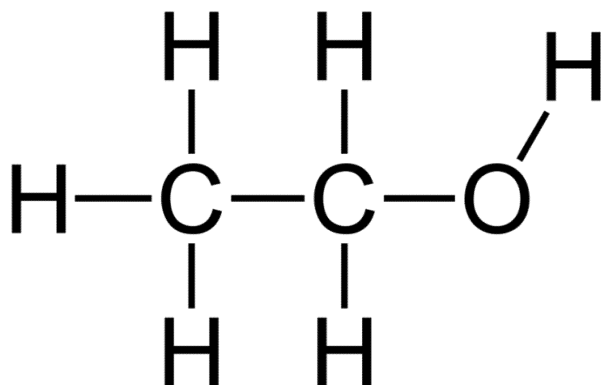
Rationale: The main focus of optimizing PCR from a teaching perspective is the importance and process of problem solving or critical thinking. Students must be able understand the many possible reasons for failure, and narrow down those possibilities to find and apply the required solution(s). It is also important to recognize how PCR amplification of a sample's DNA is an important step in the entire DNA sequencing processes.



FIGURE 6. A short fragment of three hypothetical loci, changing at different rates through time. Top sequences are ancestral. Bottom sequences are how those sequences changed over time. Asterisks mark sites that differ between the original, and the evolved state. Left panel has too much variation, such that we are not even confident of alignment or homology across sites. Middle panel does not enough variation, such that there is no signal. Right panel has the appropriate variation for phylogenetics.

Methods: The extracted DNA from the caddisfly is prepared for the PCR machine which amplifies the target DNA sequence. A master mix is made by combining specific oligonucleotide primers, and water. Most PCR is now done with commercial kits, and these kits contain *Taq* polymerase, magnesium, and dNTPs. Our reactions are in 25 ml volumes. We recommend the PuReTaq Ready-To-Go PCR beads (GE Healthcare), because although they are relatively expensive, they are reliable, simple to use, and only require the addition of primers, water, and DNA template (from Lesson 4). During PCR, the DNA is put through 35 cycles at 94°C (30 s), 50°C (30 s), and 72°C (2 min). There can be variations in these values based on the specific annealing temperature of the primer and the target DNA nucleotide composition.

A.



B.

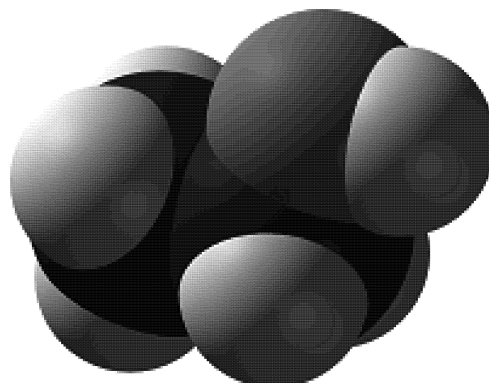


FIGURE 7. Symbolic representations of ethanol. (A) Atoms indicated with letters, with bonds indicated by lines. (B) Molecule drawn as a shape (Wikipedia: The Free Encyclopedia 2016).

Learning Objectives:

1. Learn how specific genes are targeted, and amplified with PCR.
2. Describe each step in the PCR, including melting (94 °C): the high melting temperature causes hydrogen bonds in DNA to break (denature), annealing (50 °C): the primer sticks to (anneals) the target, and extension (72 °C): the optimal temperature for *Taq* enzyme to function.
3. Know that DNA is notated as 5' to 3' by convention. It has a sugar-phosphate backbone. A growing strand of DNA extends from an active hydroxyl group on the 3' end.
4. Recognize that PCR uses the same process and mechanism of DNA replication in vitro.
5. Understand that “heat” is a concept of the speed at which molecules are moving.
6. Know that water molecules colliding with sufficient energy (speed) can disrupt hydrogen bonds, and denature DNA.
7. Realize primers collide and bind to the target region by random chance.
8. Understand that the success of PCR often varies from good quality results to no result at all. Many different factors affect the outcome.
9. Recognize that errors are usually correlated with the quality of the DNA, the match of the primers, or technique (errors made during the procedure).

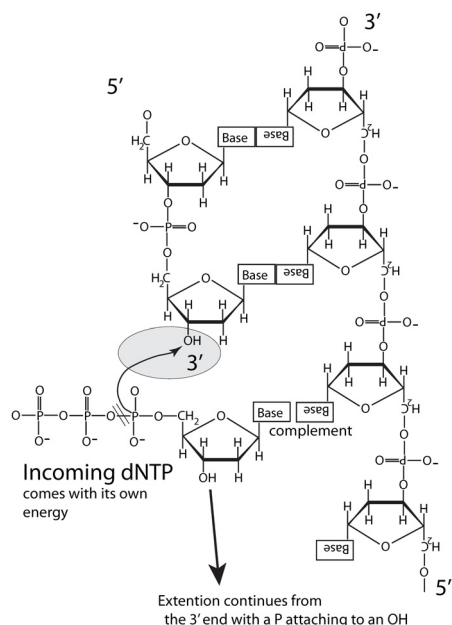


FIGURE 8. DNA structure, showing the importance of the 3' hydroxyl group. The incoming dNTP in solution finds a complementary base and attaches to the growing DNA strand from the 3' end, releasing a diphosphate group.

10. Understand how chemical structures are illustrated. For example, ethanol can be written as C_2H_5OH . This is a convenient shorthand. Alternatively, ethanol can be illustrated as in (Fig. 7a) or (Fig. 7b). Each has advantages. The symbolic representation can symbolize single-bonds, and is easier to count how many atoms make up a molecule. The three-dimensional model encourages students to think about shapes bouncing around in water. Discuss how heat is defined in terms of the speed at which molecules are moving, and the energy of collisions is familiar to us. Massive objects moving at high speeds (like a fast-moving truck) collide with other objects with greater energy than slow moving, less massive objects (like a ping pong ball, tossed underhanded). These principles of energy and motion should be discussed in terms of molecular interactions in water, related to melting and annealing temperatures of DNA.
11. Understand how the 5' and 3' orientations run in opposite directions in DNA (Fig. 8) shows why you need a primer. A primer is what it says; it "primes" the reaction for the addition of another nucleotide. (Fig. 8) shows the active hydroxyl site (grey circle). Nucleotide triphosphates attach to the hydroxyl site in an energetically favorable reaction. Students will recognize how only the 3' end can receive additional nucleotides. Mention that Deoxyribonucleic acid (DNA) is missing a hydroxyl group on the second carbon of the ribose sugar (numbered clockwise, starting with the "O" at the top). RNA has an OH there. Di-deoxy nucleotides are missing both OH groups, which will be important in discussing Sanger sequencing. Students will remember that mitochondria are the "powerhouse of the cell" from elementary school. But what does that mean? (Nothing, until they understand that mitochondria produce ATP, which they were taught is an "energy molecule"). Cleaving the 2 phosphates off the dNTP produces energy, so that no additional energy is required for the extension of DNA. The "energy molecules" are dNTPs, and they carry with them the fuel to catalyze the reaction. Taking the time to explain chemical reactions this way can move students toward realizing that chemistry is not based on abstract symbols, but happens in solutions to real things. The symbols are just shorthand for making illustrations like (Fig. 8). This would be a good time to compare PCR with DNA replication. Do a web search for "DNA replication", "Okazaki fragment" and "DNA polymerase" for illustrations and topics.

Lesson 6: Agarose Gel Electrophoresis confirmation of PCR amplification

Rationale: This phase allows students to see and analyze the work they have done, both quantitatively and qualitatively. They will have to make decisions about the results and determine whether it is of quality or needs to be thrown out and done again. Often times, the results that are collected from an experiment are not going to be clear cut or may not even work, but that is a part of the scientific process. Students will be introduced to the importance of using information to find ways to check their progress. Also, this would be a good time to discuss the properties of DNA.

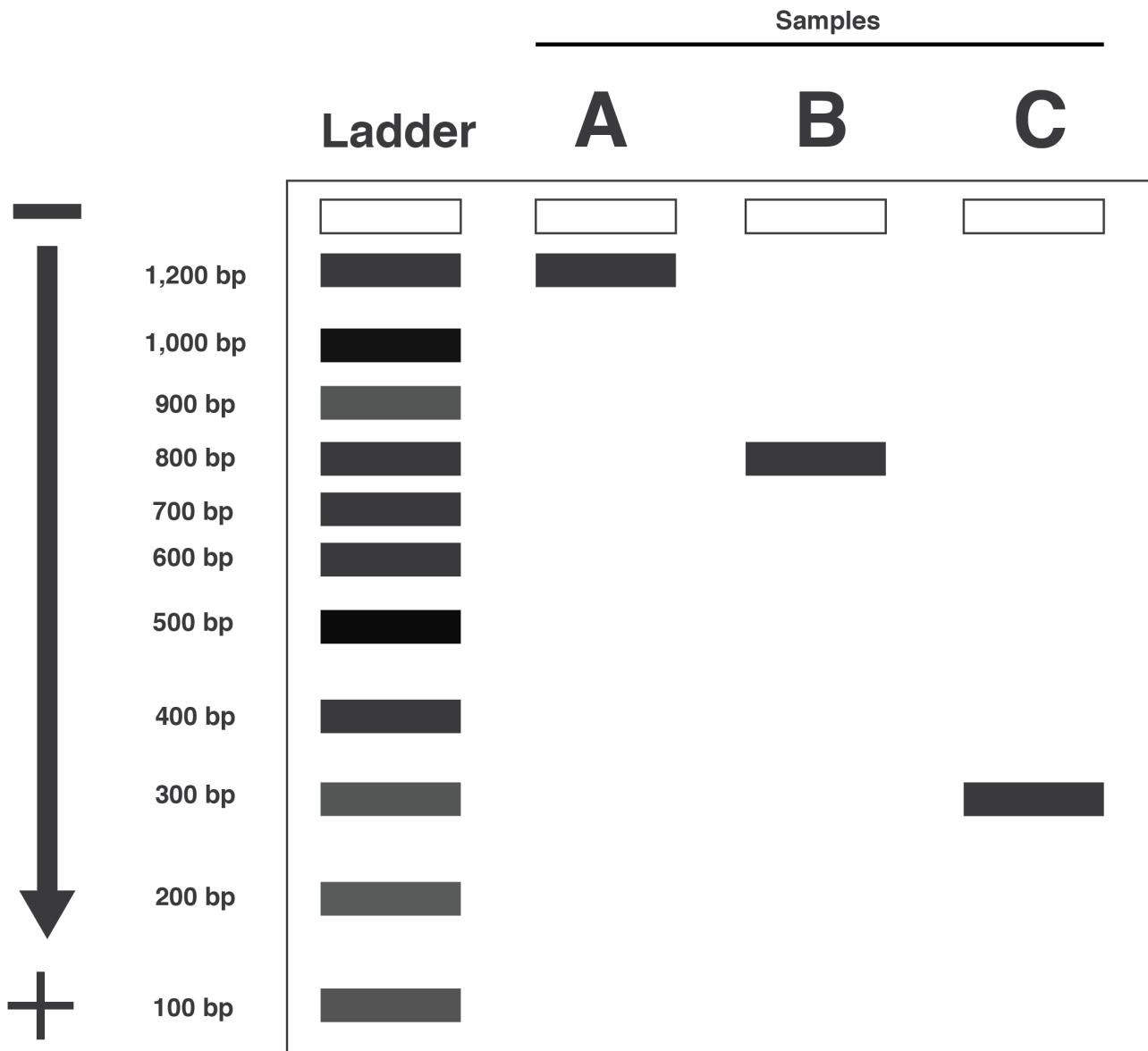


FIGURE 9. Model of gel electrophoresis. Note the mass ladder in the first well.

Methods: The amplified DNA is run through agarose gel electrophoresis to see if the process worked. A 1% solution of agarose is made in a solution of Tris, acetic acid and EDTA (TAE). The solution with agarose is boiled to dissolve the agarose and then SYBR Safe (ThermoFisher Scientific) is added as per manufacturer's instructions. The gel is poured into a mold, and cooled at room temperature for 15 min. Once set, and placed in an electrophoresis box, 5 μ l of DNA mixed with 1 μ l of loading dye is placed into the wells. Load a mass ruler, such as BioRad's "EZ Load Precision Mass Ruler" (BioRad #1708356) onto the gel as well. Run the apparatus at 120 V, making sure the negative pole is at the back. After completion, place the gel under

an ultraviolet light to qualitatively determine the concentration of the DNA, which depends on the brightness of the band, relative to the ladder (Fig. 9). Remember to put on proper eye protection before analyzing the gel. We categorize each band as either 0 (absent), 1 (faint), 2 (medium), or 3 (bright). For example, in (Fig. 10), intensities, from left to right are recorded as 1,1,1,0,0,3,2,2. Target bands of DNA are in the middle and faint bands below are primer dimers.

This value will be used in the sequencing step (Lesson 7).

Learning Objectives:

1. Discuss the chemical and physical properties of DNA that determine the speed at which the DNA travels in the gel.
2. Understand that DNA has a negative charge that is directly proportional to its length.
3. Recognize how size affects distance traveled in a set time.
4. Learn how charge affects the speed of movement through the small, evenly spaced openings in the gel.
5. Recognize how a mass ruler, prepared by digesting a plasmid into pieces of known size can be used to estimate both size and concentration of the PCR product.
6. Understand some principles of agarose gel electrophoresis will be used to separate synthesized oligonucleotides during the sequencing step (Lesson 7).
7. Understand the various factors affecting gel electrophoresis. Distribute and discuss the following illustrations: Refer to (Fig. 8) and ask "How many negative charges on each nucleotide?" (Count the negative signs. There is a single negative charge for each nucleotide). When they say "1", then ask how many negative changes on 247 nucleotides (247). Make the point that with DNA, the charge is linearly proportional to the size. This is important because the speed at which molecules move through a gel is influenced by two things: the size of the molecules, and the "push" (or pull) it gets in an electric field, which is based on its charge. Draw on the board to indicate a protein with some plus signs and negative signs. Add up the net charge. Then draw another one, slightly smaller, but with fewer negative signs than the larger one. Ask which one wins the race through the gel: the larger, more negative one, or the smaller, less negative one. Mention that the separation of proteins by size is more difficult because size and charge interact with one-another. However, with DNA, because charge is directly proportional to size, we can precisely separate it by size. Discuss primer-dimers, in a review of the PCR. Note, when the reaction fails as in the two middle lanes of (Fig. 10), extra reagents may produce a brighter primer dimer (hybridized or attached PCR primers).

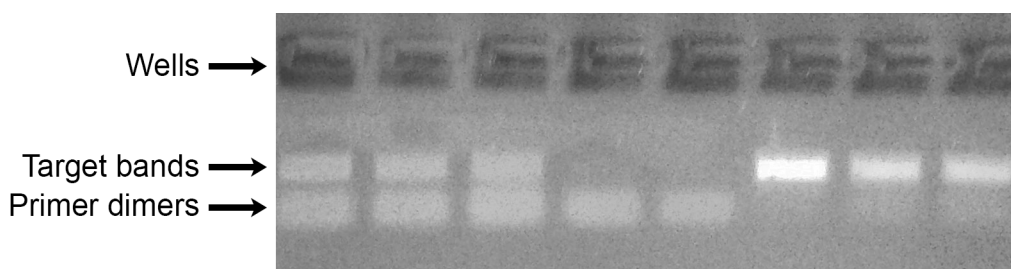


FIGURE 10. Agarose gel under UV light. Wells are at the top. Target bands of DNA are in the middle. Brighter bands contain more DNA. Faint bands below the target bands are primer dimers. See text for interpretation of band brightness.

Lesson 7: Sanger Sequencing

Rationale: Students will be introduced to a technology that revolutionized the genomics field and is currently used all around the world. They will learn how Sanger Sequencing (Sanger 1977) combines the knowledge of biological systems, PCR technology, and gel electrophoresis. Also, they will gain experience in how to book keep and submit data to a company according to certain specifications.

Methods: We used a commercial company, GeneWiz (Piscataway NJ), to sequence our DNA, and prepared our samples according to their specifications, which are available on their website (<http://www.genewiz.com>). Students dilute the PCR product according to the measured concentrations from the gel in Lesson 6. They add 12 ml of water to each sample that was measured as "1", 20 ml to each "2", and 40 ml to each "3." This provides template DNA in the ballpark concentration needed for sequencing (roughly 50 pg/ml). Students create a "map" of a 96-well plate to record where each PCR product will be placed. They then load 10 ml of the diluted samples into the appropriate wells, according to their map. A parallel plate of the appropriate primers is also loaded, according to the GeneWiz specifications (5ml of 25 pmol/ml in each well). Plates were delivered to GeneWiz via a dropbox.

Learning Objectives:

1. Know that Sanger sequencing is a method that synthesizes oligonucleotides, terminating the reaction at each nucleotide.
2. Know that the method uses similar concepts to those learned from PCR and gel electrophoresis.
3. Communicate the similarities and differences between PCR and Sanger sequencing.
4. Understand how gel electrophoresis is similar to movement through micro-capillaries, used in this process.
5. Discuss how slight modifications of PCR methods lead to a different result (sequences). Once structures and processes are understood, modifying protocols is typical of how methods are developed in molecular biology.
6. Be able to discuss the process of Sanger Sequencing (Sanger 1977). In the same manner as PCR, DNA is denatured and a complementary strand is built from the primers. Although, in this case, the strand is built from only one of the two primers at a time. The reactions are duplicated, with one set of reactions targeted at the leading and the other, the lagging strand. In the solution, there is a mixture of regular deoxyribonucleic nucleotide triphosphates (dNTPs) and dideoxynucleotides nucleotide triphosphates (ddNTPs). The ddNTPs are missing the active 3' hydroxyl group. Review (Fig. 8), and ask students what would happen if the OH group was missing (the reaction terminates). In addition, each ddNTP is tagged with one of 4 fluorescent dyes, according to whether it is a ddATP, ddCTP, ddGTP, or ddTTP. They function as regular nucleotide building blocks, but when they bind to the DNA target strand, the missing hydroxyl group causes the termination of the reaction. This happens at every nucleotide position in the sequence because the majority of nucleotides are regular dNTPs. Thus, termination of the reaction happens randomly, whenever a ddNTP hybridizes to the target DNA. Because charge is directly proportional to size in DNA, the resulting size difference among each DNA fragment allows them to be separated on a gel, or through a fine capillary. As each fragment passes by a sensor, the colored dye is recorded. Because every fragment of the same size terminates with the same ddNTP, and each of the ddNTPs is dyed with a known color, each fragment of each size is dyed with only one color. The fragments are separated in an electric field, as if in a race, where the smaller fragments win. The color is recorded as each fragment passes by a sensor. The smaller the fragment, the faster it moved through the gel or capillary, just as in the agarose gel.

Lesson 8: Editing DNA sequences

Rationale: Sequence results are presented as a series of peaks in a format that can be read by the free ApE sequence editor (<http://biologylabs.utah.edu/jorgensen/wayned/ape/>). Students edit the sequences to gain an understanding of the errors that can occur.

Methods: Both the leading and lagging strand of the same sequence were uploaded into the ApE software. The DNA sequence results are shown as consecutively colored peaks indicating the dye concentrations of the bands that passed by the sensor. The results can vary in quality. A "good" part of the sequence is characterized by even, non-overlapping peaks and low background. Next, reverse complement the lagging strand and align the two sequences. To ensure accuracy, two strands of complementary DNA are checked against each other for sequence confirmation (Fig. 11). If there is a conflict between the reads, check each read independently. If the nucleotide reading is dubious, the nucleotide is given a lowercase letter. If there are multiple peaks overlapping, then a DNA ambiguity letter is given based on which nucleotides are overlapping (e.g., Y codes

for a pyrimidine and R for a purine). A consensus sequence is constructed from the two strands. Depending on the read quality of both the leading and lagging strand, the entire sequence might be discarded or the sequence might be determined by only one strand which should be noted (Lyons 2015). Editing determines where to eliminate the unreliable sequence at both the beginning and the end of the graphs.

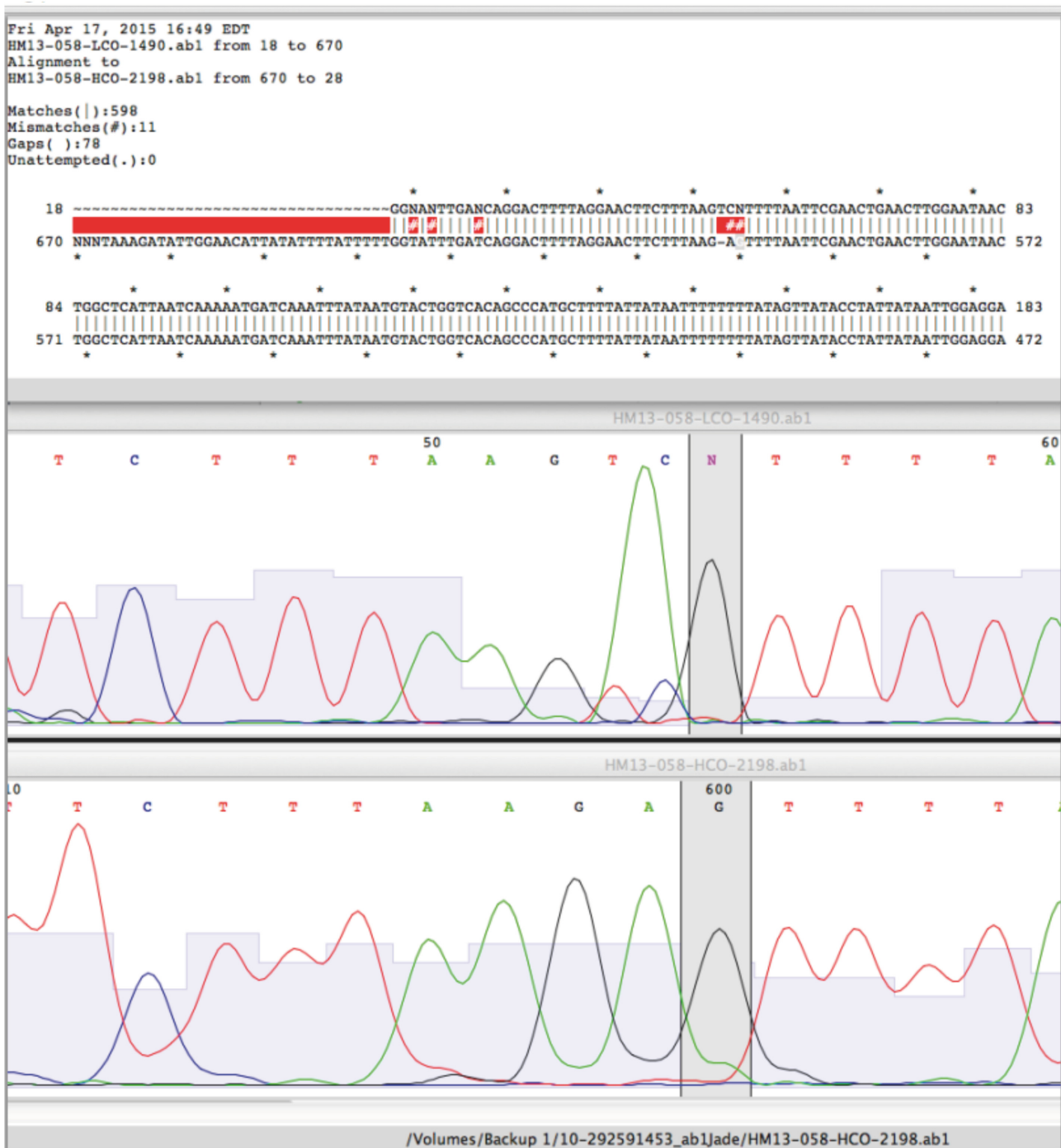


FIGURE 11. Proofreading results by reverse complementing the DNA strands. Upper panel: DNA from each strand, aligned in APE software. Missing strands and disagreements between strands are highlighted. Lower panel: Aligned data from each strand, highlighting a single nucleotide disagreement. In this case, the "N" in the upper strand appears to be a "G", which is confirmed in the lower strand.

Learning Objectives:

1. Understand why high-quality results are usually found in the middle of the read (Fig. 12). At the beginning, close to the primer, there are usually blobs of competing dye. At the end of the read, the resolution begins to fade because the peaks begin to become difficult to separate.
2. Recognize that the quality of the read depends on many factors.
3. Be informed that you may be able to get away with using sloppily edited sequences, but your long-term reputation for producing good sequences is more important. Be meticulous.

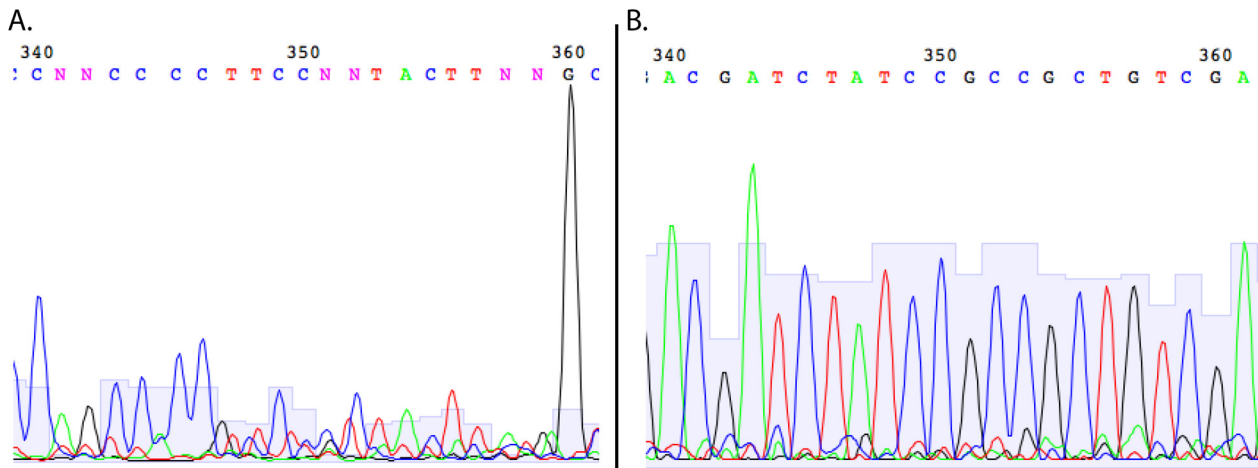


FIGURE 12. A graphic display of sequencing results. (A) Low quality data. (B) High quality data.

Lesson 9: Submitting the data to the database for identification

Rationale: Students will be exposed to large datasets and learn how to use the data in their own research studies. They will see their effort as a part of the larger scientific community and realize that advancements in research require the work of many people around the world.

Methods: The consensus sequence is submitted to the Barcode of Life Data Systems identification site (http://www.boldsystems.org/index.php/IDS_OpenIdEngine) to see if the sequence matches other sequences or if the sequence is new to the database. The identity of the sequence can be estimated by two available options in BOLD: either by how similar the sequence is in terms of pairwise percent identity (Fig. 13) to its closest match in the database or in the context of a phylogenetic tree (Fig. 14).

Learning Objectives:

1. Know that in a phylogenetic tree, the branch lengths of a phylogram corresponds to the amount of nucleotide differences between groups. These differences reflect a combination of time, and substitution rates, which vary among taxa.
2. Know that the longer the branch length, the more differences between sequences and thus, the more likely they are to belong to different species. The converse is true of shorter branch lengths.
3. Identify a species using a phylogenetic tree based on the unknown specimen being nested inside a group of identified specimens (Zhou *et al.* 2007)
4. Identify species based on percent similarity in regards to the nucleotide base sequence.
5. Recognize that in a large dataset, 2% is often used as the standard for species identification. However, because substitution rates can vary among lineages, it is important to place pairwise nucleotide difference into some meaningful species concept.
6. Know that any sequence difference other than 100% identity is an arbitrary number and may not accurately represent species (Zhou *et al.* 2007).

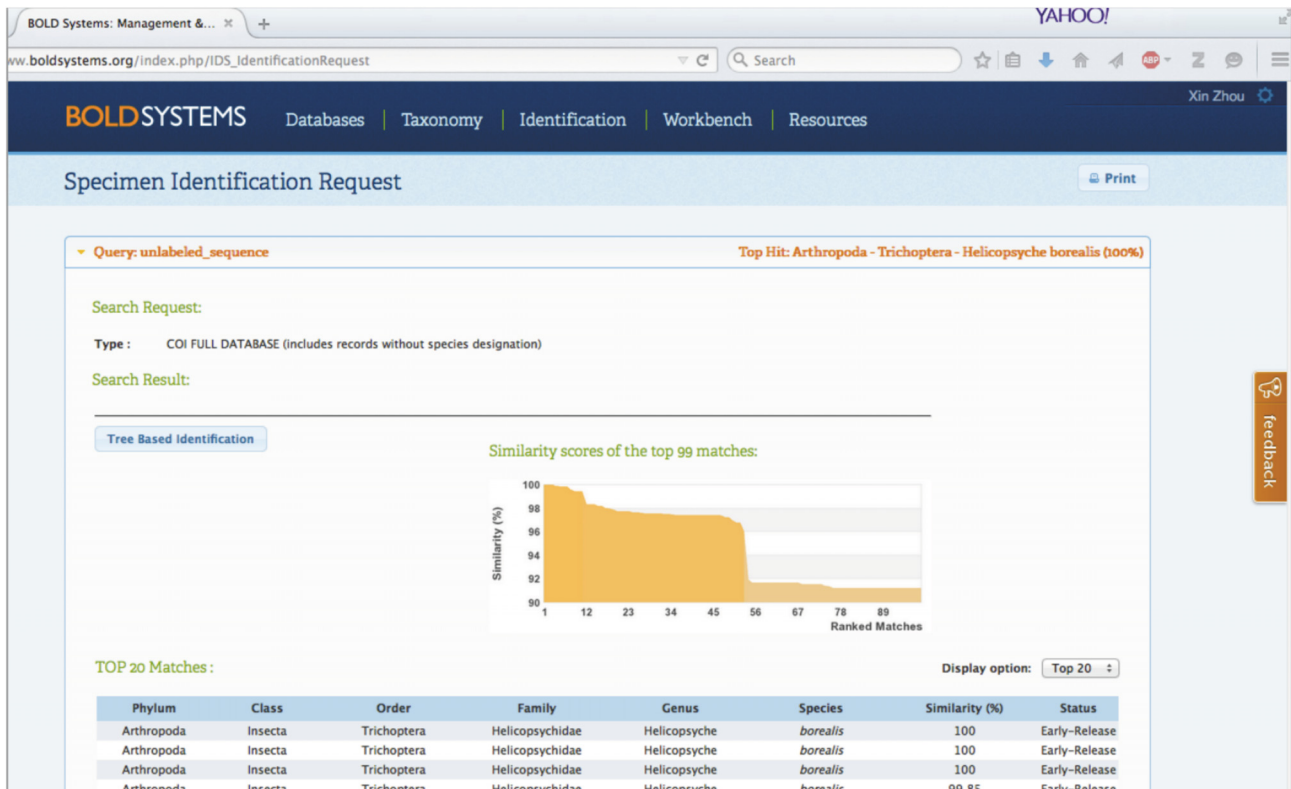


FIGURE 13. Percentage based identification. In this example, the unknown is identical to *Helicopsyche borealis* from New York State.

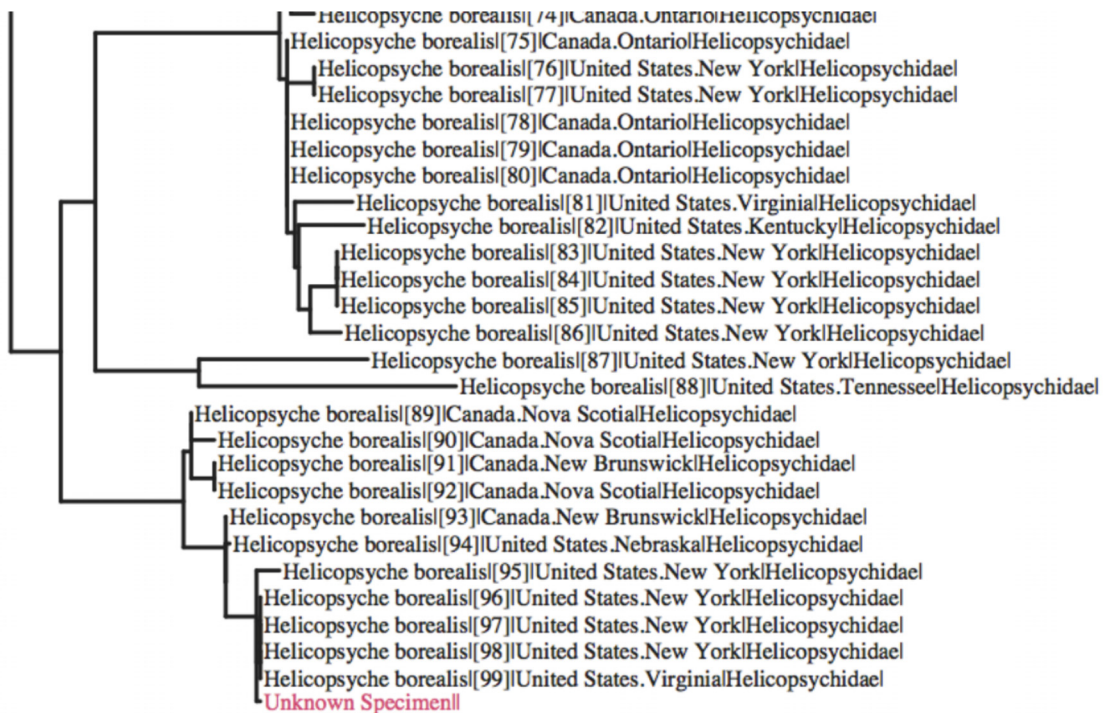


FIGURE 14. Tree based identification. In this example, the unknown is nested inside *Helicopsyche borealis*, grouping with a cluster from New York State.

Lesson 10: Phylogenetic analysis

Rationale: An in depth understanding of phylogenetic analysis is beyond the scope of a 10-lesson course. The instructor may use this lesson to explore a wide variety of phylogenetic methods and introduce students to different freely available software packages. It is good to show students that with the data they have generated, they can provide new information to building the tree of life. They are working with real data that has never been subject to a phylogenetic analysis. Therefore, the trees they create in this short lesson have never been seen before by anyone.

Methods: Students should be able to download data from BOLD and manipulate it for phylogenetics. Students could download the SeaView software (Gouy *et al.* 2010) to align data and export the alignment in a variety of formats, such as FASTA and NEXUS. Once in these formats, students would be encouraged to download a program such as PAUP* (Swofford 2003), with its user-friendly pull-down menus, to explore different methods of analysis and branch support.

Learning Objectives:

1. Use a variety of software to manipulate and analyze data to form phylogenetic trees.
2. Be able to identify errors in the data

Discussion

The scientific field is composed of many different areas of study and barcoding Trichoptera helps to bring these specialties together in a logical and systematic process. Students will be able to see and appreciate the larger picture that science is interconnected. One thing comes from, is affected by, and affects another. The process teaches that not only is knowledge necessary but it is critical for students to be able to apply the information in a systematic fashion. Future scientists need to think for themselves, to look at a problem and work towards a solution. DNA barcoding exposes students to advanced scientific equipment and gives them a chance to experience something from the beginning to the end. E.g., field collecting, molecular techniques, phylogenetic, molecular interactions, biodiversity, species identification, and big databases are invaluable genetic skills and concepts that can be applied to fields beyond Trichoptera research. These lessons were initially roughly applied in a 10-day DNA workshop held in conjunction with the XVth International Symposium on Trichoptera. Roughly 20 Scientists of varying academic level participated in the workshop including, but not limited to graduate students, PhD candidates, PhDs, and undergraduate research lab assistants who had previous experience in a DNA laboratory. Each day was a different lesson along with the associated lab task. Because of the short amount of time, participants were only able to properly learn and apply a handful of laboratory techniques.

For subsequent courses, the amount of time to learn and practice the material in each section would be extended to a few months. In addition, the class size should remain small or the number of assistants should increase to give students more personalized attention and make sure they fully understand the material. Other scientific curricula use this mixture of lecture and complementary lab to teach their students the foundational information and their applicable techniques. Although, the lecture topics and labs are usually topical, and the experiments often last a week or two. The presented curriculum combines both field work, lab, and lecture and provides a sequential order that will last the entire semester. The integration creates a scientist who understands and knows how many pieces fit into the larger picture to better solve problems. Of course, there is much more to learn and experience, and many other fields that DNA sequencing does not cover, but this approach gives students a comprehensive introduction to making connections between topics and people, and to see how science is applied and moved forward.

Acknowledgements

We thank the Rutgers Byrne Seminar program, the Douglass Project, the Aresty Research Center for Undergraduates, and the STEM summer stipend program for financial support. We thank Nicolas Firbas,

Marzia Lisa and Julianne McLaughlin for assistance in the lab, and their advice. We thank Wodzinski and Laura Stiltz for their academic support. Finally, we thank Christine Frandsen for assistance with the figures.

References

- Hebert, P.D., Cywinska, A., Ball, S.L. & deWaard, J.R. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences* 270, 313–321.
<https://doi.org/10.1098/rspb.2002.2218>
- Holzenthal, R.W., Thomson, R.E. & Ríos-Touma, B. (2015) Order Trichoptera. In: Thorp, H.J. & Rogers, C.D. (Eds.), *Thorp and Covich's Freshwater Invertebrates: Ecology and General Biology (4th Edition)*. Elsevier Inc., San Diego CA USA, 994–996 pp.
<https://doi.org/10.1016/B978-0-12-385026-3.00038-3>
- Gouy, M., Guindon, S. & Gascuel, O. (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, 27, 221–224.
<https://doi.org/10.1093/molbev/msp259>
- Lyons, R. (2015) The University of Michigan DNA Sequencing Core, The University of Michigan. Available from: <https://seqcore.bref.med.umich.edu/> (22 August 2015)
- Sanger, F., Nicklen, S. & Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74, 5463–5467.
<https://doi.org/10.1073/pnas.74.12.5463>
- Swofford, D.L. (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
- Truett, G.E., Heeger, P., Mynatt, R.L., Truett, A.A., Walker, J.A. & Warman, M.L. (2000) Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and Tris (HotSHOT). *Biotechniques*, 29, 52–54.
<https://doi.org/10.2144/00291bm09>
- Wikimedia Commons (2015) Wikimedia Foundation Inc. Available from: <https://commons.wikimedia.org/> (18 September 2015)
- Wikipedia: The Free Encyclopedia (2016) Wikimedia Foundation Inc. Available from: <https://en.wikipedia.org/> (6 January 2016).
- Zhou, X., Kjer, K.M. & Morse, J.C. (2007) Associating larvae and adults of Chinese Hydropsychidae caddisflies (Insecta:Trichoptera) using DNA sequences. *Journal of the North American Benthological Society*, 26, 719–742.
<https://doi.org/10.1899/06-089.1>