BIOCHEMISTRY AND MOLECULAR BIOLOGY EDUCATION
Vol. 36, No. 5, pp. 354–362, 2008

Laboratory Exercises

Fly Diversity Revealed by PCR-RFLP of Mitochondrial DNA

Received for publication, May 13, 2008, and in revised form, May 24, 2008

Jimmy F. Asraoui, Nancy P. Sayar, Khouzama M. Knio, and Colin A. Smith‡
From the Department of Biology, American University of Beirut, Beirut, Lebanon

In this article, we describe an inexpensive, two-session undergraduate laboratory activity that introduces important molecular biology methods in the context of biodiversity. In the first session, students bring tentatively identified flies (order Diptera, true flies) to the laboratory, extract DNA, and amplify a region of the mitochondrial gene NADH dehydrogenase subunit 1. In the second session, the students digest the PCR product with a restriction enzyme, visualize the resulting fragments by agarose gel electrophoresis, and analyze their results with comparison to known sequences. The diversity of flies and their importance as disease vectors, agriculture pests, pollinators, models of speciation, and in the case of *Drosophila melanogaster*, as a genetic model organism, offer many perspectives with which to appeal to students' interests. The laboratory exercise can be linked as a module to topics in biodiversity, bioinformatics, entomology, evolution, and mutagenesis.

Keywords: Laboratory exercises, molecular biology, nucleic acid enzymology, evolution.

INTRODUCTION

The power and widespread use of molecular biology techniques mandates their inclusion in the undergraduate laboratory. We had introduced PCR and restriction enzyme digestion into our undergraduate genetics laboratory, but informal observation indicated that our students found the plasmid-based activities arcane. Our goal was to create an inexpensive, robust laboratory activity that would introduce important molecular biology methods, interest students, and could be coupled to original scientific questions. We wanted an activity that would not merely demonstrate the procedures, but encourage students to understand their possible applications and limits. An ancillary goal was to expose students to nonnumerical and ambiguous data. We hoped practical experience in PCR and restriction site analysis in a rich biological context would help students to master in DNA analysis. Additionally, we wanted to link the activity to other courses in bioinformatics, entomology, evolution, and biodiversity.

We exploited a mitochondrial PCR-RFLP method developed for our research on distinguishing tephritid flies that infest local thistles [1]. We expected the method to function with most or all species within the order Diptera (true flies). The envisioned activity would pack DNA extraction, PCR, restriction enzyme digestion, agarose gel electrophoresis, and analysis into two sessions. Students would be responsible for collecting flies, tentative identification, and predicting digestion products of a sequenced relative using on-line resources. A similar activity has been described for commercial mushrooms [2], and we reasoned that flies could appeal to students' interests as disease vectors, agricultural pests, and familiarity as typical house or laboratory organisms. The simple question of fly identity can be expanded into discussion and projects related to biodiversity, ecology, and phylogeny, while the practical work can be expanded into bioinformatics, classification, and DNA sequencing. Importantly, controversies exist in fly phylogeny, and most fly mitochondria have not been sequenced [3]. The activity can also be used to illustrate typical analytic problems in genetics, such as identifying reading frames, the consequences of mutagenesis, sequencing, as well as general PCR and restriction enzyme problems.

The diverse order Diptera (phylum: Arthropoda; class: Insecta) has more than 150,000 species in more than 100 families originating from a most recent common ancestor 250 million years ago [4]. The order includes disease vectors such as mosquitoes, sandflies, and blow flies, agriculture pests including the medfly, the domestic housefly, and the model organism *Drosophila melanogaster* (Fig. 1). The detailed evolutionary relationships among Diptera are uncertain [4]. Identifying precise species often requires an expert, but identifying the family or subfamily can be done by a novice with access to guide books. Genetic evidence can provide corroborating evidence of identity and is needed for phylogenetic analysis, yet few species are completely sequenced and many species have no recorded sequences. Mitochondrial DNA genes are particularly attractive for distinguishing

---

‡To whom correspondence should be addressed. Department of Biology, American University of Beirut, Beirut, Lebanon. Tel.: +961-1-374-374 ext 3887; Fax: +961-1-374-374 ext 3888. E-mail: cs10@aub.edu.lb.

DOI 10.1002/bmb.20220

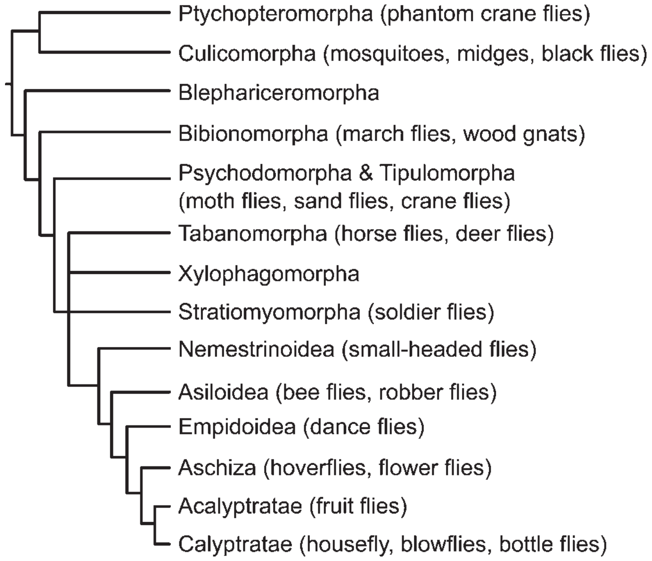


FIG. 1. **Phylogeny of Diptera**. A proposed phylogeny within the order Diptera adapted from Yeates and Wiegmann [4], using infraorders and other high-level categories. Flies with common names are indicated in their category. This is not an authoritative phylogeny; many controversies exist as to the precise relationships with Diptera.

species, because they exist in all animals, mutate relatively rapidly, and are present in multiple copies. Mitochondria are inherited almost exclusively via the oocyte in flies, and hence reveal maternal lineages. Dipteran mitochondria have circular, double-stranded DNAs, about 16 kb, and encode some proteins used for energy production, as well as some tRNAs and rRNAs [5]. Interestingly, the invertebrate mitochondrial genetic code is not identical to the standard nuclear code. Our procedure amplifies a typically 690 bp region of the mitochondrial gene NADH dehydrogenase subunit 1 (ND1),[1] which codes for 230 amino acids of a protein in the electron transport chain (Fig. 2). Observed mutations are expected to be neutral with respect to fitness, because ND1 is an essential gene. Since single nucleotide polymorphisms (SNPs) are predicted to occur at a frequency of 1–3 $\times$ 10^{-8} per bp per year [6], even recently diverged populations are likely to have distinct sequences.

The polymerase chain reaction is arguably the most widely applied technique of molecular biology. In each cycle of PCR, DNA is denatured, primers are annealed to template, and primers are extended, such that the product of one primer extension serves as a template for the other primer in the next cycle. Multiple cycling causes a chain reaction and a near doubling of the product in each cycle. Sufficient portions of the target sequence must be known for primers to be chosen. In this case, we used the ND1 reading frame from available completely sequenced dipteran mitochondria. Initially, the template DNA is denatured at high temperature and cooled in the presence of excess primers to a cooler temperature, in which the primers anneal to the template. The primed template is extended with DNA polymerase and deoxyribonucleotide triphosphates (dNTPs). The availability of synthetic DNA primers, thermostable polymerases, dNTPs, and thermal cyclers has made the procedure widely accessible.

Restriction enzymes recognize and cleave specific sequences of DNA. They have been isolated from bacteria, in which they function to digest foreign DNA. The most common laboratory restriction enzymes recognize short, palindromic sequences of 4–8 bps. Their ability to cleave DNA at specific sequences makes them a mainstay of recombinant technology. DNA sequence differences including SNPs cause sequences to be cleaved into different fragment lengths, which are detectable by electrophoresis. This method, called restriction fragment length polymorphism analysis (RFLP), has found wide application in gene mapping and forensics and is much less expensive than DNA sequencing. Although it is not possible to prove the identity of two DNAs by RFLP, it is usually possible to prove that a sequence is different from another.

The procedure uses two primers to amplify a typically 690 bp region from the open reading frame of the mitochondrial gene encoding for ND1 (Fig. 2). The primer sites are highly conserved in Diptera and may allow for amplification of ND1 from any fly with one set of primers. Our method uses Tru9I (an isoschizomer of MseI), which cleaves at 5′-TTAA-3′ to generate about 10 fragments in the very AT-rich amplified fragment of dipteran ND1 (Fig. 2). We chose to resolve digests on agarose gels, because they can be prestained with ethidium bromide and viewed throughout electrophoresis. The rate of DNA divergence in *Drosophila* (Acalyptratae; Fig. 1) has been estimated to be ~1.7% per million years [7]. We roughly estimate that a single nucleotide change has a 17% chance either to ablate an existing Tru9I site or create a new Tru9I site, suggesting that 500,000 years divergence would be sufficient to create an RFLP. Inspection of the ND1 region in four completely sequenced *Drosophila* mitochondria (*D. melanogaster*, *D. simulans*, *D. mauritiana*, and *D. yakuba*) predicts that our method would identify three distinct patterns but fail to distinguish the two close relatives, *D. mauritiana* and *D. simulans*, which are estimated to have speciated 1 million years ago [8]. Thus, our PCR-RFLP method is likely to distinguish between many species even within the same genus.

We describe here an inexpensive PCR-RFLP method applied to a variety of flies and its implementation in a second-year undergraduate genetics laboratory course. The laboratory is robust, flexible, and requires little prior experience of students. The activity can be easily expanded to include more activities and linked to other topics in biology. The students' experience and critical parameters revealed by implementation are described.

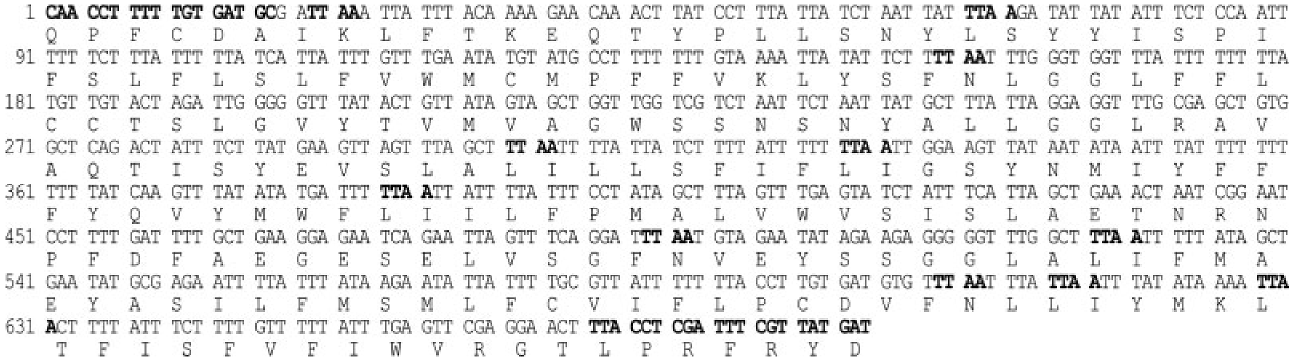
MATERIALS AND METHODS

Preparation

We modified some aspects of our established methods for student use. We switched from glass pestles to flame-sealed

---

[1] The abbreviations used are: ND1, NADH dehydrogenase subunit 1; dNTP, deoxyribonucleotide triphosphates; SNP, single nucleotide polymorphism; RFLP, restriction fragment length polymorphism.



```
  1 CAA CCT TTT TGT GAT GCG ATT AAA TTA TTT ACA AAA GAA CAA ACT TAT CCT TTA TTA TCT AAT TAT TTA AGA TAT TAT ATT TCT CCA ATT
     Q   P   F   C   D   A   I   K   L   F   T   K   E   Q   T   Y   P   L   L   S   N   Y   L   S   Y   Y   I   S   P   I
 91 TTT TCT TTA TTT TTA TCA TTA TTT GTT TGA ATA TGT ATG CCT TTT TTT GTA AAA TTA TAT TCT TTT AAT TTG GGT GGT TTA TTT TTT TTA
     F   S   L   F   L   S   L   F   V   W   M   C   M   P   F   F   V   K   L   Y   S   F   N   L   G   G   L   F   F   L
181 TGT TGT ACT AGA TTG GGG GTT TAT ACT GTT ATA GTA GCT GGT TGG TCG TCT AAT TCT AAT TAT GCT TTA TTA GGA GGT TTG CGA GCT GTG
     C   C   T   S   L   G   V   Y   T   V   M   V   A   G   W   S   S   N   S   N   Y   A   L   L   G   G   L   R   A   V
271 GCT CAG ACT ATT TCT TAT GAA GTT AGT TTA GCT TT AATT TTA TTA TCT TTT ATT TTT TTA ATT GGA AGT TAT AAT ATA ATT TAT TTT TTT
     A   Q   T   I   S   Y   E   V   S   L   A   L   I   L   L   S   F   I   F   L   I   G   S   Y   N   M   I   Y   F   F
361 TTT TAT CAA GTT TAT ATA TGA TTT TTA ATT ATT TTA TTT CCT ATA GCT TTA GTT TGA GTA TCT ATT TCA TTA GCT GAA ACT AAT CGG AAT
     F   Y   Q   V   Y   M   W   F   L   I   I   L   F   P   M   A   L   V   W   V   S   I   S   L   A   E   T   N   R   N
451 CCT TTT GAT TTT GCT GAA GGA GAA TCA GAA TTA GTT TCA GGA TTT AAT GTA GAA TAT AGA AGA GGG GGT TTG GCT TTA ATT TTT ATA GCT
     P   F   D   F   A   E   G   E   S   E   L   V   S   G   F   N   V   E   Y   S   S   G   G   L   A   L   I   F   M   A
541 GAA TAT GCG AGA ATT TTA TTT ATA AGA ATA TTA TTT TGC GTT ATT TTT TTA CCT TGT GAT GTG TTT AAT TTA TTA ATT TAT ATA AAA TTA
     E   Y   A   S   I   L   F   M   S   M   L   F   C   V   I   F   L   P   C   D   V   F   N   L   L   I   Y   M   K   L
631 ACT TTT ATT TCT TTT GTT TTT ATT TGA GTT CGA GGA ACT TTA CCT CGA TTT CGT TAT GAT
     T   F   I   S   F   V   F   I   W   V   R   G   T   L   P   R   F   R   Y   D
```

FIG. 2. **Mitochondrial ND1 amplicon and translation**. The portion of the mitochondrial ND1 sequence amplified is shown for *Drosophila melanogaster* (corresponding to RefSeq NC_001709 12508-11819). The sequence is parsed into codons with translation below. Primer sites and Tru9I recognition sites are indicated in bold. Note that invertebrate mitochondrial codon usage is nonstandard.

blue tips (1,000 μL) to avoid injuries and added more ethidium bromide to agarose to aid students' observation. Master mixes were prepared daily (Tables I and II). We prepared reference material on the course management site (Moodle, Perth, Australia) with a phylogeny to the family level and links to known fly ND1 sequences (21 species, 9 families, 4 of 11 infraorders), analysis guidance, an invertebrate mitochondrial codon chart, and links to several fly taxonomy sites: http://bugguide.net; www.sel.barc.usda.gov/Diptera/; www.cedarcreek.umn.edu/insects/albumframes/dipteraframe.html; http://www.diptera.info; www.cals.ncsu.edu/course/ent425/compendium/diptera.html. We used pDRAW32 (available free from www.acaclone.com) to analyze sequences and to prepare teaching figures. One week before the session, students were told that they would need to catch and tentatively identify a fly, and then download a sequence of a close relative. If accurate identifications are desired, assistance of an entomologist is recommended. The Genome database (http://www.ncbi.nlm.nih.gov/genomes/genlist.cgi?taxid=2759&type=4&name=Eukaryotae%20Organelles) has completely sequenced mitochondria of some species in the following dipteran genera: *Aedes*, *Anopheles*, *Bactrocera*, *Ceratitis*, *Chrysomya*, *Cochliomyia*, *Culicoides*, *Cydistomyia*, *Dermatobia*, *Drosophila*, *Haematobia*, *Lucilia*, *Simosyrphus*, and *Trichophthalma*. In the laboratory manual and during two weekly 1-hour laboratory lectures, learning objectives, scientific context, explanations of procedures, and sample gel images and their interpretation were presented. For students arriving without flies, we provided live laboratory *D. melanogaster* or locally collected flies (stored at −20 °C). Only the head was used, if the fly was estimated to exceed 100 μL volume.

We divided the activity into two weekly sessions of 3 hours each. It would be possible to complete the activity in one 8-hour session, with a 2-hour break during thermocycling, or could be expanded to more sessions if fly identification, full analysis, and students' presentations are included.

Briefly, the activity uses commonly available equipment for microcentrifugation, micropipetting, thermocycling, agarose gel electrophoresis, and a UV light source. Flies (wings can be removed for vouchers) are ground with a pestle in a 1.5-mL microcentrifuge tube with a grinding solution, a lysis solution containing SDS is added, the mixture is centrifuged to pellet debris (wings, carapace, etc), and then the supernatant is transferred to a new tube. Sodium chloride is added to precipitate protein, the samples are centrifuged, and then the resulting supernatant is transferred to a new tube. The addition of ethanol precipitates DNA and RNA, usually as a clearly visible pellet. The pellet can be washed with ethanol, dried, and resuspended in water or buffer, and 1 μL is used as template for 25–50 μL PCR reactions.

Materials

PCR primers were obtained from Thermo Electron (Ulm, Germany). *Thermus aquaticus* DNA polymerase I (Taq) was made following a published protocol [9], and commercial Taq

TABLE I

PCR master mix preparation for 25 μL reactions

Final concentration[a]	Reagent or step[b]	Stock concentration	Volume used (μL)	Mix for 20 reactions (μL)
1×	10× Taq Rxn buffer	10×	2.5	50
500 μM F, R[c]	Mix of both primers	5 μM[c]	2.5	50
200 μM[c]	dNTP mix	2 mM[c]	2.5	50
1.5 mM	$MgCl_2$	15 mM	2.5	50
Solvent	Water	Pure	13	260
	Mix gently	–	–	–
1 unit per reaction	Taq enzyme	1 unit per μL	1	20
	Mix gently[d]	–	–	–
	Add 24 μL to each reaction tube	–	–	–
[1 μL per reaction][e]	[Template in reaction tube][e]		[1][e]	[20][e]
	Total volume	–	25	500

[a] Concentration in assembled reaction.

[b] These should be added and performed in the order indicated with care to avoid foaming.

[c] Concentration of each component in the mix.

[d] The PCR master mix is completed by this step and can be stored for several hours at 4 °C before dispensing to reaction tubes.

[e] This should be previously placed in reaction tube, mixing occurs by addition of mix. Brackets indicate it is not part of the master mix; it is in the assembled reaction.

Table II

RFLP master mix preparation for 20 μL reactions

Final concentration[a]	Reagent or step[b]	Stock concentration	Volume used (μL)	Mix for 20 reactions (μL)
1×	10× Rxn buffer	10×	2	40
Solvent	Water	Pure	7.5	150
	Mix gently	–	–	–
5 units per reaction	Tru9I enzyme	10,000 units per mL	0.5	10
	Mix gently[c]	–	–	–
	Add 10 μL to each digest tube	–	–	–
	[PCR product in digest tube][d]	As is	[10][d]	[200][d]
	Mix gently	–	–	–
	Total volume	–	20	400

[a] Concentration in assembled reaction.

[b] These should be added and performed in the order indicated.

[c] The RFLP master mix is completed by this step and can be stored for several hours at 4 °C before dispensing to digest tubes.

[d] The PCR product to be digested can be placed in reaction tube earlier. Brackets indicate it is not part of the master mix; it is in the assembled reaction.

(ABgene, Epson, UK) functioned equally well. Several DNA ladders were used, such as Superladder-Mid-2 200 bp Ladder (ABgene); MassRuler DNA Ladder, Low Range (Fermentas, Vilnius, Lithuania); a homemade ladder of PCR products and synthetic oligonucleotides; GeneRuler, DNA ladder, Low Range (Fermentas). dNTPs were obtained from ABgene, Tru9I from Roche (Mannheim, Germany), and agarose, buffers, and chemicals were obtained from Amresco (Solon, OH).

Fly Collection and Identification

Flies were collected locally by instructors and students. Tentative identifications were made by students, whereas precise identifications were made by an expert (KMK) with reference to published guides. *D. melanogaster* laboratory strain was taken from a stock originally purchased from Carolina Biological Supply (Burlington, NC).

DNA Extraction

Flies were collected and stored at −20 °C. Each fly, or only the head of large flies, was placed in a 1.5-mL microfuge tube and crushed using a flame-sealed blue tip as a pestle, followed by the addition of 200 μL grinding buffer (10 mM Tris-HCl pH 7.8, 60 mM NaCl, 300 mM sucrose, 10 mM EDTA pH 8), and 30 s of grinding. The tubes were placed on ice and 200 μL lysis buffer (300 mM Tris-HCl pH 7.8, 1% SDS, 20 mM EDTA pH 8) was added, and then the tubes were left on ice for 10 to 30 minutes. Each sample was centrifuged for 5 minutes at 15,000 × *g*, 300 μL of supernatant was transferred to a fresh 1.5-mL tube, and 400 μL of 4 M NaCl was added with mixing. The samples were centrifuged 5 minutes at 15,000 × *g*, 650 μL of supernatant was transferred to a fresh 1.5-mL tube, and 1 mL of absolute ethanol was added. After 10 to 30 minutes on ice, the tubes were centrifuged for 15 minutes at 5,000 × *g*, the supernatant was discarded, 500 μL 75% ethanol was added, and then the sample was centrifuged again for 5 minutes at 15,000 × *g*. After a final wash with 100 μL of 75% ethanol, the pellet was allowed to air dry before resuspension in 200 μL TE (10 mM Tris-HCl pH 7.8, 1 mM EDTA pH 8). The DNA extracts were stored at −20 °C.

Polymerase Chain Reaction Amplification

A fragment of the mitochondrial ND1 gene (corresponding to RefSeq NC_001709 12508-11819) was amplified using the primers of Smith *et al*. [10], which was altered to include a degenerate position in the reverse primer: ND1-FD: 5′-ATCATAAC GAAAYCGAGGTAA-3′; ND1-RD: 5′-CAA CCT TTT WGT GAT GC-3′; where Y = C + T and W = A + T. The degeneracies account for two common known polymorphisms in fruit flies. PCR reactions of 25 μL used 1 μL of DNA extract and the following final reagent concentrations: 500 nM forward primer, 500 nM reverse primer, 200 μM each dNTP, 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl pH 8.3 at 25 °C, and 1 U Taq (Table I). The amplifications were performed on a Bio-Rad DNA Engine (Hercules, CA) using the following program: one cycle at 94 °C for 3 minutes, 35 cycles at 93 °C for 30 seconds; 50 °C for 30 seconds, 72 °C for 45 seconds, and one cycle at 72 °C for 10 minutes. For sequencing, ~50 pmole of gel-purified PCR product (estimated by comparison to standards on stained agarose gels) was mixed with 5 pmole of primer ND1-FD and sequenced with BigDye terminator cycle sequencing kit (Applied Biosystems, Foster City, CA) and analyzed on a ABI3100-Avant Genetic Analyser. Sequence files were analyzed using ChromasLite (available free from www.technelysium.com.au/chromas_lite.html; Technelysium, Tewantin, Australia).

RFLP Analysis

Unpurified PCR products were digested with Tru9I (Roche) in 20 μL reactions containing 5 μL PCR product, 2 μL 10× M buffer (Roche), 12.5 μL water, and 0.5 μL Tru9I (10 U/μL) at 65 °C for 60 minutes (Table II).

Electrophoresis

Samples were mixed with a loading solution containing glycerol (5% final) and bromophenol blue (0.02%). As little as 2 μL of PCR reaction and as much as 20 μL of completed digests were loaded on either 1% agarose I (Amresco) or 4% agarose (3:1 High Resolution Blend, ABgene) with 1–2 μg/mL ethidium bromide and electrophoresed using 1XTBE (0.9 M Tris-borate, 2 mM EDTA) at 5–10 volts/cm for 1 hour on a homemade gel box or using a minisub cell GT (Bio-Rad). Gels were viewed on a transilluminator using 254 nm. Photographs were taken with a Canon PowerShot A75 (Tokyo, Japan) through a Yashica O2 orange filter (Hong Kong). The photographs were enhanced for brightness and contrast in Adobe Photoshop (Adobe Systems, San Jose, CA) to better reproduce which was visible by eye.

Laboratory Safety

Ethidium bromide, a known mutagen and likely carcinogen, must be disposed in accordance with institutional guidelines. Students should wear examination gloves and eye protection. Plastic shields and goggles can protect against the UV light produced by transilluminators, which can cause sunburn, skin cancer, and cataracts. A UV–visible absorbance spectrophotometer is recommended to confirm shield function. Electropho-

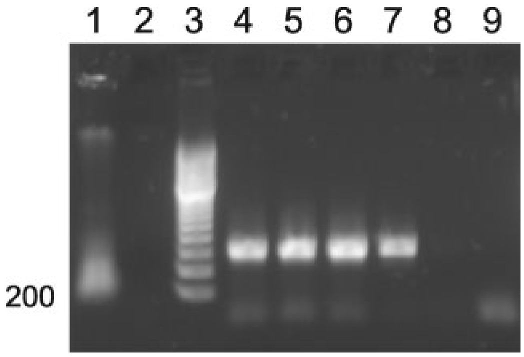


FIG. 3. **DNA extracts and PCR products**. Shown is a 1% agarose gel stained with ethidium bromide. *Lane* 1, 5% of a *Terellia fuscicornis* extract; *lane* 2, 5% of a *D. melanogaster* extract; *lane* 3, 200 bp DNA ladder (Superladder-Mid-2 200 bp Ladder); *lane* 4, 5 μL of PCR with 40 °C annealing; *lane* 5, 5 μL of PCR with 50 °C annealing; *lane* 6, 5 μL of PCR with 54 °C annealing; *lane* 7, 5 μL of PCR with 60 °C annealing; *lane* 8, 5 μL of PCR with 62 °C; *lane* 9, 5 μL of no template PCR with 50 °C annealing.

resis uses voltage capable of causing dangerous electric shock, and students need to be warned and monitored. Molten agarose, especially the 4% agarose used here, can cause serious burns. Students should exercise caution when collecting flies, as some flies are disease vectors, some bite, and some related insects are venomous.

RESULTS

Method Development

We had introduced a plasmid-based restriction enzyme activity in our undergraduate genetics laboratory course, but we were disappointed by the lack of students' enthusiasm. Informal inquiries indicated students appreciated the gel electrophoresis, but found plasmids and restriction enzymes disconnected to their interests. We decided to replace the plasmid-based activity with a PCR-RFLP method we had recently developed for distinguishing cryptic species of thistle-infesting tephritid flies [1]. Because we have ∼90 students divided in 6 or 7 sections in separate weekly 3-hour sessions each semester, we carefully considered safety, robustness, time constraints, and expenses.

We first applied our DNA extraction and PCR method developed for a tephritid genus to the much smaller laboratory fly, *D. melanogaster* (Fig. 3). No attempt was made to remove RNA. Our tephritid of interest, *Terellia fuscicornis* (artichoke fly) produced a large nucleic acid pellet, and 5% of the total extract created an easily visible stain upon agarose gel electrophoresis. In contrast, the much smaller *D. melanogaster* created a faint smear on the gel, but the nucleic acid pellet was easily visible. If the nucleic acid pellet was lost during washing, the amplification usually failed. Because we had observed some apparent primer–dimer formation in PCR products, we tried a range of annealing temperatures. We found that a wide range of annealing temperatures (40–54 °C) produced high yields of product, but primer–dimer formation was not eliminated without reduced product yield. DNA sequencing typically produced 500 bp of readable sequence after gel purification (data not shown).

Fly mitochondrial DNA is very AT-rich and is the choice of restriction enzyme results in dramatically different numbers of fragments. We choose to test Tru9I (an isoschzomer of MseI), which recognizes and cleaves at TTAA sites, because the ND1 region of dipterans would be expected to be cleaved into ∼12 fragments, of average length near 60 bp. Fragments of this size range are distinguishable by standard acrylamide or low-melt agarose blends and likely distinguish between mitochondrial lineages separated by several million years. We found Tru9I is compatible with diluted PCR reactions, thus removing any need for purification of the PCR product. Enzymes that recognize less common sites could also be used, but would be less likely to create distinguishing patterns. We preferred high-percentage agarose with TBE buffer rather than acrylamide gel electrophoresis, because it is less technically demanding and yields similar good resolution in the expected size range. Although the number of distinguishable patterns is likely to be far less than the number of dipteran species, it allows for hypothetical identifications to be supported or disproved.

To test whether the primers would amplify diverse dipterans and whether the RFLPs produced would be distinct, we assembled an assortment of flies for an undergraduate student volunteer. The results indicated that the PCR-RFLP method functioned with a wide diversity of dipterans, including a nondipteran, which suggests it is likely to work with most flies (Fig. 4*A*). The laboratory *D. melanogaster* is predicted to create fragments of 149, 109, 88, 79, 60, 57, 47, 32, 24, 22, 15, 8, of which the largest 6–10 should be observable. This matches our observations, though there appears to be some longer, possibly incompletely digested fragment in some lanes (Fig. 4A, *lane* 2, and 4*B*, *lanes* 3–5). The local *Drosophila* differs from the laboratory strain, that is, it does not have a band corresponding to 109 bp. To explore the robustness of the method, we tested the effect of fly mass, DNA pellet washing, template concentration, and reproducibility. We found that DNA yields were limiting with small flies such as *Drosophila* and mosquitoes, DNA washing was unnecessary (even with visible salt) when the final resuspension volume was 200 μL, and that the results were reproducible (Fig. 4*B*). Small flies such as mosquitoes should be reprecipitated and resuspended in 20 μL for increased yield. Unextracted fly legs placed directly in the amplification tube had erratic success, often leading to a long smear of products (not shown). We have not yet encountered any fly species whose DNA did not amplify, presumably because the low annealing temperature allows some polymorphisms in the primer sites to be tolerated.

First Semester Experience

In the first session, most students came with houseflies or other local Calyptratae identified with minimal precision and documentation. Students prepared fly DNA extracts and gave their instructor a labeled tube containing 1 μL of extract. To finish the first session within the allotted time, instructors had to schedule procedural

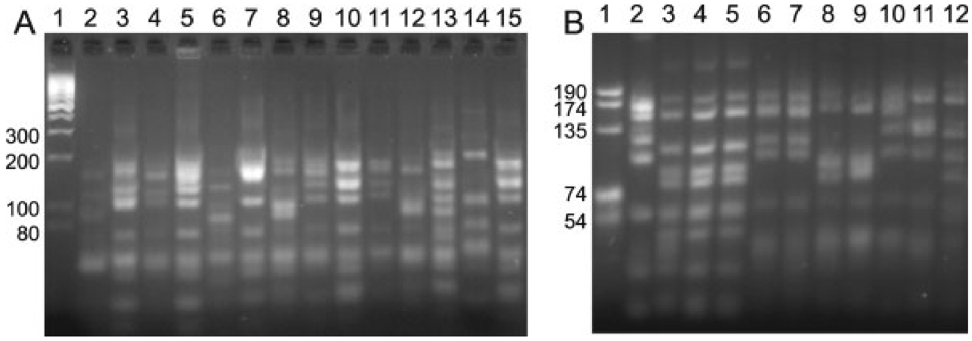


FIG. 4. **Preliminary RFLP survey and variations**. *A*: Diverse flies were extracted, amplified, and digested by an undergraduate volunteer. *Lane* 1, DNA ladder (MassRuler DNA Ladder, Low Range); *lane* 2, *D. melanogaster*; *lane* 3, house mosquito (*culex pipiens*); lane 4, housefly (*Musca domestica*); *lane* 5, local blowfly (Calliphoridae); *lane* 6, local moth fly (Psychodidae); *lane* 7, local Calyptratae, Sarcophagidae maggots; *lane* 8, local *Drosophila*; *lane* 9, local Muscidae; *lane* 10, fruit fly *Chaetostomella cylindrica*; *lane* 11, fruit fly *Chaetostomella cylindrica*; *lane* 12, fruit fly *Terellia fuscicornis*; *lane* 13, fruit fly *Urophora quadrifasciate*; *lane* 14, fruit fly *Chaetorellia succinea*; *lane* 15, lace wing, order Neuroptera. *B*: Variations and repeats of DNA extraction and amplification by instructor. *Lane* 1, DNA ladder (homemade); *lane* 2, 1 μL of 200 μL of 1 washed *Terellia fuscicornis*; *lane* 3, 1 μL of 200 μL of 1 *D. melanogaster* washed pellet; *lane* 4, 1 μL of 200 μL of 8 combined *D. melanogaster* washed pellet; *lane* 5, 1 μL of 200 μL of 8 combined *D. melanogaster* unwashed pellet; *lane* 6, 1 μL of 200 μL of 1 house mosquito (*Culex pipiens*) washed pellet; *lane* 7, 1 μL of 20 μL of 1 reprecipitated *Culex pipiens*; *lane* 8, local *Drosophila*; *lane* 9, *Urophora quadrifasciate*; *lane* 10, local Muscidae; *lane* 11, local *Chaetostomella*; *lane* 12, *Chaetorellia succinea*.

steps strictly. The instructor added the PCR mix to each tube and programmed the thermocycler in front of the students.

In the second session, students retrieved their PCR product tube, added the restriction enzyme mix prepared by the instructor, and then incubated their reactions for 1 hour. During the time the digestions were incubating, the students poured agarose gels and practiced loading gels. When the incubations were complete and the agarose gelled, students added loading solution to their samples and loaded their gels. Every 15 minutes, students observed their gel on a short-wave UV transilluminator and recorded their observations. Students requested as much as 15 minutes each to observe their gels, and they did not have time to run their gels for more than 40 minutes. Meanwhile, students ran their undigested PCR products and plasmid samples from a previous activity on a 1% agarose gel. In each section, the instructor was given a prepared sample and an undigested sample to digest. PCR success was highly variable, but failures correlated to students who reported losing their nucleic acid pellets during washing and sections with first-time instructors. Two students amplified DNA from mosquitoes, but products were weak and difficult to discern. Students without clear bands analyzed their instructor's sample. Student laboratory manuals had report forms for data collection, laboratory observations, and theoretical questions. Examples of student results are shown in Fig. 5*A*. Of six digests, four patterns are seen. Interestingly, four putative houseflies (Fig. 5*A*, *lanes* 4, 8, 6, and 10) show only three patterns, and a putative stable fly (Fig. 5*A*, *lane* 12) presents the same pattern as two putative houseflies (Fig. 5*A*, *lane* 4, 8), presumably reflecting misidentification.

The students were not able to complete their reports by the end of the second session. They finished their reports later but had difficulty with a molecular evolution question on proposing mutations to account for differences between flies (see study question 1). The last question on the report inquired whether the activity met its

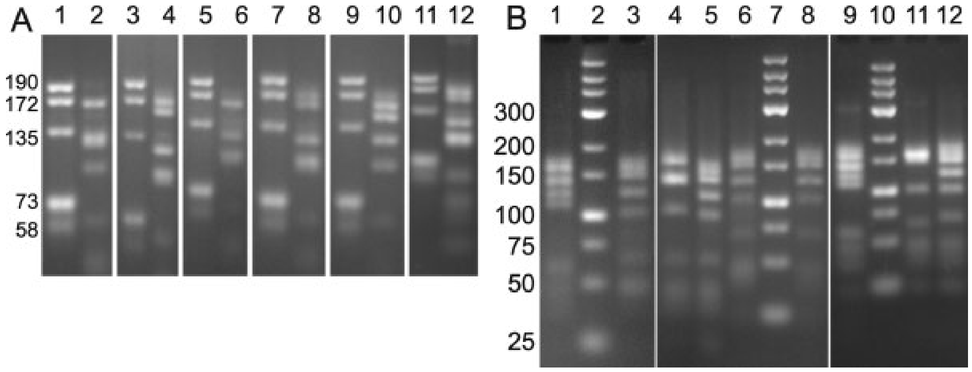


FIG. 5. **Selected students' results**. Student PCR-RFLP results photographed by the instructor. *A*: First semester: Odd number lanes, DNA ladder (home made); *lane* 2, local *Chaetostomella*; *lane* 4, putative housefly; *lane* 6, putative housefly; *lane* 8, putative housefly; *lane* 10, putative housefly; *lane* 12, putative stable fly. *B*: Second semester: *lane* 1, putative housefly; *lanes* 2, 7, 10 DNA ladder (GeneRuler, DNA ladder, Low Range); *lane* 3, putative blow fly; *lane* 4, putative housefly; *lane* 5, putative bottle fly; *lane* 6, putative stable fly; *lane* 7, DNA ladder; *lane* 8, putative housefly; *lane* 9, putative housefly; *lane* 10, DNA ladder; *lane* 11, putative Muscidae; *lane* 12, putative Calliphoridae.

objectives and how the activity could be improved. Of 93 collected reports, 55 agreed that the activity met its objectives, 3 disagreed, 35 made no statements concerning objectives, and 26 expressed unsolicited enthusiasm. The most common suggestion (36/93) was to remove a molecular evolution question, because it was too difficult or took too much time. Some suggestions were peculiar (use only flies of known sequence), but several were useful (do not wash pellet to avoid loss, include a method to estimate time since divergence between two sequences).

Second Semester Experience

Several changes were introduced in the next semester. Students were encouraged to avoid very small flies or to gather a mass equivalent to 10 *D. melanogaster*, logistics were carefully planned with laboratory instructors, the number of gels halved, and a commercial DNA ladder was used. In the first session, washing steps were eliminated to avoid pellet loss and save time. After the first session, all 82 PCR reactions were checked by the instructor. Fifteen were found to be clearly too weak to use, another 15 had only moderate yields, and 52 had high yields. High-yield samples were digested by an instructor. The day before the second session, 4% agarose gels were poured by the instructor and chilled to achieve full resolving power. At the beginning of the second session, digested samples were returned to students, and those students whose samples were deemed weak were given a duplicate sample from students in their section. Students briefly practiced loading gels and then loaded their samples. A commercial molecular weight standard was used, and students were encouraged to photograph their gels with their camera-equipped mobile phones to avoid complaints about transilluminator access. Most students' phones were capable of photographing the image displayed on the back of the instructor's camera, and a few were able to capture good images of the gel directly through the safety shield without filters. Gels were electrophoresed for a maximum of 45 minutes. The molecular evolution question on the report was replaced with a question asking for an original biological question to which PCR-RFLP could be applied. All sessions finished the activity and completed their reports by the end of the session.

Examples of students' results are shown in Fig. 5*B*. Of nine samples shown, five patterns are seen. Four putative houseflies present three different patterns, and one pattern is associated with three different flies, albeit all Calyptratae. The wide range of commercial DNA ladder eased size estimation.

Of 82 collected reports, 66 agreed that the activity met its objectives, 0 disagreed, 11 did not express a clear opinion, and 34 expressed unsolicited enthusiasm, an improvement over the previous semester. The most common suggestion for improvement was no longer more time (6/82), but for more individual equipment (17/82). Interestingly, only one student suggested that the digest can be done by students.

DISCUSSION

We have developed and implemented a two-session PCR-RFLP activity to our undergraduate genetics laboratory. The activity introduces important molecular biology methods in a biological context robustly and inexpensively, and can be linked to many other biological teaching activities. Student's response was appreciative, and students agreed that the activity achieved its learning objectives.

We consider our goal of introducing an inexpensive, robust molecular biology laboratory activity to be successfully achieved. The students used PCR, restriction enzyme digestion, and gel electrophoresis in a rich biological context. Importantly, we wanted to expose students to nonnumerical and ambiguous data, to have students understand the limits of the methods, and to encourage students to conceive original applications. Additionally, we wanted to link the activity to potential coursework in bioinformatics, entomology, evolution, and biodiversity. We originally considered human PCR activities, but we appreciated having our medically oriented students experience important molecular methods in a distant taxon.

Each student was given a laboratory manual that introduced the background theory on PCR, speciation and genetic polymorphism, mitochondria, and RFLPs. The complete procedure was described and students had a 3-page report with written questions on their fly identification, predicted results from the closest sequenced relative, observations, gel observations, analysis, and study questions. The laboratory manual stated the following learning objectives as abilities the student should acquire:

- Explain the principles by which DNA is amplified.
- Recognize and describe the purpose of PCR controls and the likely causes and solutions for PCR failure and contamination.
- Be able to follow simple PCR protocols, from DNA extraction to product.
- Design simple PCR approaches, including primer design, reaction conditions, and product analysis.
- Explain how restriction length polymorphisms can be detected.
- Choose appropriate restriction enzymes to distinguish between sequences by RFLP.
- Interpret agarose electrophoresis data.
- Explain the scientific relevance of the data generated in this activity.
- Explain with this example, how sequences can change without loss of function.
- Explain how molecular biology can be applied to biological questions.

Some objectives were met by the laboratory activity and report completion. Other objectives have been assessed by written questions connected to the lecture course, but we have not documented any specific contributions from this activity. In the second semester, most students clearly agreed, and none disputed, that the objectives were accomplished. We are greatly encouraged that a majority of students expressed enthusiasm without being asked, and we plan to maintain the activity and to expand it into additional activities.

We asked 2 study questions:

1) What is the minimum number of single nucleotide changes to the *D. melanogaster* sequence necessary to produce the results you observe? Discuss whether the sequence changes may be silent, neutral, or conservative.
2) Propose an original biological question and explain how this technique can be used to address it.

We note many potential study questions are easily generated, such as

- Which amino acid sequences can exist at a Tru9I site in an open reading frame?
- Cutting at which 4 bp palindrome would generate the most fragments?
- Assuming all known extant mutations are selectively neutral, use the sequences to rank amino acid similarity.
- How should a standard mutation rate of 1.7% per million years be converted to a number of Tru9I sites changed per million years?
- How can PCR conditions and primers be chosen to maximize the likelihood of success when the target sequence is not known?

Interestingly, many students verbally expressed frustration with not having a close match to a sequenced fly and the difficulty of properly recording gel data. We speculate that the paucity of available sequences and the nonnumerical nature of the data will deepen students' understanding of gaps in existing knowledge, the nature of experimental research, and the use of evidence that disproves, rather than proves, hypotheses. Many aspects of the activity allow for illustration of important concepts, such as the nonstandard genetic code, variations in mitochondrial genomes, the origin and consequences of polymorphisms, speciation, and controversies in classification. Finally, we note that this activity could easily be transformed into project-based learning where students would use this method to address a biological question with written and oral presentations.

Expense

When each student analyzes one fly, and eight students share one 4% agarose gel, we estimate costs of about \$3 per student (Tru9I ~\$1, agarose ~\$1, dNTPs ~\$0.25, DNA ladder ~\$0.25, chemicals, disposables, <\$0.50). Costs can be halved by assigning 1 digest per student pair and using fewer agarose gels of less percentage. We have used homemade polymerase and homemade DNA ladder to save funds, but use of commercial PCR master mixes should increase costs by <\$0.50 per student. Reduced amounts of Tru9I would also save costs, but we sometimes had incomplete digestions using less enzyme. Few inexpensive restriction enzymes exist that would cleave the amplified sequence, and the use of inexpensive normal agarose would reveal far less complex and distinguishing patterns.

Student Pitfalls

In the first semester, students' mistakes in DNA extraction prevented some students from observing the digest they loaded on the gel. Our students were in their third or fourth semester of university, and their only relevant prior experience was a plasmid mini prep. Robustness of the activity was insured by the inclusion of verified instructor samples for those students. In the second semester, we used the time between the sessions to check amplification products, select those that had complete yields, digest them, and redistribute successful reactions to all students. Confined to a 3-hour session, we decided the missing activities of adding the Tru9I master mix and pouring gels was more than offset by the increased robustness, time for gel observation, and time for analysis. The students were scheduled to pour agarose gels in a later activity, and the total time taken by one skilled instructor to check student PCR reactions, prepare one Tru9I master mix, and perform the digests, was only 3 hours. This format also suggests that precisely identified local flies could be redistributed as unknowns for students to match to their predictions.

Finally, we consider how the activity can be developed into original research projects for undergraduates. There is a large diversity of flies in Lebanon, including an *Anopheles* malaria vector, a sandfly leishmania vector (also linked to a recent outbreak of sandfly fever virus), as well as mosquito vectors of West Nile Encephalitis. Molecular typing of these vectors or other local flies would be appropriate for sustained undergraduate research projects, and it is assumed that most teaching institutions will have similarly relevant local research opportunities. We intend to link precise fly identification from our entomology course to this activity in our genetics course and include sequencing and submission of novel sequences to GenBank in our bioinformatics course. Lastly, we note that with appropriately designed primers, the activity could be converted from flies to other genes and taxa.

Acknowledgments— We thank instructors Akram Ghantous, Layane Hajjar, and Ansam Sinjab for assistance in conducting the laboratory sessions.

REFERENCES

[1] C. A. Smith, M. S. Al-Zein, N. P. Sayar, K. K. Knio (2008) Host races in *Chaetostomella cylindrica* (Diptera: Tephritidae): *Genetic and behavioural evidence,* Bull. Entomol. Res. (in press).

[2] P. Martin, M. Muruke, K. Hosea, A. Kivais, N. Zerwas, C. Bauerle (2004) A rapid PCR-RFLP method for monitoring genetic variation among commercial mushroom species, *Biochem. Mol. Biol. Ed.* **32**, 390–394.

[3] S. L. Cameron, C. L. Lambkin, S. C. Barker, M. F. Whiting (2007) A mitochondrial genome phylogeny of Diptera: Whole genome sequence data accurately resolve relationships over broad timescales with high precision, *Systematic Entomol.* **32**, 40–59.

[4] D. K. Yeates, B. M. Wiegmann (1999) Congruence and controversy: Toward a higher-level phylogeny of Diptera, *Annu. Rev. Entomol.* **44**, 397–428.

[5] Y. Satta, S. I. Chigusa, in M. Kimura, N. Takahata, Eds. (1991) New Aspects of the Genetics of Molecular Evolution, Japan Scientific Societies Press, Tokyo, pp. 117–134.

[6] K. H. Wolfe, P. M. Sharp, W. H. Li (1989) Mutation rates differ among regions of the mammalian genome, *Nature* **337**, 283–285.

[7] A. Caccone, G. D. Amato, J. R. Powell (1988) Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup, *Genetics* **118**, 671–683.

[8] K. Tamura, S. Subramanian, S. Kumar (2004) Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks, *Mol. Biol. Evol.* **21**, 36–44.

[9] D. R. Engelke, A. Krikos, M. E. Bruck, D. Ginsburg (1990) Purification of *Thermus aquaticus* DNA polymerase expressed in *Escherichia coli*, *Anal. Biochem*. **191**, 396–400.

[10] P. T. Smith, K. Krager, J. T. Cronin, S. Kambhampati (2002) Mitochondrial DNA variation among host races of *Eurosta solidaginis* Fitch (Diptera: Tephritidae), *Mol. Phylogenet. Evol.* **25**, 372–376.