BIOCHEMISTRY AND MOLECULAR BIOLOGY EDUCATION
Vol. 32, No. 2, pp. 108–114, 2004

Laboratory Exercises

An Hypothesis-driven, Molecular Phylogenetics Exercise for College Biology Students*

Received for publication, June 23, 2003, and in revised form, November 19, 2003

Joel D. Parker‡§¶, Robert E. Ziemba‡‖, Sara Helms Cahan‡¶, and Steven W. Rissing‡**

From the ‡Department of Biology, Arizona State University, Tempe, Arizona 85287

This hypothesis-driven laboratory exercise teaches how DNA evidence can be used to investigate an organism's evolutionary history while providing practical modeling of the fundamental processes of gene transcription and translation. We used an inquiry-based approach to construct a laboratory around a nontrivial, open-ended evolutionary question about the relationship of five species of *Drosophila*. In the course of answering this question, students at the early college biology level learn how the information in DNA can be extracted and used by both the cell and scientists. This dual proximate-ultimate approach introduces students to the techniques of PCR, DNA sequencing, and phylogenetic sequence analysis while simultaneously providing a concrete pen-and-paper model of the cellular processes of transcription and translation. The laboratory has been successfully employed over 3 years with first-year college students and has proven its versatility by being easily adapted to a "dry lab" form with advanced high school students.

Keywords: DNA sequencing, phylogenetics, transcription, translation, desert *Drosophila.*

One of the main challenges of introductory biology courses is establishing connections between scientific concepts across different levels of organization, from genes to organisms to ecosystems. This is particularly the case for genetics, where the mechanisms, function, evolutionary dynamics, and phylogenetic value of genes are often presented as disjointed subjects throughout a typical survey course. Here we describe an hypothesis-driven laboratory exercise, appropriate for students who have learned the basics of DNA structure and function, designed to promote more integrated understanding of the use of molecular genetics in evolutionary biology. In this three- to four-period exercise, students extract, amplify, and sequence DNA in order to determine the phylogenetic relationships among five species of *Drosophila*. In so doing, they are introduced to the concepts of phylogenetics while reinforcing the proximate mechanisms of gene function.

The fruit fly *Drosophila melanogaster*, routinely used in genetics laboratories for a century, has many interesting and diverse relatives. A number of species of *Drosophila* are endemic to the deserts of southwestern North America. These cactophilic species are all highly adapted to xeric conditions, but vary considerably in ecological, morphological, and reproductive traits, making them ideal taxa for molecular phylogenetic investigation [1–3]. Here we use *D. melanogaster* and four of these desert-dwelling relatives to understand more about the nature of the information stored in DNA sequences. Students use ecological and physiological data provided by the instructor, along with their own morphological observations, to generate hypothetical relationships among the species. To test their hypotheses, the students sequence a part of the mitochondrial cytochrome *b* gene. They then interpret and analyze the nucleotide and protein sequence information. The result is a satisfying answer to their question, practical experience with DNA sequencing techniques, and reinforcement of the proximate mechanisms of gene expression. A modified version of this laboratory, omitting the more technical and expensive PCR and sequencing, has also been successfully employed. This "dry lab" version relies on a sample dataset that is distributed to the students directly after hypothesis generation.

* This laboratory was designed and tested with the assistance of grants from The Hughes Foundation and the Math-Science Honors Program run by the Strengthening Understanding of Mathematics and Science Institute at Arizona State University. J. D. P. was supported by the Achievement Rewards for College Students Foundation during the development of the laboratory. Manuscript preparation was supported by grants to Laurent Keller from the AETAS Foundation for Research into Aging, The Fondation A.R.&J. Leenards (Lausanne), and the Swiss National Science Foundation. The cost of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed. E-mail: joel.parker@ie-zea.unil.ch.

¶ Current address: Department of Ecology and Evolution, Biology Building, University of Lausanne, CH-1015, Lausanne, Switzerland.

‖ Current address: Department of Biology, Centre College, 600 West Walnut Street, Danville, Kentucky 40402.

** Current address: Director, Introductory Biology Program, Botany and Zoology Building, The Ohio State University, Columbus, Ohio 43210-1293.

EXPERIMENTAL PROCEDURES

General Outline—We divided the classes into five groups no bigger than four persons each to encourage discussion and facilitate group learning. The laboratory was organized into four 2.5-h sessions. We found that this format divided the information content into discrete and manageable units, as well as allowing for necessary but time-consuming steps (PCR, sequencing) between sessions.

Session 1: Hypothesis generation (1 h), DNA isolation/PCR (1.5 h)
Session 2: PCR visualization and purification (2 h)
Session 3: Transcription and translation (2.5 h)
Session 4: Phylogenetic analysis (2.5 h)

The number of laboratory periods required can be reduced by combining sessions 3 and 4 or by having some components of the process (*e.g.* visualization and purification) performed by the instructor outside of class time. Alternatively, extraction and sequencing of the DNA can be omitted entirely (the dry lab version) by using the sample sequences provided, which can reduce the exercise to two laboratory sessions.

Major Equipment Required—All of the major equipment can be found in most molecular biology research laboratories. For DNA isolation and PCR, we used a microcentrifuge, PCR machine, and pipettes. For electrophoresis, we used an agarose gel electrophoresis box, power supply, ultraviolet (UV)[1] light box, laboratory coats, latex gloves, UV face shield, and pipettes. Access to a sequencing facility or a fast commercial sequencing laboratory is also required for sequencing the student's samples. Other required materials include the flies (*D. arizonae, D. melanogaster, D. mettleri, D. mojavensis, D. nigrospiracula*, available from the *Drosophila* Stock Center, Tucson, AZ; stockcenter.arl.arizona.edu), dissecting scopes, note cards, Fly Nap, drawing paper, 0.5-inch graph paper cut into strips two rows high, four different color highlighters, and cellophane tape. There are example class handouts in the "Appendix." A supplementary material file containing a printable sample set of color ABI sequence files for all species, the aligned sequences, and distance matrices in several formats are available from any of the authors by request.

Hypothesis Generation—Live representative flies from each species, anesthetized by brief exposure to Fly Nap, were given to the students for examination under dissecting scopes. In the case of the shortened dry lab version (omitting actual DNA isolation and sequencing), flies stored in alcohol can be substituted for the live flies. Handout 1 provides an ecological and a reproductive trait [1–3] and asks the students to complete a third column based on their own morphological observations. Using these data, each group was asked to agree on one best guess as to the correct evolutionary relationships among the fly species and to present their hypothesis as a tree to the rest of the class. The trees were posted and left up to be tested in "DNA Visualization and Purification."

DNA Isolation/PCR—PCR requires very little DNA, hence we used a very basic isolation protocol with "squishy buffer" [4]. One fly was placed into a PCR tube and squished for 5–10 s with a pipette tip in 50-μl squishing buffer (10 mM Tris-Cl, pH 8.2, 1 mM EDTA, 25 mM NaCl, and 200 μg/ml proteinase K freshly diluted). The flies were then digested for 20 min at 37 °C followed by a heat denaturation step (95 °C for 2 min) to inactivate the proteinase K. These incubations are most easily done in a PCR machine.

We used two universal primers, CB1 (5′-TATGTACTACCATGAGGACAAATATC-3′) and CB2 (5′-ATTACACCTCCTAATTTATTAGGAAT-3′) [5], to amplify a part of the coding region of the mitochondrial cytochrome *b* gene of the fly. Reactions were set up for 50 μl to minimize pipetting errors with 1 μl of the DNA prep, 500 μM primers, 250 μM dNTPs, 1× reaction buffer, and 1 μl Taq polymerase (Invitrogen, San Diego, CA). It is also a good idea to include a negative and positive PCR control that you have tested before class. The program used was denaturation at 92 °C for 2 min followed by 30 cycles of 30 s denaturation at 92 °C, 30 s annealing at 55 °C, 1 min extension at 72 °C, followed by a 5-min extension at 72 °C, then hold at 4 °C on a Perkin-Elmer 2400 PCR machine (Perkin-Elmer, Wellesley, MA). Fortunately, PCR does not require high-quality DNA, so the experiment is extremely resistant to errors in isolation. The most important factor in success is making sure that the students label their tubes clearly and do not mix up the flies.

DNA Visualization and Purification—Five microliters of PCR product was checked on a 1% agarose gel stained with ethidium bromide. This should only be attempted by those who are aware of the associated hazards and safety precautions. We let the students load, run, and visualize the gels under close supervision with laboratory coats, gloves, and a UV face shield. Ethidium bromide is a likely teratogen and mutagen and must be disposed of appropriately. There is a risk of electric shock with the gel box and burns to exposed skin and eyes with the UV light box.

Reactions producing single clean bands were then purified with spin columns as described by the manufacturer (Micron, Westboro, MA). Our sequencing was done by a core facility on an ABI 377 (Applied Biosystems, Foster City, CA) with DNA and primer concentrations as per their directions. These conditions will vary across sequencing facilities and companies.

Transcription and Translation—Each group was given a handout with the invertebrate mitochondrial codon table and ABI color printouts of the cytochrome *b* forward (CB1) and reverse (CB2) sequence output files generated from their sample. If the output of any of the groups were not usable, the sample files for the same species were substituted. They were asked to write the sequence 5′ to 3′ in the CB1 direction on strips of graph paper. The strips were taped together as they went. The graph paper is used to keep the sequences written at equal intervals so that the strips can eventually be copied and aligned from bench to bench. The students were told that they need to reverse complement the CB2 sequence to check the areas of CB1 that are hard to interpret or show very small peaks. All of the sample sequences have problematic areas, and it is essential that both printouts be used. When reading ABI files, watch for even spacing, low peak heights for "G," and problems with repeats of the same base pairs. If the class is doing the sequencing, then they must use CB1 and CB2 printouts from exactly the same fly due to possible allelic differences. When a group made an insertion or deletion reading error, we had them cut and paste the tape to avoid having to rewrite the entire sequence downstream of the error. Once they had all of the sequence, the primer sequences were removed. They then taped this strip to their bench top and proceeded to translate the sequence using the provided codon table (see supplementary materials), recording the amino acids on another strip of graph paper as above. At this point, note that U = T in the table provided because this is an mRNA table. When the first student noticed a problem with identifying the reading frame, we passed out the honey bee cytochrome *b* protein sequence to help guide them (GenBank P34845, positions 142–285). The protein sequence strip was then taped to the other side of the bench.

Phylogenetic Analysis—The students copied each of their sequences onto other strips of paper and then took them around to the other groups to count the number of nucleotide differences and the number of amino acid differences between their species and each of the other four species. The students were instructed to highlight the differences with color highlighters. Two distance tables were then drawn on the black board (one for nucleotides, one for proteins), and the students filled in the squares with raw difference counts. At this point, we calculated tables of either proportion distance *p* (number of differences divided by total number of bases compared) or the more complicated Jukes-Cantor corrected distance for the nucleotide sequences ($d = -(3/4)\ln[1 - (3/4)p]$ and Poisson corrected distance for the protein distance ($d = -\ln(1 - p)$). The choice depends on the level of the class. The latter two distances assume a Poisson distribution of mutations, and the Jukes-Cantor distance corrects for multiple

[1] The abbreviations used are: UV, ultraviolet; UPGMA, unweighted pair-group method using arithmetic averages.

changes at the same positions. These formulae and assumptions are given in handout 3. Derivations can be found in most phylogenetic textbooks (see Chapters 2 and 3 of Nei and Kumar [6] for example). As a class, we went through the construction of a tree with the unweighted pair-group method using arithmetic averages (UPGMA) for the two datasets and posted the trees.

For the first time, students should do all of these calculations by hand with a calculator instead of relying on computer programs that they will not immediately understand. For instructors, there are many computer packages available that can do these calculations and construct a UPGMA tree. We used Mega2.1 (free at www.megasoftware.net), which is specifically designed for distance-based analysis to analyze and check the calculations. The sample data and results are provided in Mega format files in the supplementary materials. UPGMA tree construction is done by taking the grouping with the smallest difference, grouping them, and drawing a node at a depth of half the distance. We then crossed out the distance from the matrix and repeated the procedure. When multiple taxa are connected, their distances are averaged. For example, the first grouping with the Jukes-Cantor distance in Table I places the node between *D. arizonae* and at a depth of $0.05/2 = 0.0025$ or rounded to 0.002 in Fig. 2*A*. The next node is at a depth of the average distance between *D. nigrospiracula* and *D. arizonae*, *D. nigrospiracula* and *D. mojavensis*, or $[(0.123 + 0.123)/2]/2 = 0.0615$, which was rounded to 0.061 by Mega2.1. Notice that the branch lengths sum to 0.061 from that node to *D. arizonae* and *D. mojavensis* ($0.059 + 0.002$), equaling the branch distance to *D. nigrospiracula*. The distances between taxa are the total distance going back in time from one taxon to the common ancestor then forward to the other taxon. An example of the UPGMA procedure for the students is given in handout 4. Once we have both the nucleotide and protein trees, we then compared these trees to the trees predicted by the student groups.

RESULTS

Expect one clean band at 485 bp from all of the PCRs. The aligned and edited sequences are shown in Fig. 1, revealing mostly conservative amino acid changes and silent substitutions. A table of the raw differences and distances reveals a lower number of amino acid substitutions relative to nucleic acid changes (Table I). The final UPGMA tree recovered after converting to Jukes-Cantor and Poisson distances reveal that in this case, the reproductive trait and the sequence data closely agree (Fig. 2). The same topology is recovered if *p* distances are used. The greatest problems encountered were mixed-up samples and bad sequencing runs, resulting in the need to resort to the sample data set; however, it is worth noting that these sample data presented here and in the supplementary files were student-generated in an actual class.

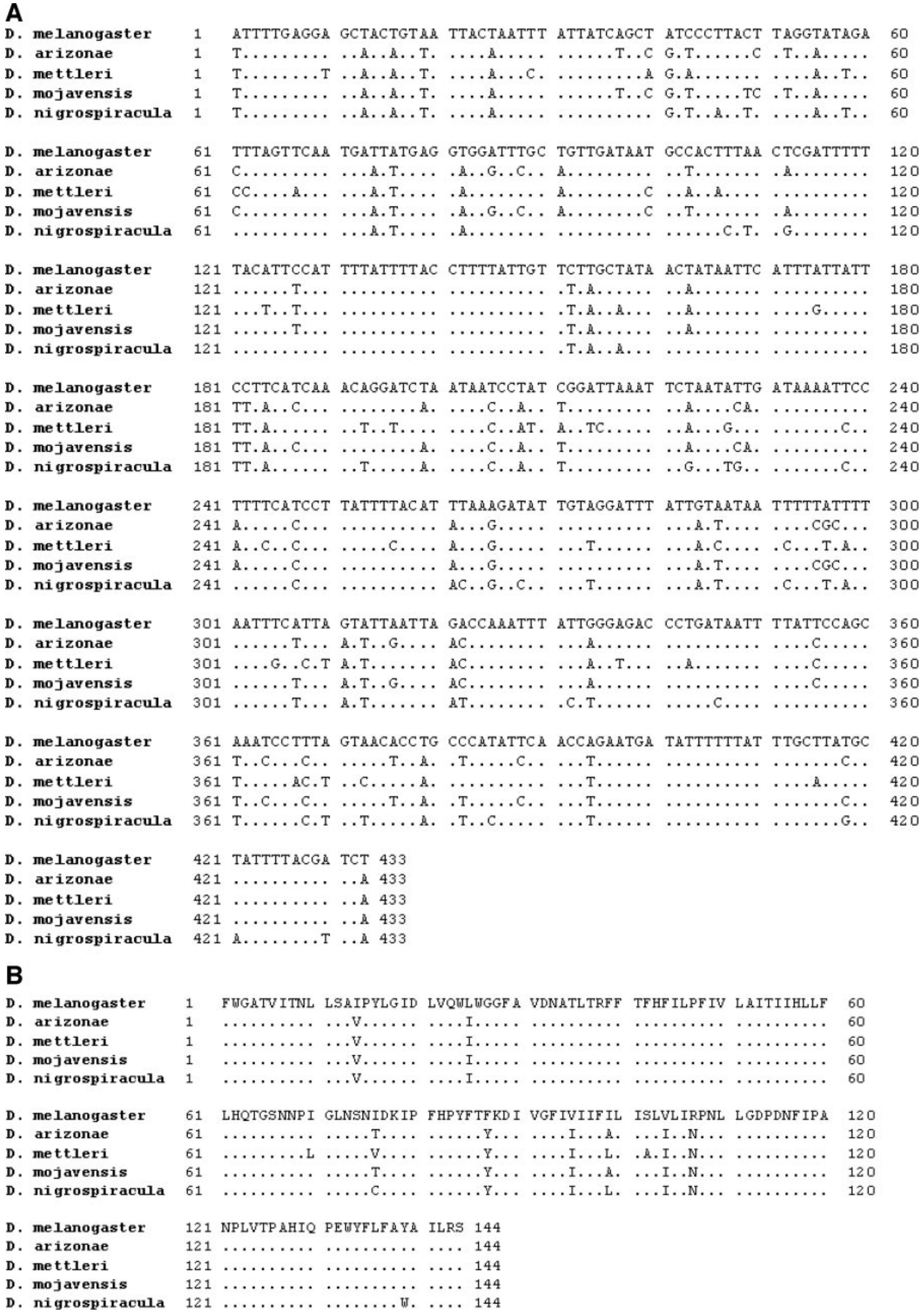
DISCUSSION TOPICS

How Are Desert Drosophila Related to One Another?—The first handout is a table describing an ecological character and a reproductive character with space for students to add morphological characters of the five species. The ecological character is the substrate where the larvae develop. These larva feed on yeast growing in the rotting plant tissue. Cactus have extensive chemical defenses and the flies have adapted to deal with the chemical mix in their own particular hosts [1]. One can suggest to the students that more closely related flies might share similar ecological constraints and urge them to think about how expansion into another environment might be related to speciation. It is helpful to have pictures or a plant handbook with the various cacti available in the classroom. Incompatible reproductive organs can be a strong species barrier, and hence the evolution of reproductive traits can also be important for species isolation. The reproductive trait is the total amount of protein incorporated into the female's body from proteins deposited by the male with sperm [3]. This trait correlates with size and morphology of both male and female reproductive organ structure. Females of many insects metabolize and incorporate the protein in ejaculate to help offset the cost of making eggs. In this study, incorporated proteins were measured by radioactively labeling the proteins from the male; hence the units are in decompositions per minute (dpm). Finally, the instructor can point out that most insect taxonomy is based on morphological characters. The large majority of insect species are defined by experts measuring and comparing morphological characters in museum specimens. For less advanced classes, we found it useful to begin by describing *D. melanogaster* as the outgroup and to build the trees with this species as the root. Doing it this way makes the trees easier to compare with the final trees that will be obtained by UPGMA. The goal of the discussion is to generate as much controversy within and among student groups as to what the correct relationship will be and to obtain alternative trees for the groups. Systematics is one of the most contentious fields in biology and controversy is the norm. No one knows with absolute certainty the correct answer as to the relationship among these species. In fact, this laboratory exercise is the one of the first published DNA sequence studies ever to address this particular question. The posted trees will eventually be evaluated against the recovered molecular phylogenies from "Hypothesis Generation."

DNA Isolation/PCR—The students were given a lecture about how PCR works, noting that CB1 binds 5′ or upstream of CB2. It is essential to introduce the polarity of each DNA strand and exactly how each primer is binding as this will be a very important point later on. Going into details about the contents of the reaction mix will depend on the level of the class. The migration of DNA in an electric field needs to be explained as well as how the agarose gel separates DNA molecules according to size as these principles apply to sequencing gels.

Transcription and Translation—This exercise has the hidden agenda of reinforcing the concepts of transcription and translation. The CB1 sequence, which is 5′ of CB2 relative to the start of the gene, corresponds to the mRNA sequence (substituting U for T) and is in the conventional orientation used in the literature and databases. We highlighted the analogy of the two ABI printouts to the double-stranded DNA, the first paper tape to mRNA, and the translated tape to a protein showing the students that they have effectively mimicked the processes of transcription and translation. We also introduced and discussed more detailed aspects of this process, including where and how transcription begins and ends, how the correct reading frame is determined, and the causes and consequences of redundancy in codon usage.

Alignment, Distance Calculation, and Tree Construction—We discussed why certain subsets of species



A

```
D. melanogaster     1   ATTTTGAGGA GCTACTGTAA TTACTAATTT ATTATCAGCT ATCCCTTACT TAGGTATAGA  60
D. arizonae         1   T......... ..A..A..T. ....A..... ......T..C G.T......C .T..A.....  60
D. mettleri         1   T........T ..A..A..T. ....A...C. .........A G.A....... ....A..T..  60
D. mojavensis       1   T......... ..A..A..T. ....A..... ......T..C G.T.....TC .T..A.....  60
D. nigrospiracula   1   T......... ..A..A..T. ....A..... .......... G.T..A..T. ....A..T..  60

D. melanogaster     61  TTTAGTTCAA TGATTATGAG GTGGATTTGC TGTTGATAAT GCCACTTTAA CTCGATTTTT  120
D. arizonae         61  C......... ...A.T.... .A..G..C.. A......... ..T....... .A........  120
D. mettleri         61  CC....A... ...A.T.... .A........ A........C ..A..A.... ..........  120
D. mojavensis       61  C......... ...A.T.... .A..G..C.. A........C ..T....... .A........  120
D. nigrospiracula   61  .......... ...A.T.... .A........ .......... ......C.T. .G........  120

D. melanogaster     121 TACATTCCAT TTTATTTTAC CTTTTATTGT TCTTGCTATA ACTATAATTC ATTTATTATT  180
D. arizonae         121 ......T... .......... .......... .T.A...... ..A....... ..........  180
D. mettleri         121 ...T..T... .......... .......... .T.A..A... ..A....... ....G.....  180
D. mojavensis       121 ......T... .......... .......... .T.A...... ..A....... ..........  180
D. nigrospiracula   121 .......... .......... .......... .T.A..A... .......... ..........  180

D. melanogaster     181 CCTTCATCAA ACAGGATCTA ATAATCCTAT CGGATTAAAT TCTAATATTG ATAAAATTCC  240
D. arizonae         181 TT.A..C... ........A. ....C..A.. T......... ..A....CA. ..........  240
D. mettleri         181 TT.A...... ..T..T.... ....C..AT. A..TC..... ..A...G... .......C..  240
D. mojavensis       181 TT.A..C... ........A. ....C..A.. T......... ..A....CA. ..........  240
D. nigrospiracula   181 TT.A...... ..T.....A. ....C..A.. T......... ..G...TG.. .......C..  240

D. melanogaster     241 TTTTCATCCT TATTTTACAT TTAAAGATAT TGTAGGATTT ATTGTAATAA TTTTTATTTT  300
D. arizonae         241 A.....C... .......... A...G..... .......... ...A.T.... ....CGC...  300
D. mettleri         241 A..C..C... .....C.... A...G..... ...T...... ...A.C.... .C...T.A..  300
D. mojavensis       241 A.....C... .......... A...G..... .......... ...A.T.... ....CGC...  300
D. nigrospiracula   241 ......C... .......... AC..G..C.. ...T...... ...A.T.... .C...T.A..  300

D. melanogaster     301 AATTTCATTA GTATTAATTA GACCAAATTT ATTGGGAGAC CCTGATAATT TTATTCCAGC  360
D. arizonae         301 ......T... A.T..G.... AC........ ...A...... .......... ....C.....  360
D. mettleri         301 ....G..C.T A.T....... AC........ ...A..T... ..A....... ....C.....  360
D. mojavensis       301 ......T... A.T..G.... AC........ ...A...... .......... ....C.....  360
D. nigrospiracula   301 ......T... A.T....... AT........ .C.T...... .....C.... ..........  360

D. melanogaster     361 AAATCCTTTA GTAACACCTG CCCATATTCA ACCAGAATGA TATTTTTTAT TTGCTTATGC  420
D. arizonae         361 T..C...C.. .....T..A. .T.....C.. ...T...... .......... .......C..  420
D. mettleri         361 T.....AC.T ..C.....A. .......... ...T...... .......... ....A.....  420
D. mojavensis       361 T..C...C.. .....T..A. .T.....C.. ...T...... .......... .......C..  420
D. nigrospiracula   361 T......C.T ..T.....A. .T..C..... ...T...... .......... .......G..  420

D. melanogaster     421 TATTTTACGA TCT 433
D. arizonae         421 .......... ..A 433
D. mettleri         421 .......... ..A 433
D. mojavensis       421 .......... ..A 433
D. nigrospiracula   421 A........T ..A 433
```

B

```
D. melanogaster     1   FWGATVITNL LSAIPYLGID LVQWLWGGFA VDNATLTRFF TFHFILPFIV LAITIIHLLF  60
D. arizonae         1   .......... ...V...... ....I..... .......... .......... ..........  60
D. mettleri         1   .......... ...V...... ....I..... .......... .......... ..........  60
D. mojavensis       1   .......... ...V...... ....I..... .......... .......... ..........  60
D. nigrospiracula   1   .......... ...V...... ....I..... .......... .......... ..........  60

D. melanogaster     61  LHQTGSNNPI GLNSNIDKIP FHPYFTFKDI VGFIVIIFIL ISLVLIRPNL LGDPDNFIPA  120
D. arizonae         61  .......... .....T.... ......Y... ....I...A. ...I..N... ..........  120
D. mettleri         61  .........L .....V.... ......Y... ....I...L. .A.I..N... ..........  120
D. mojavensis       61  .......... .....T.... ......Y... ....I...A. ...I..N... ..........  120
D. nigrospiracula   61  .......... .....C.... ......Y... ....I...L. ...I..N... ..........  120

D. melanogaster     121 NPLVTPAHIQ PEWYFLFAYA ILRS 144
D. arizonae         121 .......... .......... .... 144
D. mettleri         121 .......... .......... .... 144
D. mojavensis       121 .......... .......... .... 144
D. nigrospiracula   121 .......... ........W. .... 144
```

FIG. 1. **Aligned DNA (*A*) and translated protein sequences (*B*) from the five species of *Drosophila*.** The reading frame begins at +2 in the DNA sequence.

tended to have the same specific mutations and introduced the concept of common ancestry. This led naturally to the idea that more genetically similar species should share a more recent common ancestor. We then discussed any patterns or properties of the sequences that the students noticed. Two patterns are particularly important. First, we asked them why more changes were at the third codon position than at the first or second and intro-

TABLE I
Jukes-Cantor and Poisson corrected distances for the sample data set

	D. melanogaster	*D. arizonae*	*D. mettleri*	*D. mojavensis*	*D. nigrospiracula*
Nucleotide Jukes-Cantor corrected distances[a]					
D. melanogaster	–	0.162	0.197	0.168	.0162
D. arizonae	63 (0.145)	–	0.142	0.005	0.123
D. mettleri	75 (0.173)	56 (0.129)	–	0.142	0.136
D. mojavensis	65 (0.150)	2 (0.005)	56 (0.129)	–	0.123
D. nigrospiracula	63 (0.145)	49 (0.113)	54 (0.125)	49 (0.113)	–
Protein Poisson corrected distances[b]					
D. melanogaster	–	0.057	0.072	0.057	0.065
D. arizonae	8 (0.056)	–	0.028	0	0.021
D. mettleri	10 (0.069)	4 (0.028)	–	0.028	0.028
D. mojavensis	8 (0.056)	0 (0)	4 (0.028)	–	0.021
D. nigrospiracula	9 (0.063)	3 (0.021)	4 (0.028)	3 (0.021)	–

[a] Nucleotide differences and proportion difference of 433 sites in parentheses.
[b] Protein differences and proportion difference of 144 sites in parentheses.

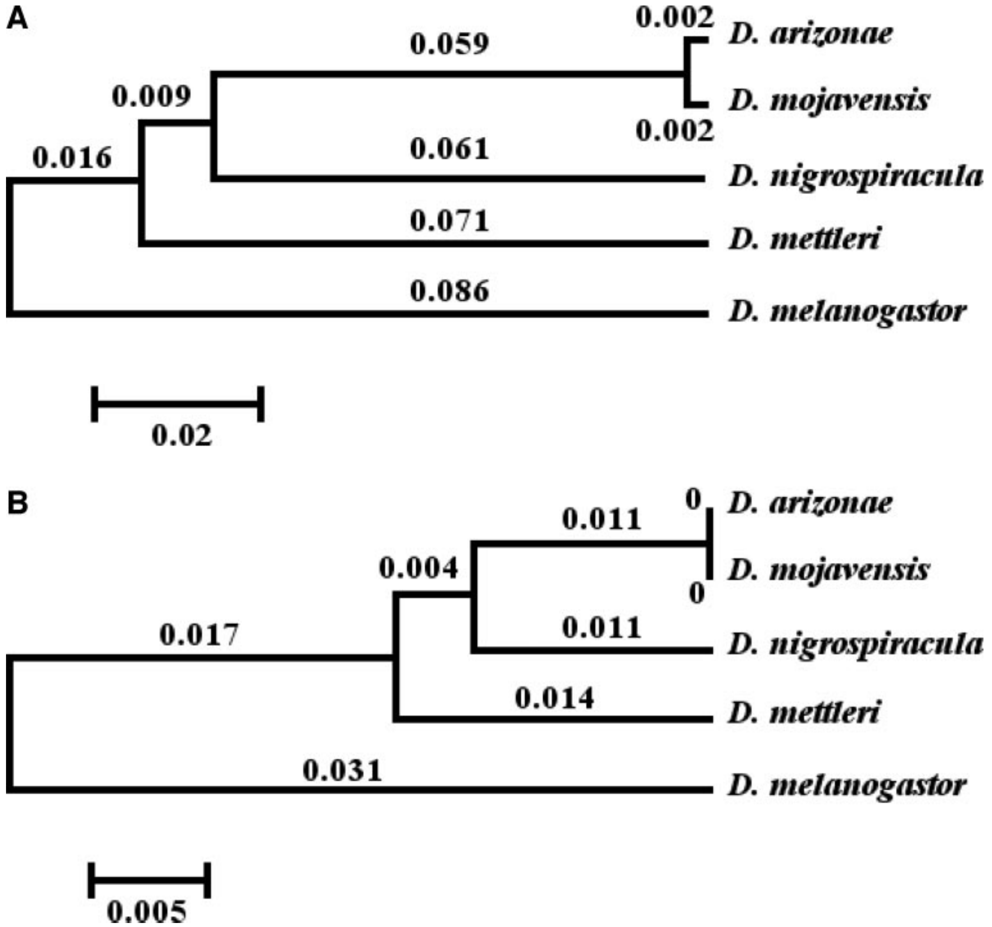


FIG. 2. **The UPGMA trees constructed with Jukes-Cantor distances for the nucleotide data (*A*) and with Poisson distances for the protein data (*B*) are congruent.** The topology is generally as predicted by the reproductive traits in handout 1.

duced the idea of synonymous and nonsynonymous mutations. Here we explained that selection can only see the protein changes, so even with an initial distribution of random mutations over time, those with a harmful phenotype are more likely to be removed. At this point the assumptions underlying the distance definitions and tree reconstruction should be discussed. Most importantly, these include constant mutation rates across all lineages and that any position has an equally random chance to change to any other character. The first invokes the controversial idea of the molecular clock and the latter is clearly a simplification that would not be expected to hold for the functional reasons cited above, but is good enough for many cases. The molecular clock could be a good discussion point by questioning whether selection and mutation are constant for each species. An instructor could ask how colonization, repeated bottle necks in population size, and selective sweeps might effect the clock. If the class is advanced enough to understand basic probability, one can explain the rationale for the Jukes-Cantor and Poisson corrections [6]. The formulae and the major assumptions they are based upon are included in handout 3. For the most advanced classes, one could present other distances that take into account synonomous/nonsynon-

Table II
Example table for Handout 1 with ecological reproductive characters for the sample taxa

Species	Ecological character (where their larva develop)	Reproductive character (amount of sperm proteins incorporated by the female)	Morphological characters
		dpm	
D. arizonae	Rotting prickly pear cactus fruit or pads	185	
D. melanogaster	Rotting soft sweet fruit (grapes, bananas, etc.)	34	
D. mettleri	Soil soaked in juices of rotting cardón cactus, sagauro cactus, or organ pipe cactus	68	
D. mojavensis	Rotting organ pipe cactus, California barrel cactus, or agria	446	
D. nigrospiracula	Rotting cardón cactus or saguaro cactus	49	

mous mutations, transition/transversion ratios, or protein distance matrices [6].

For lower-level classes, even the UPGMA tree method can be too complicated. Instead, we simply connected the most closely related groups sequentially from closest to farthest and rooted them with *D. melanogaster*. In other words, ignore the branch lengths and have the students just reconstruct the topology. This approach was effectively used with the high school groups. Finally, we compared the molecular trees to the hypothesized trees generated earlier. The students were asked which traits they think are most likely to reveal the true evolutionary relationship and why. We also addressed the strengths and weaknesses of the approaches and pointed out that the mitochondrial tree is not necessarily the best or final answer.

GENERAL CONCLUSION

Phylogenetics has gone through a renaissance with the advance in DNA sequencing technology, and a basic understanding of this process should be a part of any undergraduate introductory biology course. It is also a tool of growing importance to medicine for tracking the evolution of viruses and resistant bacteria as well as for mining information from inherited disorders. The full version of this laboratory was taught over 3 years in three separate sections each of freshman biology students, while the dry lab version (“Hypothesis Generation,” “Transcription and Translation,” and “Phylogenetic Analysis”) was successfully employed with an advanced summer high school class. An open-ended, question-driven pedagogical approach helps hold the students attention and is a more realistic way to teach science as hypothesis testing. The specific example could be changed and the same approach used for any other phylogenetic question if the materials and sequencing printouts are available. This laboratory shows how such an approach can be successfully employed, even with a subject with a high density and diversity of information.

Acknowledgments—We thank Greg Hocutt and Theresa Markow for providing the flies and background information and Scott Bingham and Kevin Garvey at the Arizona State University sequencing facility.

REFERENCES

[1] W. B. Heeb (1978) *Ecological Genetics: The Interface* (R. H. Robichaux, ed) pp. 164–208, University of Arizona Press, Tucson, AZ.

[2] E. Pfeiler, T. A. Markow (2001) Ecology and population genetics of Sonoran Desert *Drosophila*, *Mol. Ecol*. **10**, 1787–1791.

[3] S. Pitnick, G. S. Spicer, T. Markow (1997) Phylogenetic examination of female incorporation of ejaculate in *Drosophila*, *Evolution.* **51**, 833–845.

[4] W. R. Engels, D. M. Johnson, W. B. Eggleston, J. Sved (1990) High-frequency p-element loss in *Drosophila* is homolog dependent, *Cell* **62**, 515–525.

[5] L. Jermiin, R. H. Crozier (1994) The cytochrome b region in the mitochondrial DNA of the ant *Tetraponera rufoniger*: Sequence divergence in Hymenoptera may be associated with nucleotide content, *J. Mol. Evol.* **38**, 282–294.

[6] M. Nei, S. Kumar (2000) *Molecular Evolution and Phylogenetics*, Oxford University Press, London.

APPENDIX

Handout 1: How Are These Drosophila Species Related to One Another?

The goal is to reconstruct the family relationships of these five species of *Drosophila*. One of them is the very familiar *D. melanogaster* that most people know and has been used for many years in genetics research. The other four are free-living desert *Drosophila* from southwestern North America. Which types of characteristics are most relevant for deciding their relationships? Each group should examine the flies and finish the table with morphological traits, then propose a branching tree depicting a best guess about the ancestral relationship of these species. Each group should discuss and eventually decide on one hypothesis tree to present to the class and to post.

Table II shows a example table for Handout 1.

Handout 2

Primer Sequences—CB1: 5′-TATGTACTACCATGAG GACAAATATC-3′; CB2: 5′-ATTACACCTCCTAATTTATT AGGAAT-3′.

The Invertebrate Mitochondrial Code—The first column is the three-letter codon, the second is the one-letter amino acid code, and the third is the three-letter amino acid code. This table is compiled from translation table 5 from the National Center for Biotechnology Information taxonomy database at www3.ncbi.nlm.nih.gov/Taxonomy/.

UUU	F	Phe	UCU	S	Ser	UAU	Y	Tyr	UGU	C	Cys
UUC	F	Phe	UCC	S	Ser	UAC	Y	Tyr	UGC	C	Cys
UUA	L	Leu	UCA	S	Ser	UAA	*	Ter	UGA	W	Trp
UUG	L	Leu	UCG	S	Ser	UAG	*	Ter	UGG	W	Trp
CUU	L	Leu	CCU	P	Pro	CAU	H	His	CGU	R	Arg
CUC	L	Leu	CCC	P	Pro	CAC	H	His	CGC	R	Arg
CUA	L	Leu	CCA	P	Pro	CAA	Q	Gln	CGA	R	Arg
CUG	L	Leu	CCG	P	Pro	CAG	Q	Gln	CGG	R	Arg
AUU	I	Ile	ACU	T	Thr	AAU	N	Asn	AGU	S	Ser
AUC	I	Ile	ACC	T	Thr	AAC	N	Asn	AGC	S	Ser
AUA	M	Met	ACA	T	Thr	AAA	K	Lys	AGA	S	Ser
AUG	M	Met	ACG	T	Thr	AAG	K	Lys	AGG	S	Ser
GUU	V	Val	GCU	A	Ala	GAU	D	Asp	GGU	G	Gly

GUC	V	Val	GCC	A	Ala	GAC	D	Asp	GGC	G	Gly
GUA	V	Val	GCA	A	Ala	GAA	E	Glu	GGA	G	Gly
GUG	V	Val	GCG	A	Ala	GAG	E	Glu	GGG	G	Gly

Honey Bee Cytochrome b Sequence from Amino Acid Position 142–285

```
YWGATVITNLLSAIPYIGDTIVLWIWGGFSINNAT
LNRFFSLHFILPLLILFMVILHLFALHLTGSSNPLG
SNFNNYKISFHPYFSIKDLLGFYIILFIFMFINFQFPYHL
GDPDNFKIANPMNTPTHIKPEWYFLFAYSILRA
```

Handout 3: Distance Formulae

p Distance—Number of differences (N_d) divided by the number of nucleotides or amino acids compared (N), or $p = N_d/N$.

Jukes-Cantor Distance (Nucleotide Comparisons)—The Jukes-Cantor distance (d) takes into account the random nature of substitutions (Poisson distributed) and the possibility of reversals. It corrects the above p distance: $d = -(3/4)\ln[1 - (4/3)p]$

The assumptions include:

1. The mutation rate is constant over evolutionary time.
2. The mutation rate does not vary among any of the branches.
3. The mutation rate is the same regardless of nucleotide position.
4. Any nucleotide can change to any other nucleotide with equal probability.

Poisson Distance (Amino Acid Comparisons)—The Poisson corrected distance (also symbolized by d) for proteins assumes random substitutions that are Poisson distributed: $d = -\ln(1 - p)$.

The assumptions include:

1. The mutation rate is constant over evolutionary time.
2. The mutation rate does not vary among any of the branches.
3. The mutation rate is the same regardless of amino acid position.
4. Any amino acid can change to any other amino acid with equal probability.
5. Reversals of mutations are ignored because they are infrequent.

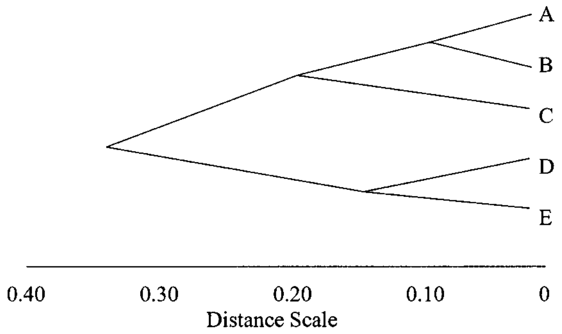
For more detailed explanations and derivations see Chapters 2 and 3 of Nei and Kumar [6].

Handout 4: UPGMA Cluster Analysis

The most simple clustering method is the "unweighted pair group method with arithmetic averages" (UPGMA). Start with a distance matrix listing the taxonomic units (species of *Drosophila*) across the top and side. Fill in one half of the matrix with the corresponding genetic distances. Here is an example matrix with letters instead of fly species.

	A	B	C	D	E
A					
B	0.20				
C	0.36	0.44			
D	0.70	.75	.60		
E	0.65	.80	.70	.30	

This matrix is scanned for the smallest element, and the two taxa are joined at an internal node drawn at a depth of 1/2 the distance back from 0 on a distance axis as shown below so that the distance along the line equals 0.08.



Thus *A* and *B* are joined at 0.20/2 = 0.10. The next group, *D* and *E*, are joined at 0.30/2 = 0.15. The next smallest distance contains a taxon that is already in a group and is joined by taking the average difference between it and all members of the group. The average distance between *A* and *B* and *C* is (0.36 + 0.44)/2 = 0.40, so take half of this and the next node joining *C* to *A* and *B* is at 0.20. Finally, to join the remaining groups we use the mean of all of the pair-wise distances between clusters (distance = 0.70 + 0.65 + 0.75 + 0.70 + 0.80 + 0.60)/6 = 0.70, so the node is drawn at 0.70/2 = 0.35. Note that all of the distances in the matrix are used only once. It can be helpful to cross out each distance as it is used.