

Chapter 6

DNA sequencing used to illustrate mutations and evolution

Theodore Gurney, Jr., Russell LeMon and Kathleen Nolan

Theodore Gurney, Jr.
Russell LeMon
University of Utah
Department of Biology
257 South 1400 East
Salt Lake City UT 84112-0840
tedgurney@bioscience.utah.edu
russlemon@hotmail.com

Kathleen Nolan
Department of Biology
St. Francis College
180 Remsen Street
Brooklyn, NY 11201
knolan@stfranciscollege.edu

Theodore Gurney, Jr. received an AB in Physics from Harvard, a PhD in Molecular Biophysics and Biochemistry from Yale, and did postdoctoral work in molecular biology at MIT and Berkeley. He has taught cell biology in lecture and lab at Utah for 26 years. His main research project concerns spontaneous rearrangements of integrated SV40 DNA in SV40-transformed mouse cells. He began a collaboration in insect PCR with Robert Elbel in 1997 on using DNA to identify fleas which can carry plague. Telephone: 801-581-8724, e-mail: tedgurney@bioscience.utah.edu, FAX: 801-581-4668

Russell LeMon received a BS in Biology May, 2000. His interests are wide, ranging from genetics and cell biology to ecology and evolution. He is especially interested in technological applications including PCR, sequencing, and computer-aided analysis of sequences. He found the Web sites MultAlin and GeneBee used in this workshop. He plans to work for a year or two before beginning graduate school. E-mail: russlemon@hotmail.com

Kathleen A. Nolan, PhD is an assistant professor in biology at St. Francis College, Brooklyn, NY. She has been teaching biology there for five years. Previous to this, she taught biology at Columbia University and Yeshiva University. She is also currently a Visiting Scientist at the American Museum of Natural History in the Molecular Systematics Laboratory. Her research interests are fisheries biology and science education. E-mail: knolan@stfranciscollege.edu

© 2001 University of Utah

Reprinted From: Gurney, T. jr., R. Le Mon, and K. Nolan. 2000. DNA sequence used to illustrate mutations and evolution. Pages 100-119, *in* Tested studies for laboratory teaching, Volume 22 (S. J. Karcher, Editor). Proceedings of the 22nd Workshop/Conference of the Association for Biology Laboratory Education (ABLE), 489 pages.

- Copyright policy: <http://www.zoo.utoronto.ca/able/volumes/copyright.htm>

Although the laboratory exercises in ABLE proceedings volumes have been tested and due consideration has been given to safety, individuals performing these exercises must assume all responsibility for risk. The Association for Biology Laboratory Education (ABLE) disclaims any liability with regards to safety in connection with the use of the exercises in its proceedings volumes.

Contents

Introduction.....	101
Materials	103
Notes to Instructors.....	104
Student Outline	110
Bibliography	119
Acknowledgments.....	119

Introduction

Sequencing for undergraduates DNA sequencing is now possible to do in undergraduate labs because of recent advances in technology. A consequence of the enormous Human Genome Project is extra sequencing capacity at many sequencing centers throughout the country. The sequencing centers are often willing to sell us a little of their spare sequencing time.

DNA sequences tell us many things about gene expression, mutations, evolution, and comparisons between species and individuals. Phylogenetic trees can be constructed from sequence data alone. Court decisions depend on sequences. We could make a case for including sequencing as a component of an ideal liberal arts education in the 21st century. At least, a well-educated person should know more about sequencing than the confusing accounts in the news media.

PCR Our lab relies on the polymerase chain reaction (PCR) which is already popular in undergraduate biology labs. To review, PCR amplifies a specific tiny fraction of the total DNA of an organism, only a part of one gene, typically 200-800 nucleotides out of many millions of nucleotides. The amplified sequence, called the *PCR product*, is specified by and bounded by two single-stranded DNA *PCR primers*, each about 25 nucleotides long, which hybridize to the cellular DNA at unique primer binding sites 100 to a few thousand nucleotides apart. PCR amplifies the sequence between the two primer binding sites exponentially because the template sequence is amplified. The ultimate amplification is typically 30 doublings or 10^9 times. PCR from a template of 50 ng of whole eukaryotic cell DNA will typically produce about 100 ng of PCR product and that is enough to see on an agarose gel, to isolate from the gel, and to sequence.

Sequence variability From the OJ trial and "Monicagate" we know that individuals can be identified by their DNA sequences because sequences have variability. By choosing different PCR primers we can produce PCR products with different degrees of sequence variability

between species or individuals.

ITS primers The ribosomal DNA sequences coding for 18S, 28S, and 5.8S ribosomal RNA are very similar in all of biological organisms; they are highly conserved. Very near the conserved ribosomal DNA sequences are the highly divergent *intervening transcribed sequences* or ITS. ITS RNA is part of the initial ribosomal RNA transcript which is discarded during ribosomal RNA processing; it codes for no essential information. The *ITS primers* are PCR primers in the conserved ribosomal sequences but the *ITS products* are PCR amplifications of the variable sequences between the primers. That means ITS primers are nearly universal in all eukaryotes but ITS products are highly variable, even between closely related species and maybe even between individuals. If we sequenced PCR products in the conserved 18S or 28S regions instead of in the variable ITS products, sequences between species would have been nearly constant instead of highly variable. ITS products from different species are often so variable that they contain DNA deletions or insertions, making clear differences in size observable using simple low resolution agarose gel electrophoresis without DNA sequencing.

COI primers We use the ITS primers and products in this lab but in addition we use some mitochondrial DNA sequences, in the COI coding region (subunit one of cytochrome oxidase) of mitochondrial DNA. COI sequences are of intermediate variability compared with ITS sequences (high variability) and mature ribosomal sequences (low variability). Unlike ITS products, the COI PCR products are all of the same size, so we must sequence them to see the differences between species. Mitochondrial DNA is inherited maternally (because mitochondria are) while the ITS sequences are nuclear and so are inherited from both parents.

Sequencing technology The automated Applied Biosystems machines at the sequencing facilities use dideoxy or Sanger sequencing in a procedure called *cycle sequencing* which is a version of PCR. Cycle sequencing uses our DNA template (our PCR product) but only one of our primers so the synthesized DNA is copied repeatedly from the original template but in only one direction. It is amplified linearly, not logarithmically, because the template is copied repeatedly to make a complementary sequence but unlike regular PCR the template itself is not amplified. The synthesis system contains a special mixture of deoxynucleoside triphosphates which are mostly the standard deoxyribonucleotide triphosphates (dGTP, dATP, dCTP, and TTP) but with added small amounts (about one part in 500) of non-standard and synthetically fluorescent 2', 3' dideoxynucleoside triphosphates (ddGTP, ddATP, ddCTP, and ddTTP). DNA synthesis terminates at random lengths when by chance it incorporates one of the ddXTP's instead of a garden variety dXTP. The ratio of each ddXTP to dXTP is adjusted by the facility to give an average terminated fragment length of 300-800 nucleotides. (Specify the PCR product size so the sequencing facility can optimize the triphosphate mixture and other conditions for the sequencing job.) The end result of cycle sequencing is a large collection of DNA fragments ranging in size from about 30 nucleotides to full length which can be over 800 nucleotides long. Each DNA fragment is colored by the ddXTP that terminated it. The sequencing facility then electrophoreses the collection of colored fragments on a polyacrylamide slab gel or on a tiny fiber matrix to separate fragments by size. Smallest fragments run fastest. The electrophoresis is good enough to resolve fragments differing by a single nucleotide in size, up to about 800 nucleotides. At the bottom of the gel or the fiber, photo multipliers read the color of the DNA fragments as they migrate past during electrophoresis. The raw sequence data are the photo multiplier output signals as the series of colored peaks. The sequence of those colored peaks gives the DNA sequence directly. We get the primary data in digital form as a graphics file of those colored peaks. We also get a computer interpretation of the primary data--a four letter

DNA Sequencing

(ACGT) text file.

Outline of our lab We start with DNA in the form of PCR products from insects which have been electrophoresed on agarose-ethidium bromide gels. (Instructions for extracting DNA from insects, making the PCR products, and doing the electrophoresis are in *Notes for Instructors* below.) We cut the PCR products out of the gels, extract the DNA from the gel slices, mix the purified PCR products with one of the two PCR primers used to make them to send them to an automated sequencing facility. The cost of sequencing is about \$15 per sequence, perhaps less if done in large quantities. In the real world the sequencing facility would send us back our sequences over the Web in about a week, but in our 3 hour workshop we have to pretend a little and work with typical data downloaded from a sequencing facility. We use some clever Web sites which facilitate sequence comparisons and make a phylogenetic tree from sequence data. We use DNA from insects here but the same techniques could be used for any species.

Materials

Horizontal slab agarose gel containing fluorescent PCR products and DNA size markers

We use 0.7% agarose containing TBE buffer and 800 ng/mL ethidium bromide.
A gel 20 cm long is run for 500 volt-hours in TBE buffer.
Overnight slow gel running is convenient, for example 20 hours at 25 volts.

Recipe for 1 liter of 10x concentrated TBE buffer

121 g Tris Base, also called Sigma 7-9
61.8 g boric acid

7.4 g tetra sodium EDTA•2H₂O

Autoclave 10x TBE to prevent a precipitate which otherwise slowly develops.

UV light box for agarose-ethidium bromide gels

Lab coats, gloves, and face shields to protect skin from UV light

Digital camera and computer or Polaroid camera and 667 film

This requires a red or orange filter which blocks UV excitation light and passes fluorescence.

Plastic knife for cutting agarose without scratching the light box. A *dull* scalpel also works well.

Forceps for picking up agarose blocks.

Waterbath set at 50°C degrees for DNA extraction (A 50 °C degree air incubator containing an end-over-end tube rotator is better.)

Microfuge tubes, 1.7 mL. You will need to weigh some of them and mark them with tare weight ± 1 mg. Use non-leakers; you don't want them to leak during shipment and some brands are terrible leakers. National Scientific brand is recommended. (Phone 1-800-459-1779.) Or order through Van Waters and Rogers, VWR #20172-698.

Microfuges, one for every 5 or 10 students.

Pipetmen: P1000, P200, P20, and P2, one each for every 10 students, plus non sterile tips for them.

Balance for weighing tubes to ± 1 mg.

BioRad Prep-a-Gene kit (stored at 4°C) for extracting DNA from agarose gels. The kit contains:

Binding buffer

50 mM Tris-HCl pH 7.4
1 mM EDTA
6 M NaClO₄ (Sodium perchlorate is sometimes 15% water; read the label.)

Wash buffer

20 mM Tris-HCl pH 7.4
2 mM EDTA
400 mM NaCl
50% (v/v) ethanol (adjust pH to 7.4 before adding ethanol)

Elution buffer

10 mM Tris-HCl pH 8.0
0.1 mM EDTA

Slurry of Prep-a-Gene silica matrix, diatomaceous earth, in water, purchased from BioRad

Staedler Lumicolor 313 permanent black fine line marker for writing on microfuge tubes. Macintosh PowerPC computers. Special software might be needed for graphics output from the sequencing facility but that can be obtained from the facility after establishing an account with them.

Notes to Instructors

Rationale This lab demonstrates some of the practical details of DNA sequencing and computer-aided sequence comparisons for undergraduate labs. Students' understanding of genetics, molecular biology, and evolution expands enormously after doing just a little DNA sequencing. Seeing those primary sequence data, those four-color peaks, is wonderfully impressive, the first time anyway. The next levels of analysis, finding open reading frames, protein coding start and stop signals, RNA splice sites, and making sequence comparisons between species or individuals are equally enlightening. We do just a little of the wet lab sequencing details and elementary data handling here but you could easily build a great semester-long lab course for Biology majors around PCR and sequencing. Elements of the lab could even be used in liberal arts courses for non-majors.

Recommended experience Students should probably have had at least one year of college biology plus lab experience with micropipettes, slab gel electrophoresis of DNA, polymerase chain reaction (PCR), graphics on PC's or Macs, and Internet use. With careful instruction (that probably means lots of good TA's) it should be possible for a beginning student or a non-major to learn all of these topics in this lab but the chances of success are much better with students who have more experience.

Continuation from last year This workshop is a sequel to three ABLE workshops offered in

DNA Sequencing

1999, [and published in *Tested Studies for Laboratory Education: Proceedings of the 21st Workshop/Conference of the Association for Biology Laboratory Education (ABLE)*, S.J. Karcher, editor]. Those workshops are in Volume 21: Chapter 5, pp. 102-112, "Introducing Students to Conservation Genetics Using Sturgeon Caviar " by K. Nolan, P. Doukakis, V. Birstein, and R. DeSalle; Chapter 1, pp. 1-32, "What I could Teach Darwin using "Darwin 2000", an Interactive Web Site for Student Research into the Evolution of Genes and Proteins" by R. Hershberger; and Chapter 3, pp. 63-79, "Introduction to the Molecular Phylogeny of Insects" by T. Gurney, R. Elbel, D. Ratnapradipa, and R. Bossard. Last year's workshops Chapters 3 and 5 started with biological material (insects or caviar) and ended with the production and visualization of PCR products as fluorescent bands on agarose slab gels. Last year's workshop Chapter 1 introduced students to computer-aided sequence analysis using sequences from available databases.

Last year we extracted DNA from different sturgeon eggs in commercial caviar, and from different species of insects caught in the field. We used the extracted DNA as templates in the PCR. The PCR worked as expected; it amplified the short segment of DNA between the two primer sites well enough to see bands on gels.

Sequencing is practical and cheap If you can see a fluorescent band of a PCR product on an agarose ethidium bromide gel, you should be able to recover it and sequence it. Processing the PCR product by electrophoresis eliminates both of the PCR primers added to the PCR reaction so after gel-purifying the PCR product we add back just one of the PCR primers to serve as a sequencing primer. We send the sequencing facility a microfuge tube containing our DNA (the PCR product) and one of the two PCR primers. The next part of the wet lab procedure, sequencing at the sequencing facility, is fortunately out of our hands.

A sequencing facility for you The University of Utah sequencing facility provides an off-campus service to nonprofit educational institutions. See <http://www.hci.utah.edu/groups/sequencing>. You will have to establish an account with them before sending material. They have a quantity discount. If you can generate 96 sequencing samples at once you can cut your sequencing costs in half by sending your DNA in a 96-well microtiter plate instead of in 96 microfuge tubes. We have not yet scaled up enough to take advantage of that discount ourselves but it illustrates the impressive capacity of modern sequencing facilities. As yet, our turnaround time for sequencing has never been longer than 4 days. You should be able to send tubes by FedEx, or maybe even by ordinary first class mail, and still get the data back over the Web within a week, in time for the next week's lab class.

The cost of sequencing is currently \$15 per sequence for an off-campus nonprofit educational institution. There are probably cheaper facilities we do not yet know about, but in any case sequencing cost is not a major expense in our lab course. We usually do one sequence for every 2 students in the class and we make sure that each sequence is different to get the most information from our sequencing dollars. Then all of the students share all of the sequencing data.

Primer sequences Use primers in the pairs given below. The first primer pair amplifies ribosomal ITS sequences (intervening transcribed sequences) and the next two amplify a tandemly arranged pair of sequences in the COI region (subunit one of cytochrome oxidase) of the mitochondrial DNA. The nucleotide order of the sequences is 5' to 3'.

ITS 3: GCATCGATGAAGAACGCAGC

ITS 4: TCCTCCGCTTATTGATATGC

mtD-6: GGAGGATTTGGAAATTGATTAGTTCC

mtD-9: CCCGGTAAAATTAATAAACTTC

mtD-10: TTGATTTTTTGGTCATCCAGAAGT

mtD-12: TCCAATGCACTAATCTGCCATATTA

You can get small amounts of the primers from the authors at no cost or you can order larger amounts through the sequencing facility. The cost will be about \$40 per batch of one primer. One batch would be a lifetime supply for our own uses but not for sharing with the whole world. There are probably cheaper primer-making services we do not know about.

DNA extraction from insects We have updated our procedure from last year to keep the insect exoskeleton intact for good morphological identification in the microscope. With this modification you can recover an intact exoskeleton from an insect in one tube and its DNA in another.

1. Put a live insect into a large excess of 95% ethanol or 70% isopropanol overnight or longer. For example, use 0.5 mL for a fruit fly or 1 mL for a mosquito. The insect in alcohol is stable indefinitely at room temperature, for the purposes of DNA sequencing.
2. Dry the fixed insect for 30 min at room temperature.
3. Soak the dried insect in 0.5 M KOH plus 0.1% Joy detergent for 1-7 days at room temperature. The volume of KOH-Joy depends on the size of the insect.

40 μ L for a flea, e. g. *C. felis*

80 μ L for a fruit fly, *D. melanogaster*

160 μ L for a mosquito, *A. dorsalis*

Don't try to process whole larger insects such as bees, wasps, big flies, beetles, grasshoppers, or dragonflies but you could do a flea-sized piece of a leg or a wing.

4. Neutralize KOH by adding 37.5% of the initial volume of 1 N HCl and 12.5% volume of 1 M Tris-HCl, pH 7.2. (For example, we neutralize 40 μ L KOH-Joy with 15 μ L HCl and 5 μ L Tris.)
5. Vortex hard with an equal volume (after the additions above) of water-saturated phenol:chloroform, 1:1. Centrifuge briefly.
6. Recover the aqueous (upper) phase into a clean microfuge tube.
7. Add sodium acetate to a 50 mM final concentration.
8. Add glycogen (precipitation carrier) to 50 ng/ μ L final concentration. See the ordering information below. We use glycogen with tiny insects such as fleas but we can skip it with larger insects such as mosquitoes; apparently KOH extracts enough endogenous

DNA Sequencing

precipitation carriers.

9. Add 95% ethanol to 67% final concentration; place in -20° freezer for >2 hours.
10. Microfuge (12,000 rpm) at room temperature for 1 min, decant liquid and drain 1 min.
11. Add 500 μL 70% ethanol, vortex, spin in microfuge 30 sec, decant liquid and dry inverted tube 30-60 min, until bone dry.
12. Re-dissolve overnight at 4°C in 10 mM Tris-HCl pH 8, 0.1 mM EDTA. Volume depends on size:
 - 10 μL for a flea
 - 20 μL for a fruit fly
 - 40 μL for a mosquito
13. Store DNA at 4°C or frozen at -20°C . Be sure to dissolve the DNA before freezing it. Be sure to vortex thawed DNA before using it.

Assembling and running the PCR

1. Put 4 μL of insect DNA prepared as above in a PCR tube and chill the sample on ice. Prepare at least five different insects to run together. Try to do at least one insect for each student. Small volumes minimize costs.
2. Add 6 μL of the following polymerase mixture to the 4 μL of DNA. Keep cold, on ice. This recipe is for just 6 μL of mixture for one 10 μL PCR but make at least five PCR's worth of the mixture because it is not possible to pipet 0.05 μL accurately, even with a P2 Pipetman. Your recipe might be different from this one if your 10x PCR buffer contains magnesium or if your primers are at different concentrations from ours.

1.95 μL H_2O

1 μL of 10x PCR buffer minus magnesium

1 μL of 20 mM MgSO_4 or MgCl_2

1 μL of triphosphate mixture, 2 mM of each dXTP

0.5 μL of each of the 2 primers (see sequences above) at 10 μM each

0.05 μL Taq polymerase (or equivalent) at 5 u/ μL . Use a P2 Pipetman to measure this small volume.

3. Vortex each PCR reaction and centrifuge the tube to get all the liquid to the bottom of the tube. Keep the sample cold.
4. If your PCR machine has no heated lid, add 12 μL mineral oil to each tube and to each well in the heating block. Microfuge the tubes 10 sec after adding oil. Adding cold oil is tricky, so practice! The oil should cover the PCR but to a depth of only 1-2 mm; more oil causes problems.

5. Put the tubes into the PCR machine.
6. Start PCR program mt-1; it takes 1 h 15 min in an MJ Research PCR machine.

PCR program mt-1 for MJ Research PCR machines

1. 94°C for 1 min
 2. 94°C for 10 sec
 3. 40°C for 10 sec
 4. 72°C for 50 sec
 5. Go to step 2 -- 29 times
 6. 72°C for 1 min
 7. 30°C for 10 sec
 8. End
7. At end of PCR add 2 μ L PCR stop mixture: 50 mM Tris-HCl, pH 8, 1 mM EDTA, 30% glycerol, 0.1% bromphenol blue. The stopped PCR's are stable in the refrigerator.
 8. If you have had to use oil, de-grease your PCR by rolling the sample down a fold in Parafilm.
 9. Vortex the sample and load everything, 12 μ L, into one agarose gel slot.

The mitochondrial PCR program mt-1 works for both mitochondrial and ITS primers. It is designed for 10 microliter reactions.

The program mt-1 uses a low annealing temperature (step 3, 40 degrees) because of the low GC content of mitochondrial primers and possible mis-matches between primers and templates. The low temperature sometimes leads to spurious PCR products in addition to the expected ones, due to false priming. You can try experimenting with a higher annealing temperature, such as 50 or 55 degrees, if you get multiple products. If you cut out the expected PCR products from the gel before sequencing, however, the spurious products should not matter unless you are very unlucky and get a spurious product of exactly the same size as the expected one. The approximate expected sizes of PCR products for insects are as follows:

ITS 3&4: 500 nucleotides
 mtD6&9: 500 nucleotides
 mtD10&12: 800 nucleotides

Agarose gel electrophoresis of DNA This recipe is for a flat gel bed 0.7% gel of 12.5 x 20 cm. If your gel is different, scale the recipe up or down proportionately to the area of your gel bed. Level the gel bed and put in end pieces to hold the liquid agarose. Insert the slot-forming comb. Small comb teeth, \approx 1.5 x 4 mm each, are adequate for 10 μ L of PCR plus 2 μ L stop buffer.

1. Mix 1.05 g agarose (special agarose for DNA electrophoresis) with 135 mL deionized water in a 500 mL Erlenmeyer flask.

DNA Sequencing

2. Boil in microwave oven, full power for 1.5 min.
3. Simmer at 30% power until all agarose is melted, about 2 min.
4. Add 15 mL 10x TBE buffer (see recipe above).
5. Add 13 μ L 10 mg/mL ethidium bromide in formamide. (Ethidium bromide is a mutagen, danger.)
6. Mix and cool to ~40 degrees.
7. Pour the gel; remove bubbles. The gel should set within 30 min; then remove the end pieces.
8. Fill the electrode reservoirs and cover the gel to a depth of 3 mm with 1X TBE buffer.
9. Do not pull the comb out until just before loading the gel.
10. Set the voltage before loading and turn it off until the gel is loaded. Plan on 500 volt-hours, for a gel 20 cm long. If your gel is shorter than 20 cm use proportionally fewer volt-hours.
11. The positive electrode should be at the far end of the gel, away from the slots.
12. Load the samples together with size markers such as Gibco/BRL "1 kb" markers. If you have to remove oil from the samples (no heated lid on PCR machine), do it at the time of gel loading.
13. You may use an appliance timer to turn the power on after all students have left the lab (This is safer.) and then off again after 500 volt-hours. (This avoids rushing in to the lab to turn it off.)

Slow overnight (12-20 hour) gel running makes somewhat broader bands than fast (1-3 hour) gel running because there is more time for diffusion of small PCR products but the slow gel has its advantages. The bands are always at reliable distances from the slots because there is no distortion due to heating of the gel so PCR product size estimates are more accurate. Overnight runs are much easier to schedule; the gel might as well be running while you are sleeping. If you teach the same class on successive days you can have the students in one class prepare a gel for the next day's class.

Photographing the gel If you have a commercial DNA gel photographing system, that is wonderful, but you do not have to spend that kind of money. The cheaper Polaroid version (Fotodyne 5-5344) works very well with Polaroid 667 film. If you have your own Polaroid, or 35 mm camera, or digital camera, you will need a red-orange filter to block the excitation wavelength of the UV light box and pass the red-orange fluorescence. The Epson Photo-PC 650 digital camera works well in this application. If you use a Polaroid camera to photograph the gel, you can use a digital camera to make copies of the Polaroid print.

Prep-a-Gene Vs Qiagen We use the BioRad Prep-a-Gene slurry system instead of Qiagen columns for extracting DNA from gel slices because it is cheaper and we get more concentrated DNA. It does take more work but we always have lots of eager young hands so the work goes quickly and the cooperative experience actually builds morale. We can elute DNA from the Prep-a-Gene matrix in only 20 μ L; that detail alone argues strongly for Prep-a-Gene.

The catalog number for the complete Prep-a-Gene kit is BioRad 732-6010, \$103.00 for 100 purifications. To be more economical, purchase only the matrix, DNA gel purification matrix, BioRad 732-6012, \$47.00 for 2 mL which should be enough for over 200 purifications. Make up the rest of the kit reagents yourself.

Rotators If you plan on using Prep-a-Gene a lot, you might consider buying two end-over-end microfuge tube rotators or some other kind of mechanical microfuge tube agitator. The rotator helps to dissolve agarose, bind DNA to matrix, and elute DNA from matrix by keeping solids suspended. You need one rotator in a 50°C incubator for the two 50°C steps (melting agarose and eluting DNA from matrix) and another rotator for the room temperature DNA binding step. Using rotators makes the process easier and more reliable than using stationary waterbaths.

Precipitation Sequencing usually works well with DNA directly eluted from diatoms without precipitation if students are careful to remove all dregs of the washing buffer and if they have enough material. But accidents happen and there is a way to salvage the work if you lose most of the material or if the salt concentration is too high due to careless washing. Try precipitating and re-dissolving the eluted DNA to increase the DNA concentration and to reduce the salt concentration. To the 20 µL of eluted DNA prepared as described, add 2 µL of 1 mg/mL glycogen, 2 µL of 3 M sodium acetate, pH 7, and 60 µL of 95% ethanol. Put the tube at -20° for at least one hour and then microfuge it at 12,000 rpm at room temperature for 1 minute. Decant the supernatant as soon as the rotor stops. Add 500 µL of 70% ethanol to rinse the pellet. Don't even try to look for a pellet; you will not see one unless you have carried over some diatoms. Microfuge the DNA again for 30 seconds. Decant again as soon as the rotor stops and keep the tube upside down at room temperature until it is dry, 30-60 minutes. Re-dissolve the DNA in 8 µL of elution buffer to give the highest possible DNA concentration. Precipitation and re-dissolving has a yield of ~80%.

The glycogen serves as a carrier for precipitation; it helps to precipitate tiny quantities of DNA without adding more nucleic acids to the tube. Glycogen from mussels, Roche Molecular Biochemicals, cat. no. 901-393, 20 mg in 1 mL water, \$93.00, is a lifetime supply.

In our experience, DNA recovered from a bright gel band and eluted from the Prep-a-Gene diatom matrix in 20 µL should give good sequencing performance. If the fluorescent PCR product is easily visible (through protective glasses) you should be able to sequence it.

Student Outline

Protective clothing Before turning on the UV light box put on protective clothing and use eye protection. Cover all exposed skin with a lab coat and gloves. Be sure to wear protective glasses or a face shield. The potentially worst UV damage would be to your eyes. Ultraviolet light is the most dangerous thing in this lab.

Photographing the gel We start the lab by observing fluorescent DNA PCR products on agarose slab gels. The PCR products were made from insect DNA templates using ribosomal ITS primers or mitochondrial mtD primers described in the Introduction. The gels were run slowly overnight for 500 volt-hours. You should see a ladder of size markers in one lane and then PCR products as single red fluorescent bands in each of the other lanes. We will photograph the gel and make copies. It is always safer to examine the photograph than to examine the fluorescent gel itself.

Cutting out the bands Wearing all of your protective clothing and eye protection, use a dull scalpel (to avoid scratching the light box) to cut out a band of PCR products. Cut the gel on the four sides of the fluorescent band as close as possible to the band without losing band material. Do all of the cutting on a particular gel before trying to pick up the agarose chunks; then turn the UV off. With the UV off pick up the cut-out chunks of agarose with a forceps and put each in a

DNA Sequencing

pre-weighed microfuge tube. Turn the UV back on to see if you missed any DNA. Maybe photograph the gel again. Weight the tubes again and compute the net weights of the agarose in the tube.

Melting the agarose Add binding buffer to the agarose in the microfuge tube. The volume in microliters to add is 3.5 times the weight of the gel slice in milligrams; for instance, for a 100 mg add slice gets 350 μL of binding buffer. Place the tube in a 50°C waterbath, vortex it every 5 minutes, and put it back in the waterbath. Do this for a total of 15 minutes of heating but don't count the time that the tube was out of the waterbath. An end-over-end tube rotator inside a 50°C incubator is a more efficient way to do this step.

Binding DNA to diatoms Suspend the BioRad diatomaceous earth matrix by vortexing it and then add 6 μL of suspended matrix to each tube of melted agarose in binding buffer. (BioRad has a patent on diatoms for DNA use!) Vortex the tube every 1 minute for 15 minutes at room temperature. In high salt binding buffer DNA, but not RNA, protein, carbohydrates, or lipids, binds selectively to silica which is what diatom skeletons are made of. DNA also binds to silica in the form of tiny glass beads, as in Qiagen columns. The mechanism of selective DNA binding to silica is still debated but it works wonderfully and has made some people very rich.

Washing the diatoms This step is best done by a team of students in an assembly line, with a division of labor. Microfuge all of tubes at once for 15 seconds, at 12,000 rpm, at room temperature. Decant and discard the supernatant from each tube but then put the tube right back in the microfuge. When all of the tubes are back in, spin again for 15 seconds to get the traces of liquid in the bottom of the tube; then remove the traces of liquid with a P200 Pipetman. If the microfuge tube hinge is positioned towards the outside of a microfuge angle head rotor, you can safely place a P200 Pipetman tip all the way to the bottom of the tube on the side opposite the hinge without disturbing the pellet. (This trick works well with Eppendorf and similar angle head microfuges but does not work with old Beckman Model B horizontal microfuges.) When removing supernatants do not bother to change Pipetman tips between tubes; we never do and we do not have problems. If you do change tips it takes forever to take off the supernatants. But we do change tips between the different rinses and we always use a clean tip to go into the binding buffer and wash buffer, to avoid possible contamination of our reagents. After removing all of the liquid from a tube with a Pipetman, add 100 μL of fresh binding buffer using a clean tip. Vortex hard for 2 seconds to suspend the diatoms completely and put the tubes back in the microfuge, hinge side out. Microfuge for 15 seconds again. Remove the supernatants with a P200 as above. You have now washed the diatoms once with binding buffer. Next you want to wash the diatoms twice with the washing buffer. There is lots of spinning and pipetting to do but if you get organized the work will go fast. While some students are washing the diatoms, other students could be marking tubes to contain the DNA eluted in the next step. After the last wash be very careful to remove all traces of the washing buffer, to reduce the salt concentration as much as possible in preparation for elution in low salt.

Eluting DNA from diatoms Add 20 μL of elution buffer to each tube of washed diatoms. Use a P200 in the "to deliver" mode of operation without touching anything in the tube so that you can add elution buffer to all of the tubes with a single Pipetman tip. Vortex slowly to suspend the diatoms without spattering. Incubate at 50°C for 10 minutes with occasional vortexing. The 10 minutes should not include time out of the 50°C waterbath. Store the eluted DNA in the tube with the pellet in the refrigerator (+4°C) until you are ready to do the next step.

Preparing DNA for sequencing These instructions are current (Fall, 2000) for the University of Utah's sequencing facility. Check their Web site: <http://www.hci.utah.edu/groups/sequencing> for updates. Turn a 1.7 mL microfuge tube on its side to use the rectangular frosted patch on the side of the tube for labeling. The lab's name you use has to be the same as the one your

DNA Sequencing

instructor uses to write the sequencing order on the Web to the sequencing facility, so get it right. Write the lab's name on the top of the patch and the DNA template name underneath the lab's name. Label the same tube on its top with the primer name. Keep all names short and use all caps. Using a P2 Pipetman, spot 0.3 μL of 10 μM primer on the inside wall of the tube. (Pipetting 0.3 μL takes talent so practice first.) Then add 6.7 μL of DNA eluted from the diatoms. Cap the tube tightly and ship according to instructions from the Web site.

Downloading data You will be notified by e-mail when your data are ready. The lab should have established an account which will give you reading privileges; you will need a username and a password to download. Each sequence will come in two forms, the primary data as a graphics file of about 230 K, which can be opened with EditView software, and a text file of about 20 K, which can be opened with many word processing programs.

Interpreting data Check the graphics first to see if the data are reliable. You should get clear reads of sequence with just AGCT and very few if any N's. The N's are signs that their computer could not interpret your sequence. If you have more than 10% N's the sequence is probably not reliable. Remember that the text file is an interpretation of the graphics file, so if there are ambiguities, consult the graphics.

Using BLAST2 This National Library of Medicine, National Institutes of Health Web site (<http://www.ncbi.nlm.nih.gov/gorf/bl2.html>) is useful for comparing two sequences to find differences between them. You will have to remove some of the sequencing facility's formatting. You want just 2 plain text-only files for this Web site. Each sequence is a single string of capital letters with no spaces, no returns, and no punctuation. Here is an example of two files prepared for BLAST2; the first is the *Aedes dorsalis* (mosquito) mtD6&9 product and the second is the *Drosophila melanogaster* (fruit fly) mtD6&9 product. Both were sequenced with mtD6.

```
TAGGAGCTCCAGATATAGCATTTCCTCGAATAAATAATATAAGTTTTTGAATACTACCTCCTTCA
TTAACTACTGCTACTTTCAAGTAGTATAGTAGAAAATGGATCAGGAACAGGATGAACAGTTTTATCC
ACCTCTTTCATCTGGAAGTCTCATGCAGGAGCTTCAGTTGATTTAACAATTTTTTCTTTACATT
TAGCNTGGAGTATCATCAATTTTAGGAGCAGTAAATTTTATTACTACTTGTAATTAATATACGAT
CAGCAGGAATTACTTTAGATCGATTACCTTTATTTGTTTGATCTGTTGTAATTACAGCAGTATTA
TTACTTTTATCATTAACCTGTTTTAGCCGGAGCTATTAATAATTAATACTGATCGAAATTTAAA
TACTTCATTCTTTGATCCTATTGGAGGAGGAGACCCTATTTTATATCAACATTTATTTTGATTCT
TTGGACATCCTGAAGTTATTTATTT
```

```
TAGGTGCTCCTGATATAGCATTCCCACGAATAAATAATATAAGATTTTGAATACTACCTCCTGCT
CTTTCTTTACTATTTAGTAAGTAGAATAGTTGAAAATGGAGCTGGAACAGGATGAACGTTTTATCC
ACCTTTATCCGCTGGAATTTGCTCATGGTGGAGCTTCAGTTGATTTAGCTATTTTTTCTCTACATT
TAGCAGGGATTTCTTCAATTTTAGGAGCTGTAAATTTTATTACAACGTAATTAATATACGATCA
ACAGGAATTTCATTAGATCGTATACCTTTATTTGTTTGATCAGTAGTTATTACTGCTTTATTATT
ATTATTATCACTCCAGTACTAGCAGGAGCTATTAATAATTAATAACAGATCGAAATTTAAATA
CATCATTTTTTGACCCAGCGGGAGGAGATCCTATTTTATATCAACATTTATTTTGATTTTTT
GGTCACCCTGAAGTTATTTATTT
```

Open the Web site. You will find two boxes. Cut and paste one sequence into one box and the other into the other box. Click "start". The analysis should take only about a minute. Download the output file onto your hard disk or floppy. Quit. That is it.

Using other BLAST sites Explore all of the BLAST sites (<http://www.ncbi.nlm.nih.gov/BLAST/>) to see what they offer. The BLASTN site can be used with an unknown sequence to scan a large database to identify an unknown. An extensive database search with BLASTN will take longer than the pair-wise comparisons with BLAST2,

DNA Sequencing

probably several minutes. If you are using the mitochondrial primers, select the mitochondrial database instead of the general database to save searching time. Use the FASTA format for a submitted sequence for searching; that is just like the format for MultAlin given below.

Two searching examples: 1) We were having a hard time matching an insect ITS sequence with other insects but BLAST told us why: Our sequence was fungal, not insect, probably from fungus growing in our insect. 2) A mystery insect flying around potted plants turned out to be a *Sciara* gnat, identified by ITS sequence.

Using MultAlin This Web site at the French National Agricultural Institute at Toulouse (<http://www.toulouse.inra.fr/multalin.html>) compares several sequences at once. It orders the sequences by placing the most similar sequences next to each other and then it makes a consensus sequence. The preparation of data for MultAlin is different from BLAST2; you paste in just one file with all of the sequences in it and you must name each sequence in FASTA format. Here are the same two sequences as above prepared in FASTA format for MultAlin. Notice the > at the front of the name line and notice the single blank line between the sequences.

```
>AD-MTD6 . seq
TAGGAGCTCCAGATATAGCATTTCCTCGAATAAATAATATAAGTTTTTTGAATACTACCTCCTTCA
TTAACACTGCTACTTTCAAGTAGTATAGTAGAAAATGGATCAGGAACAGGATGAACAGTTTTATCC
ACCTCTTTCATCTGGAAGCTGCTCATGCAGGAGCTTCAGTTGATTTAACAAATTTTTTCTTTACATT
TAGCNTGGAGTATCATCAATTTTAGGAGCAGTAAATTTTATTACTACTTGTAATTAATATACGAT
CAGCAGGAATTACTTTAGATCGATTACCTTTATTTGTTTGATCTGTTGTAATTACAGCAGTATTA
TTACTTTTATCATTACCTGTTTTAGCCGGAGCTATTACTATATATTAACTGATCGAAATTTAAA
TACTTCATTCTTTGATCCTATTGGAGGAGGAGACCCTATTTTATATCAACATTTATTTTGATTCT
TTGGACATCCTGAAGTTATTTATTT
```

```
>DM-MTD6 . seq
TAGGTGCTCCTGATATAGCATTCCCACGAATAAATAATATAAGATTTTTGACTACTACCTCCTGCT
CTTCTTTACTATTTAGTAAGTAGAATAGTTGAAAATGGAGCTGGAACAGGATGAACGTTTTATCC
ACCTTTATCCGCTGGAATTGCTCATGGTGGAGCTTCAGTTGATTTAGCTATTTTTTCTCTACATT
TAGCAGGGATTTCTTCAATTTTAGGAGCTGTAATTTTATTACAACGTAATTAATATACGATCA
ACAGGAATTTCATTTAGATCGTATACCTTTATTTGTTTGATCAGTAGTTATTACTGCTTTATTATT
ATTATTATCACTTCCAGTACTAGCAGGAGCTATTACTATATTATTAACAGATCGAAATTTAAATA
CATCATTTTTTTGACCCAGCGGGAGGAGATCCTATTTTATATCAACATTTATTTTGATTTTTT
GGTCACCCCTGAAGTTATTTATTT
```

The MultAlin output can be several forms; we use their text (.mtf), their graphics (.gif) output, and their web (.htm) output.

Here is MultAlin text output for nucleotides 101-150, out of about 500, for the ITS 3&4 sequences of 24 different fleas. We pasted in all 24 sequences in the one box on MultAlin. The PCR products were made with ITS 3&4 and sequenced with ITS 3.

To make the output look reasonable, we changed the font to a constant-spacing font, Courier, from the proportional spacing font, Times. MultAlin added all the names on left and repeated them on every page of 50 nucleotides, and it added the numbers of the nucleotides at the top, the space every 10 nucleotides and the consensus sequence at the bottom. The consensus sequence gives conservation information: Upper case in the consensus line means completely conserved, lower case means mostly conserved, and dot means weakly conserved.

	101	150
OM.seq	ACCAGACTGC TACTTGCC.. . . .TCCG..	GGCTCGTT.A GCGAACGATG
3433.seq	ACCAGACTGC TACTTGCC.. . . .TTCG..	GGCTCGTT.A GCGAACGATG
OT.seq	ACCAGACTGC TACTTGCC.. . . .CTCG..	GGCTCGTT.A GCGAACGATG
5041.seq	ACCAGACTGC TGCTTGCC.. . . .TTCG..	GGCTCGTA.A GCGAACGATG
PC5.seq	ACCAGACTGC TGCTTGCC.. . . .TTCG..	GGCTCGTG.A GCGAACGATG
299.seq	ACCAGACTGC TACTTGCT.. . . .TTCG..	GGCTCGTT.A GCGTACGATG
319F.seq	ACCAGACTGC TGCTTGCC.. . . .TTCG..	GGCTCGTT.A GCGAACGATG
SD.seq	ACCAGACTGC CACTTGCC.. . . .TTCG..	GGCTTGTC.G GCGAACAGTG
MG.seq	ACCAGACTGC TACTTGCC.. . . .ATCG..	AGCATGTT.A GCGAACGATG
345B.seq	ACCAGACTGC TACTTGCC.. . . .TTCG..	GGCTCGTT.A GCGTAAGATG
310.seq	ACCAGACTGC TACTTGCC.. . . .TTCG..	GGCTCGTT.A GCGTAAGATG
304C.seq	ACCAGACTGC TACTTGCC.. . . .TTCG..	GGCTCGTT.A GCGTAAGATG
331A.seq	ACCAGACTGC TACTTGCC.. . . .TTCG..	GGCTCGTT.A GCGTAAGATG
345A.seq	ACCAGACTGT CGCTTGCT.. . . .TTCG..	GGCTTGCC.G ACGAATGTTG
3432.seq	ACCAGACTGC TACTTGCC.. . . .TTAC..	GGCTCGCT.A GCGAATATTG
3434.seq	ACCAGACTGC TACTTGCC.. . . .TTAC..	GGCTCGCT.A GCGAATATTG
EW.seq	ACCAGACTGC TACTTGCC.. . . .TTAC..	GGCTCGCT.A GCTAATATTG
297.seq	ACCAGACTGC TGCTTGCT.. . . .CTCG..	GGCTCGCT.A GCGAATGATG
DF.seq	ACCAGACTGC TGCTTGCT.. . . .TTCG..	GGCTTGCT.A GCGCACGGTG
ST.seq	ATCAGACTGC CGTTTGTCGC CTTGTGCG..	GTCATTCT.G GCGAATGATG
HAA.seq	ACCAGACTGC CGCTTGCCGC TTCATGTG..	GACTTGCTCG GCGAATGTTG
PS.seq	ATCAGACTGC CGCTTGCTCG CTTCTACGCT	TGCACGC.CG GCGTATGATG
CF.seq	ACCAGACTGC CGCTTGCT.. .TTAAACG..	.GCTTGC.CG GCGAATGATG
XC.seq	ACAAGACTGT CGCTTGCC.. . .CTCCG..	.GCTCG.TCG ACGAATGATG
Consensus	AcCAGACTGc tgCTTGCCc.. . . .ttcg..	ggCtcG.t.a gCGaAtgaTG

Only two sequences turned out to be exactly the same. Notice the deletions/insertions which are typical of ITS sequences but not mitochondrial sequences.

The MultAlin graphics file output is in gif format which is readable on Macintoshes with current versions of PictureReader in the Apple QuickTime package that comes with new iMacs and other Apple computers. There are equivalent PC programs that can open gif files. The html output is readable with Netscape or Internet Explorer. MultAlin color-codes the data to indicate conservation; red is completely conserved, blue is mostly conserved, and gray is weakly conserved. In comparisons of ITS sequences, the color-coding is dramatic because sequences start and end in conserved regions of 5.8S and 28S coding DNA but the middle parts are highly variable ITS. Then there are surprising conserved islands in the middles of the variable regions.

Unlike ITS sequences, the mitochondrial sequences code for a subunit of cytochrome oxidase, a gene product essential for (aerobic) life. (Remember that the mitochondrial genetic code is different from the standard genetic code!) There are no insertions/deletions because of the essential function but there are single nucleotide substitutions. In comparisons of mitochondrial sequences which are all of the same lengths, the color coding highlights mutations we might miss otherwise. You could find which reading frame is the protein coding frame and if the mutations are third base substitutions resulting in no amino acid change, or if they do make

DNA Sequencing

amino acid changes, and if so if the changes are conservative or radical. This is just a sample of the kinds of analysis you can do with these tools.

Using GeneBee This site, at Moscow State University in Russia, makes simple distance-based phylogenetic trees from a collection of sequences that you submit. The URL is http://www.genebee.msu.su/services/phtree_reduced.html. The instructions are pretty simple and, as teachers sometimes say, we leave most of the details as exercises for the student.

You should re-format data for GeneBee. As a formatting example, here below is the same piece of that 24 x 500 flea file we received as MultAlin text output above, but we reformatted it for GeneBee. Notice that 1) The names column has its margin on the left instead of the right, 2) We shortened the names, 3) There are 4 spaces between the longest name and the data, 4) The space between every 10 nucleotides is gone, 5) The dots are all changed to dashes, 6) The nucleotide numbers are gone, and 7) The consensus line is gone. A colleague, Eric Kofoid, has just written a little program that does the trick; it can be obtained from us by e-mail.

```
OM      ACCAGACTGCTACTTGCC-----TCCG--GGCTCGTT-AGCGAACGATG
3433    ACCAGACTGCTACTTGCC-----TTCG--GGCTCGTT-AGCGAACGATG
OT      ACCAGACTGCTACTTGCC-----CTCG--GGCTCGTT-AGCGAACGATG
5041    ACCAGACTGCTGCTTGCC-----TTCG--GGCTCGTA-AGCGAACGATG
PC5     ACCAGACTGCTGCTTGCC-----TTCG--GGCTCGTG-AGCGAACGATG
299     ACCAGACTGCTACTTGCT-----TTCG--GGCTCGTT-AGCGTACGATG
319F    ACCAGACTGCTGCTTGCC-----TTCG--GGCTCGTT-AGCGAACGATG
SD      ACCAGACTGCCACTTGCC-----TTCG--GGCTTGTC-GGCGAACAGTG
MG      ACCAGACTGCTACTTGCC-----ATCG--AGCATGTT-AGCGAACGATG
345B    ACCAGACTGCTACTTGCC-----TTCG--GGCTCGTT-AGCGTAAGATG
310     ACCAGACTGCTACTTGCC-----TTCG--GGCTCGTT-AGCGTAAGATG
304C    ACCAGACTGCTACTTGCC-----TTCG--GGCTCGTT-AGCGTAAGATG
331A    ACCAGACTGCTACTTGCC-----TTCG--GGCTCGTT-AGCGTAAGATG
345A    ACCAGACTGTCGCTTGCT-----TTCG--GGCTTGCC-GACGAATGTTG
3432    ACCAGACTGCTACTTGCC-----TTAC--GGCTCGCT-AGCGAATATTG
3434    ACCAGACTGCTACTTGCC-----TTAC--GGCTCGCT-AGCGAATATTG
EW      ACCAGACTGCTACTTGCC-----TTAC--GGCTCGCT-AGCTAATATTG
297     ACCAGACTGCTGCTTGCT-----CTCG--GGCTCGCT-AGCGAATGATG
DF      ACCAGACTGCTGCTTGCT-----TTCG--GGCTTGCT-AGCGCACGGTG
ST      ATCAGACTGCCGTTTGTGCCTTGTCG--GTCATTCT-GGCGAATGATG
```

DNA Sequencing

```

HAA   ACCAGACTGCCGCTTGCCGCTTCATGTG--GACTTGCTCGGCGAATGTTG
      PS   ATCAGACTGCCGCTTGCTCGCTTCTACGCTTGACACGC-CGGCGTATGATG
      CF   ACCAGACTGCCGCTTGCT---TTAAACG---GCTTGC-CGGCGAATGATG
      XC   ACAAGACTGTGCTTGCC-----CTCCG---GCTCG-TCGACGAATGATG
  
```

The file shown above is only about 10% of the total file we submitted to GeneBee, just nucleotides 101 to 150 out of 500. As with the other programs, we just paste the re-formatted file into a box, but with GeneBee we also have to give a return e-mail address. Then just we push the start button. In our experience so far, the data have come back to our e-mail address within a few minutes. Below are GeneBee results for the complete 24 x 500 matrix of the 24 ITS 3&4 sequences aligned by MultAlin. GeneBee also gives a distance matrix based on differences between the sequences.

```

DISTANCE MATRIX FROM GENESEE
      1      2      3      4      5      6      7      8      9     10     11     12
1 OM   0.000 0.170 0.170 0.207 0.201 0.215 0.220 0.252 0.263 0.339 0.339 0.317
2 3433 0.170 0.000 0.034 0.127 0.118 0.133 0.129 0.172 0.199 0.278 0.278 0.266
3 OT   0.170 0.034 0.000 0.139 0.130 0.145 0.148 0.178 0.199 0.276 0.276 0.258
4 5041 0.207 0.127 0.139 0.000 0.024 0.112 0.172 0.176 0.196 0.269 0.269 0.254
5 PC5  0.201 0.118 0.130 0.024 0.000 0.112 0.163 0.164 0.182 0.254 0.254 0.240
6 299  0.215 0.133 0.145 0.112 0.112 0.000 0.195 0.199 0.219 0.240 0.240 0.222
7 319F 0.220 0.129 0.148 0.172 0.163 0.195 0.000 0.205 0.227 0.308 0.308 0.308
8 SD   0.252 0.172 0.178 0.176 0.164 0.199 0.205 0.000 0.105 0.319 0.319 0.305
9 MG   0.263 0.199 0.199 0.196 0.182 0.219 0.227 0.105 0.000 0.315 0.315 0.306
10 345B 0.339 0.278 0.276 0.269 0.254 0.240 0.308 0.319 0.315 0.000 0.000 0.073
11 310  0.339 0.278 0.276 0.269 0.254 0.240 0.308 0.319 0.315 0.000 0.000 0.073
12 304C 0.317 0.266 0.258 0.254 0.240 0.222 0.308 0.305 0.306 0.073 0.073 0.000
13 331A 0.311 0.261 0.253 0.249 0.234 0.217 0.302 0.299 0.301 0.067 0.067 0.012
14 345A 0.408 0.370 0.368 0.343 0.337 0.378 0.376 0.327 0.364 0.426 0.426 0.417
15 3432 0.350 0.318 0.316 0.291 0.291 0.329 0.341 0.271 0.300 0.389 0.389 0.380
16 3434 0.348 0.315 0.313 0.288 0.288 0.326 0.339 0.268 0.297 0.386 0.386 0.377
17 EW  0.378 0.340 0.341 0.322 0.319 0.357 0.363 0.302 0.328 0.410 0.410 0.402
18 297  0.360 0.321 0.313 0.282 0.282 0.315 0.324 0.289 0.312 0.383 0.383 0.378
19 DF  0.363 0.318 0.328 0.297 0.297 0.321 0.340 0.283 0.306 0.410 0.410 0.393
20 ST  0.550 0.518 0.530 0.512 0.518 0.526 0.517 0.523 0.523 0.527 0.527 0.530
21 HAA 0.594 0.571 0.570 0.576 0.571 0.573 0.575 0.566 0.579 0.580 0.580 0.575
22 PS  0.541 0.506 0.518 0.517 0.511 0.513 0.502 0.503 0.525 0.515 0.515 0.520
23 CF  0.500 0.450 0.465 0.458 0.450 0.453 0.458 0.425 0.447 0.496 0.496 0.493
24 XC  0.590 0.559 0.571 0.556 0.559 0.555 0.545 0.569 0.574 0.567 0.567 0.585
  
```

DNA Sequencing

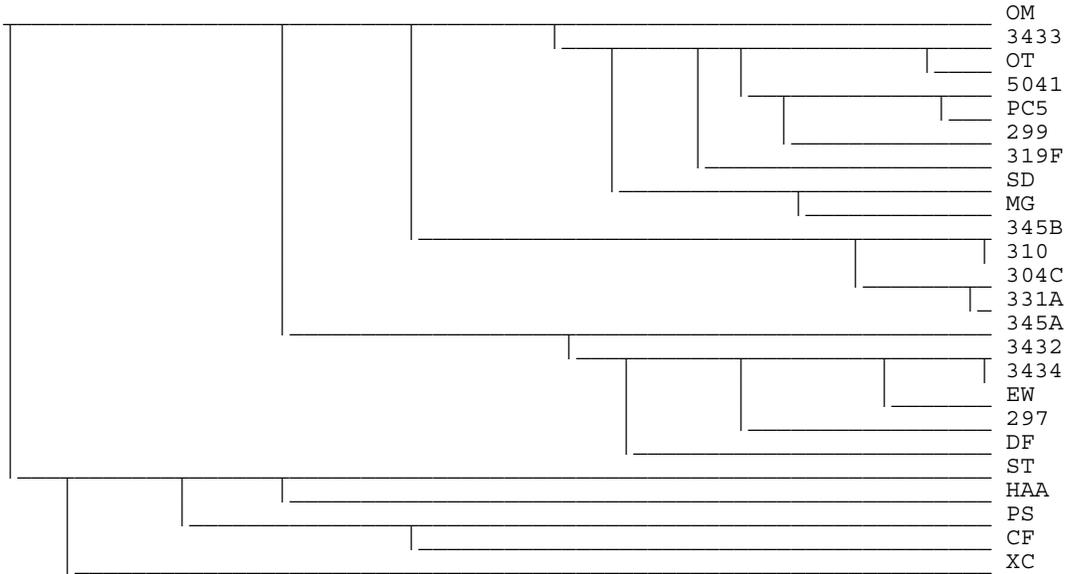
	13	14	15	16	17	18	19	20	21	22	23	24
1 OM	0.311	0.408	0.350	0.348	0.378	0.360	0.363	0.550	0.594	0.541	0.500	0.590
2 3433	0.261	0.370	0.318	0.315	0.340	0.321	0.318	0.518	0.571	0.506	0.450	0.559
3 OT	0.253	0.368	0.316	0.313	0.341	0.313	0.328	0.530	0.570	0.518	0.465	0.571
4 5041	0.249	0.343	0.291	0.288	0.322	0.282	0.297	0.512	0.576	0.517	0.458	0.556
5 PC5	0.234	0.337	0.291	0.288	0.319	0.282	0.297	0.518	0.571	0.511	0.450	0.559
6 299	0.217	0.378	0.329	0.326	0.357	0.315	0.321	0.526	0.573	0.513	0.453	0.555
7 319F	0.302	0.376	0.341	0.339	0.363	0.324	0.340	0.517	0.575	0.502	0.458	0.545
8 SD	0.299	0.327	0.271	0.268	0.302	0.289	0.283	0.523	0.566	0.503	0.425	0.569
9 MG	0.301	0.364	0.300	0.297	0.328	0.312	0.306	0.523	0.579	0.525	0.447	0.574
10 345B	0.067	0.426	0.389	0.386	0.410	0.383	0.410	0.527	0.580	0.515	0.496	0.567
11 310	0.067	0.426	0.389	0.386	0.410	0.383	0.410	0.527	0.580	0.515	0.496	0.567
12 304C	0.012	0.417	0.380	0.377	0.402	0.378	0.393	0.530	0.575	0.520	0.493	0.585
13 331A	0.000	0.420	0.372	0.369	0.393	0.369	0.393	0.524	0.567	0.517	0.485	0.577
14 345A	0.420	0.000	0.217	0.214	0.215	0.201	0.247	0.500	0.539	0.442	0.431	0.513
15 3432	0.372	0.217	0.000	0.003	0.057	0.122	0.182	0.512	0.550	0.474	0.437	0.532
16 3434	0.369	0.214	0.003	0.000	0.054	0.119	0.179	0.509	0.547	0.471	0.434	0.529
17 EW	0.393	0.215	0.057	0.054	0.000	0.144	0.201	0.518	0.551	0.475	0.447	0.533
18 297	0.369	0.201	0.122	0.119	0.144	0.000	0.208	0.512	0.570	0.460	0.429	0.527
19 DF	0.393	0.247	0.182	0.179	0.201	0.208	0.000	0.529	0.564	0.475	0.461	0.557
20 ST	0.524	0.500	0.512	0.509	0.518	0.512	0.529	0.000	0.379	0.386	0.427	0.495
21 HAA	0.567	0.539	0.550	0.547	0.551	0.570	0.564	0.379	0.000	0.454	0.467	0.578
22 PS	0.517	0.442	0.474	0.471	0.475	0.460	0.475	0.386	0.454	0.000	0.311	0.453
23 CF	0.485	0.431	0.437	0.434	0.447	0.429	0.461	0.427	0.467	0.311	0.000	0.445
24 XC	0.577	0.513	0.532	0.529	0.533	0.527	0.557	0.495	0.578	0.453	0.445	0.000

DNA Sequencing

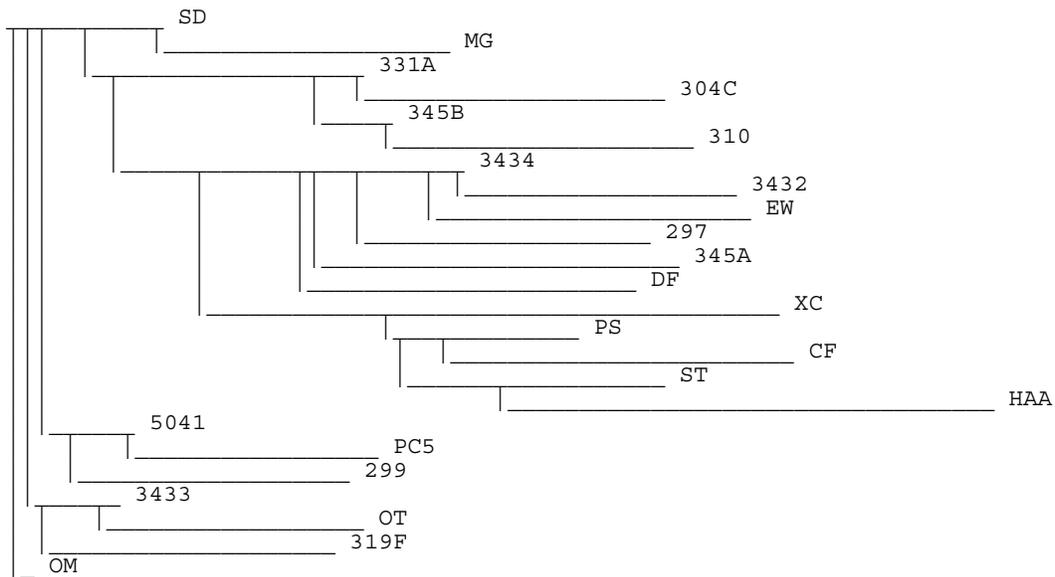
Then we get two phylogenetic trees; each is based on a particular mathematical algorithm.

CLUSTER ALGORITHM

0.530267



TOPOLOGICAL ALGORITHM



Bibliography

Students need to have access to up-to-date information on sequencing, ribosomal DNA, and mitochondrial DNA. Any one of several good modern texts in genetics or cell biology should suffice but we especially recommend the following text and its chapter 9, "Molecular Structure of Genes and Chromosomes" and the many links to related subjects.

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D. and Darnell, J., 2000, "Molecular Cell Biology", W. H. Freeman & Co., New York, 4th ed., 1084p.

The reference for the Web site Multalin is the following:

Corpet, F. 1988, Multiple sequence alignment with hierarchical clustering. Nucl. Acids Res. 16:10881-10890.

An excellent article which discusses molecular phylogeny and gives several mitochondrial primer sequences is the following:

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., and Flook, P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Annals of the Entomological Society of America. 87: 651-701.

Acknowledgments

We are grateful to Robert Elbel for much information about collecting and analyzing insects, to Jon Seger for introducing us to mitochondrial PCR and insect mitochondrial primers, to Patrice Corneli for advice about phylogenetic trees, to Eric Kofoid for programming, and to Margaret Robertson and her staff at the Huntsman Cancer Institute sequencing lab for much help with sequencing.