**From Bugs to Barcodes: Using Molecular Tools to Study Biodiversity[[1]](#footnote-1)**

Background

Biodiversity, usually defined as the number of species in a specific ecosystem or area, is important for numerous reasons. A diversity of organisms support ecosystem services such as purification of air, climate control, water purification, food production, pollination, and erosion prevention. Many people feel that biodiversity is important for aesthetic, ethical, and cultural reasons as well. But biodiversity is at risk. Habitat destruction is probably the most serious threat, although overexploitation of natural resources and invasive species play a role, and climate change will become increasingly important in the future.

However, we cannot assess the impact of any of these threats if we do not know what is being threatened. Current estimates of the total number of eukaryote species on Earth range from five to ten million (May, 2010). Of this, less than two million species have been named or described, and this lack of information is not distributed evenly among taxa despite 250 years of modern taxonomy. Although groups such as mammals and birds are fairly well known, it is estimated that 70% of arthropod species have yet to be discovered (Hamilton *et al*., 2010). This is particularly troubling at a time when human activities are impacting virtually every organism on the planet. This knowledge gap is a huge obstacle for conservation efforts. It is critical that we develop a better understanding of what organisms exist if we want to conserve species, or even know the effect of our conservation strategies. With DNA barcoding, students can help in this effort.

Species are usually identified by their morphology. It is possible for students or other non-experts to identify large organisms such as birds or mammals this way, but identifying smaller organisms such as invertebrates can be very difficult. Morphological differences are almost always so subtle it takes an expert to distinguish species. Also, as is the case with cryptic species, there may not be any morphological differences even when the genetic evidence suggests that the organisms do not interbreed, and thus by definition are different species.

Molecular taxonomists proposed using a specific DNA sequence called the barcode as an identifier for species (Herbert *et al*., 2003). There has been debate about reliability of the barcode sequence to detect taxonomic subtleties, but many taxonomists have embraced the use of barcode sequences as an additional tool. The Consortium for the Barcode of Life (CBOL) is an international collaboration of experts in genomics, taxonomy, and computer science whose mission is to create a reference library of DNA barcodes in the form of the Barcode of Life Database (BOLD, <http://www.barcodeoflife.org/>).

Students at any academic institution can become involved in barcoding projects and even contribute novel sequences to the BOLD database. Barcoding is presently being used in a variety of educational settings as a means of involving students in discovery-based science (Santschi *et al*., 2013; and <http://www.urbanbarcodeproject.org/>). Barcoding can also be used to integrate concepts and provide hands-one exposure to techniques in a variety of different disciplines such as ecology, taxonomy, genetics, molecular biology, and bioinformatics.

As part of our effort to bring authentic research into large undergraduate biology labs, we have initiated a DNA barcoding project in which students are documenting biodiversity at the UC San Diego Scripps Coastal Reserve. We have used barcoding both to discover the diversity of species in a particular habitat as well as to test specific hypotheses. For example, one quarter we documented the vegetation-inhabiting spiders at our reserve and another quarter we looked at the intraspecies diversity in the honeybee *Apis mellifera,* which has a number of subspecies. Our hypothesis-testing projects have explored whether polychaete worms from different intertidal zones of the sandy beach are different species and whether flower-inhabiting thrip species specialize on different species of plant hosts. Other groups have used barcoding for similar studies, both basic and applied. For example, students have done barcoding to detect whether fish being sold in stores are the actual advertised species (Stockle *et al*., 2010).

In order to do barcoding, students collect specimens, extract DNA, then amplify the DNA barcoding region using Polymerase Chain Reaction (PCR). After running a gel to verify that they have a PCR product of the correct size, students then purify the PCR product and send it for Sanger sequencing. The DNA sequences are then analyzed using several free bioinformatics programs. The methods are all straightforward and require only basic specimen collection and molecular biology lab equipment. We typically spread the experiments out over four lab periods of about 3 hours each, and the schedule can be adjusted to fit lab periods of shorter or longer times. The experimental protocols in this paper have been designed for use with insects but they can be used with other invertebrates as well. However, the DNA extraction methods and primers should be tested with the type of animals to be studied in your class before actual implementation of the module.

**DNA Barcoding:**

In order to barcode an organism, DNA is first extracted from an organism and then the barcode sequence is amplified using PCR. Because we want to simply amplify and sequence the DNA without having to clone it, it is important to use a haploid gene. Mitochondrial DNA is only inherited from the mother, and thus all the genes on the mitochondrial DNA are haploid. Mitochondrial genes also have a low level of intraspecies diversity and a high level of interspecies diversity, which makes it useful for differentiating species based on DNA sequence differences. Also, there are many copies of DNA per mitochondria, and there are many mitochondria per cell so the copy number of mitochondrial genes is higher than nuclear genes.

We will be amplifying DNA using part of the cytochrome c oxidase (CO1) gene located on mitochondrial DNA. This gene has been accepted by scientists as the standard gene to be used for all animal barcoding studies (Herbert *et al*., 2003). For the barcode PCR, we use what are known as “universal primers” which are designed to recognize conserved areas in the CO1 gene in many invertebrate species. Because the primer sequences will not be an exact match to the CO1 target sequence in all invertebrate species, the PCR reaction is performed at a low annealing temperature. This should allow primers that are not an exact match to still anneal well enough to form a stable duplex for the PCR reaction.

After running the PCR, some of the PCR product is run on a gel to make sure it is the expected size which is about 660-680 base pairs. The remainder of the PCR sample is then cleaned up to remove the free nucleotides, primers, and enzyme and it is then sent for Sanger sequencing. The sequencing results are analyzed using several free bioinformatics programs and databases.

Week 1: Bug Collection

*Places to find insects*

Flowers – Pollinators (bees, etc.) are very common at flowers. Tap a flower over a white tray, white paper plate, or white pad of paper to find thrips, spiders, and other species that lurk therein.

Leaves – Look on the underside of leaves for plant sucking insects. Look for evidence of leaf chewing, which suggests that herbivorous insects might be active. Galls are oddly shaped plant growths caused by the immature insect developing inside. Break open the gall to see if the insect is still there.

Underneath logs – Turn over any object that creates damp, protected conditions - stones, logs, old lumber, or trash. You will be sure to find earwigs and other moisture-loving insects. Ants, termites, roaches, beetles, and bristletails are common.

Lights – Lots of insects are attracted to lights at night, especially moths and lacewings. Look around your porch light at night.

Water – Look under stones in running streams for immature mayflies, stoneflies, and the cases of caddisflies. Water striders are common walking on water, and look in the shallows along the edges of ponds for various aquatic beetles and immature dragonflies, midges, and mosquitoes.

Basements – Look in old books and newspapers for silverfish and booklice which are primarily feeding on the mold that grows in humid conditions.

Traps – Put out fruit, rotting or otherwise, as baits to attract insects. Try meat, cookies, or a soda. Create habitat by putting out pieces of wood.

**Protocol**

1. Have the students collect insects using the above techniques.
2. Place the bags containing the insects in the freezer. Make sure the bag is labeled with the date, temperature, and GPS location.

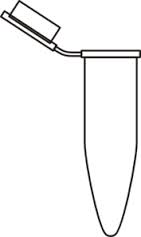
Bug Identification using Microscope

**Protocol**

1. Select a bug to use for your experiment and place it in a petri dish.
2. Place the petri dish containing your bug under the dissecting microscope.
3. Using the provided bug ID guide, determine which Order you selected bug is in. You may need to use forceps to gently move the bug to around under the microscope.

Order\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Place your bug under the microscope with the cell phone camera mount. Using your cell phone, or a group member’s phone, take a picture of the bug. Make sure to save the image as you will need it for completing the project in a few weeks.

 C:\Users\bockhs\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\PUS8R4ZN\MC900346899[1].wmf

1

1. For large insects, carefully cut off one of the back legs with scissors and save the rest of the specimen. You should have about 3 to 4 mm of leg so if the specimen legs are very small, use more than one. Try to cut as close to the body as possible. Do not simply pull off the leg.
2. Place the leg (or the entire specimen if your bug was very small) in a blue microfuge tube and label your tube with your group number.
3. Add 180 μL of ATL buffer to the tube
4. Add 20 μL of proteinase K to the microfuge tube.
5. Place your tube on the heat block that is set to 56oC. This will incubate overnight and your instructor will remove it and place it in the freezer for use in the next lab meeting.

Week 2: DNA Extraction

**Protocol**

1. Collect your bug digestion from the instructor.
2. Vortex the tube containing your digested bug tissue for 20 seconds. (check that the tube lid is closed!)
3. Add 200uL of binding buffer and vortex for 5 seconds. (check that the tube lid is closed!)
4. Add 200uL of ethanol and vortex for 5 seconds. (check that the tube lid is closed!)
5. Place a column in a collection tube. Now pipette all the liquid from your bug tissue onto this column. Label the lid of your tube.



Filter in collection tube

Filter

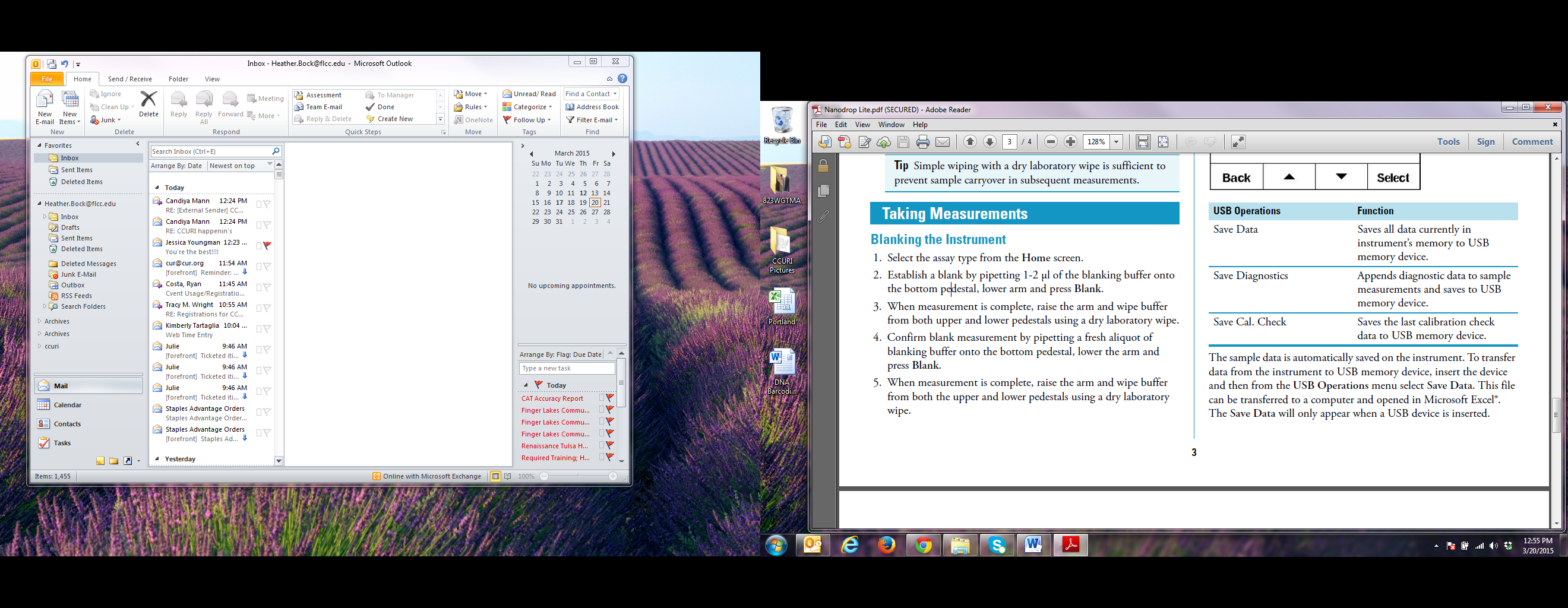
1. Centrifuge at 8,000 rpm for 1 minute. Make sure the centrifuge is balanced and the plastic lid is on. This allows the DNA to bind to the filter column.
2. Place the filter column into a new collection tube and discard the old collection tube. **Keep the filter column, this has collected all the DNA from your bug.**
3. Add 500uL of Wash buffer 1
4. Centrifuge for 1 minute at 8,000 rpm. Make sure the centrifuge is balanced and the plastic lid is on.
5. Place the filter column into a new collection tube and discard the old collection tube. **Keep the filter column, this has collected all the DNA from your bug.**
6. Add 500uL of Wash buffer 2
7. Centrifuge for 3 minutes at 13000 rpm. **Note the change in centrifuge speed!** Make sure the centrifuge is balanced and the plastic lid is on.
8. Place the filter column into a new collection tube and discard the old collection tube. **Keep the filter column, this has collected all the DNA from your bug.**
9. Centrifuge for1 minute at 13000 rpm to remove any remaining buffer. Make sure the centrifuge is balanced and the plastic lid is on.
10. Place the column in a new 1.5mL microfuge tube
11. Discard collection tube.
12. Add 50uL of elution buffer to the **center of the filter column**, not near the side
13. Let sit for 5 minute
14. Centrifuge for 1 minute at 8,000 rpm which will elute the DNA from the filter column. **Note the change in centrifuge speed!** Make sure the centrifuge is balanced and the plastic lid is on.

Note: You will not be able to close the lid of the 1.5ml microcentrifuge tube over the top of the filter column. When you place the tube with the column into the centrifuge, point the tube lid into the center of the centrifuge so that the centrifuge cap will fit.

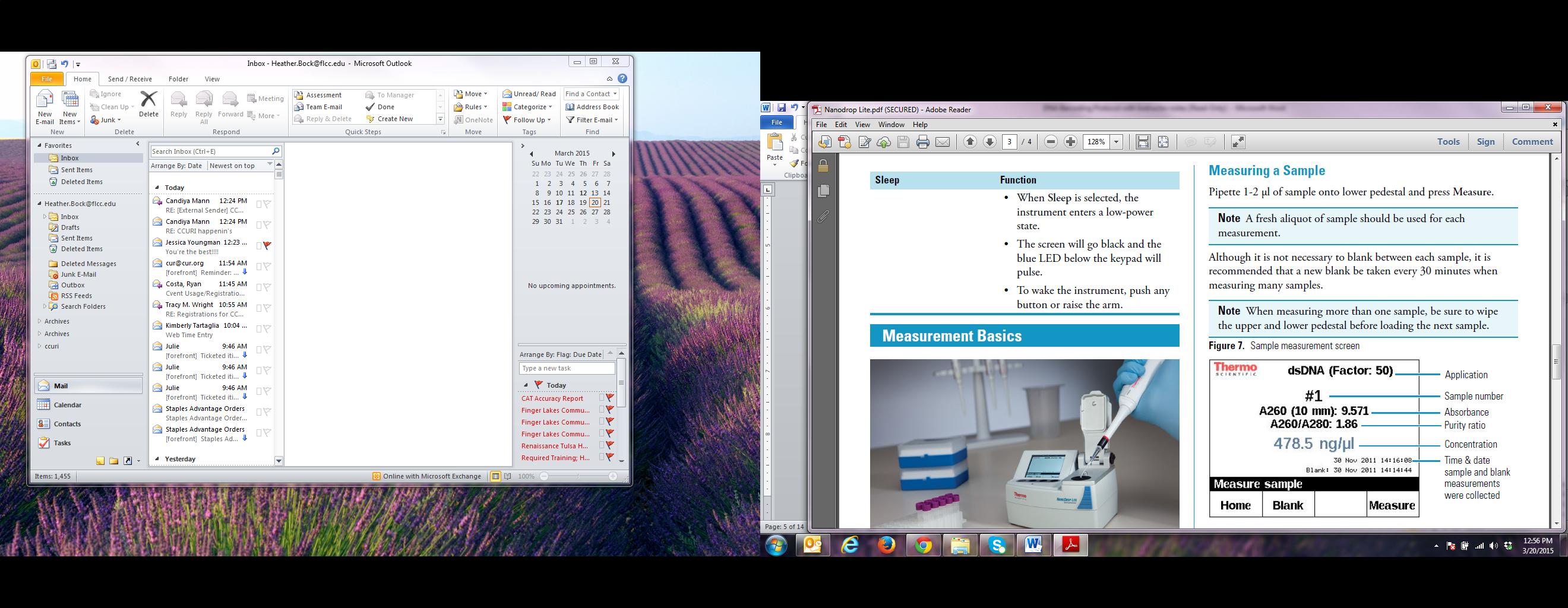
1. The flowthrough that has collected in the microcentrifuge tube contains your bug DNA. You want to save this. Remove the filter column and close the lid to the microcentrifuge tube. **SAVE THE FLOWTHROUGH – THIS HAS THE DNA.** You will use this DNA to set up the PCR reaction.
2. Calculate the DNA concentration using the Nanodrop

**Nanodrop Instructions – continued on next page**

**Your instructor will blank the nanodrop for you**



**You will start here**



DNA Concentration \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ng/ul

A260/280 purity ratio\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

A DNA concentration of more than 10ng/ul is good, 30ng/ul or higher is ideal. The A260/208 purity ratio should be close to 2.0

PCR

**Protocol**

Your instructor has prepared a mastermix with all the components required for the polymerase chain reaction (PCR).

1. Your instructor will pipet 47.5ul of the mastermix into a PCR tube.
2. Label the PCR tube you receive from your instructor.
3. Add 2.5uL of your bug DNA to the PCR tube.
4. Close the PCR tube. Be careful that you do not squeeze the tube sides as they will crack.
5. Place your PCR tube on the class rack that is on ice.
6. Place your bug DNA on the class rack that is on ice.

PCR Conditions

|  |  |  |  |
| --- | --- | --- | --- |
|  | Temperature | Time | Cycles |
| Denature | 94°C | 3 minutes | 1 |
| Annealing | 95°C | 45 seconds | 35 |
| 42°C | 45 seconds |
| 72°C | 60 seconds |
| Final Extension | 72°C | 7 minutes | 1 |
| Hold | 4°C | ∞ | N/A |

**Notes**

* The expected PCR product is 648 base pairs
* We order the primers from IDT (<http://www.idtdna.com/site>) and then store the stocks at 100 uM water. For the PCR reaction, we use 2X GoTaq Green mix from Promega. It is relatively inexpensive and already has a gel loading dye in the master mix.
* The sequences of the Folmer primers LCO1490 and HCO2198, which amplify a 660-680 bp fragment of the COI gene in a wide range of invertebrate taxa, are shown in Table 1.

|  |  |  |
| --- | --- | --- |
| Primer | Name | Sequence |
| Forward | LCO1490 | 5'-GGTCAACAAATCATAAAGATATTGG-3' |
| Reverse | HCO2198 | 5'-TAAACTTCAGGGTGACCAAAAAATCA-3' |

PCR mastermix

* Below is the mastermix reaction for a single PCR reaction. The total reaction volume per reaction will be 50 µL with the insect DNA added.

Components of and respective volumes for a single PCR reaction

|  |  |  |
| --- | --- | --- |
| Volume (µL) | Component | Stock Concentration (µM) |
| 25 | GoTaq green | na |
| 2.5 | forward primer | 10 |
| 2.5 | reverse primer | 10 |
| 17.5 | sterile water | na |

Week 3: Electrophoresis and PCR Purification using QIAGEN QIAquick PCR Purification Kit

**Electrophoresis of PCR Product**

**Protocol**

1. Obtain a flask with 0.8% agarose and pour a gel for electrophoresis.
2. Wait for the gel to solidify (this should take 10-15 minutes). During this time you will practice loading a gel.
3. Place the gel into a electrophoresis chamber and cover with TAE buffer
4. Load 5ul of ladder into the first lane
5. Load 5ul of PCR product into the third lane
6. Place the lid on the gel box
7. Connect the gel box to the power source
8. Run your gel for 30 minutes 150mV (Set timer for 30 minutes on power supply so gel run will stop automatically)
9. Turn off the power source and unplug the gel lid.
10. Carefully lift the gel tray out of the buffer. Hold the gel from the ends so that it does not slide out of the tray.
11. Take your gel to the Bio-Rad gel imager and follow instructions to capture an image of the PCR product.

**PCR Clean Up**

**Protocol**

1. Add 225ul of Binding Buffer to a 1.5ml microcentrifuge tube and label the tube with your group name.
2. Add 45ul of PCR product to the microcentrifuge tube containing the Binding Buffer.
3. Close the lid and vortex for 5 seconds
4. Place a QIAquick column into a 2ml collection tube. Label the tube with your group name.
5. Add the entire sample (about 270ul) to the column
6. Close the lid and centrifuge for 1 minute at 13,000 rpm
7. Discard the flow-through by dumping into a waste beaker and place the QIAquick column back into the same 2ml collection tube. **Keep the filter column, this has collected all the DNA from your PCR reaction.**
8. Add 750ul Wash Buffer to the QIAquick column
9. Centrifuge for 1 minute at 13,000 rpm
10. Discard the flow-through into the waste beaker and place the QIAquick column back into the same 2ml collection tube
11. Centrifuge the QIAquick column once more in the 2ml collection tube for 1 minute at 13,000 rpm. This will remove residual Wash Buffer.
12. Place the QIAquick column into a **clean** 1.5ml microcentrifuge tube. Label the side of the tube with your group name.
13. Add 50ul elution buffer to the center of the column
14. Incubate at room temperature for 1-3 minutes.
15. Centrifuge for 1 minute at 13,000 rpm.

Note: You will not be able to close the lid of the 1.5ml microcentrifuge tube over the top of the filter column. When you place the tube with the column into the centrifuge, point the tube lid into the center of the centrifuge so that the centrifuge cap will fit.

1. After the spin is complete, measure your DNA concentration on the Nanodrop.

DNA Concentration \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ng/ul

A260/280 purity ratio\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Label your tube with your group number and the DNA concentration. Bring your labeled tube (group number and DNA concentration) with the DNA to your instructor for sequencing.

**DNA Sequencing Calculation**

Your instructor does this for you.

Concentration = Mass/Volume

C=M/V

CV=M so V=M/C

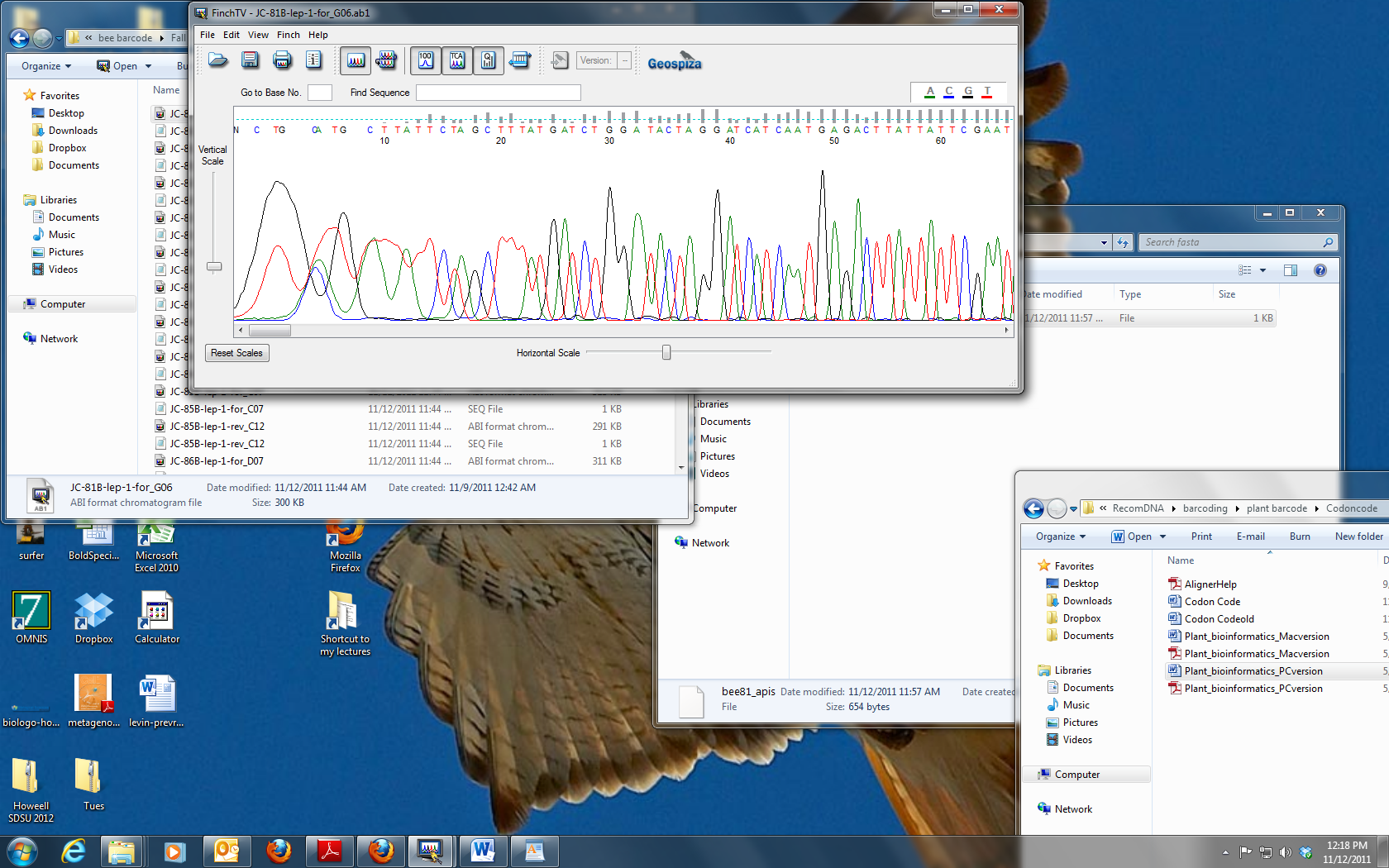
Need 20ng of sample so divide 20ng by the concentration of your sample to get the volume of sample that needs to be sequenced.

**Week 4: Sequencing Analysis**

**Protocol**

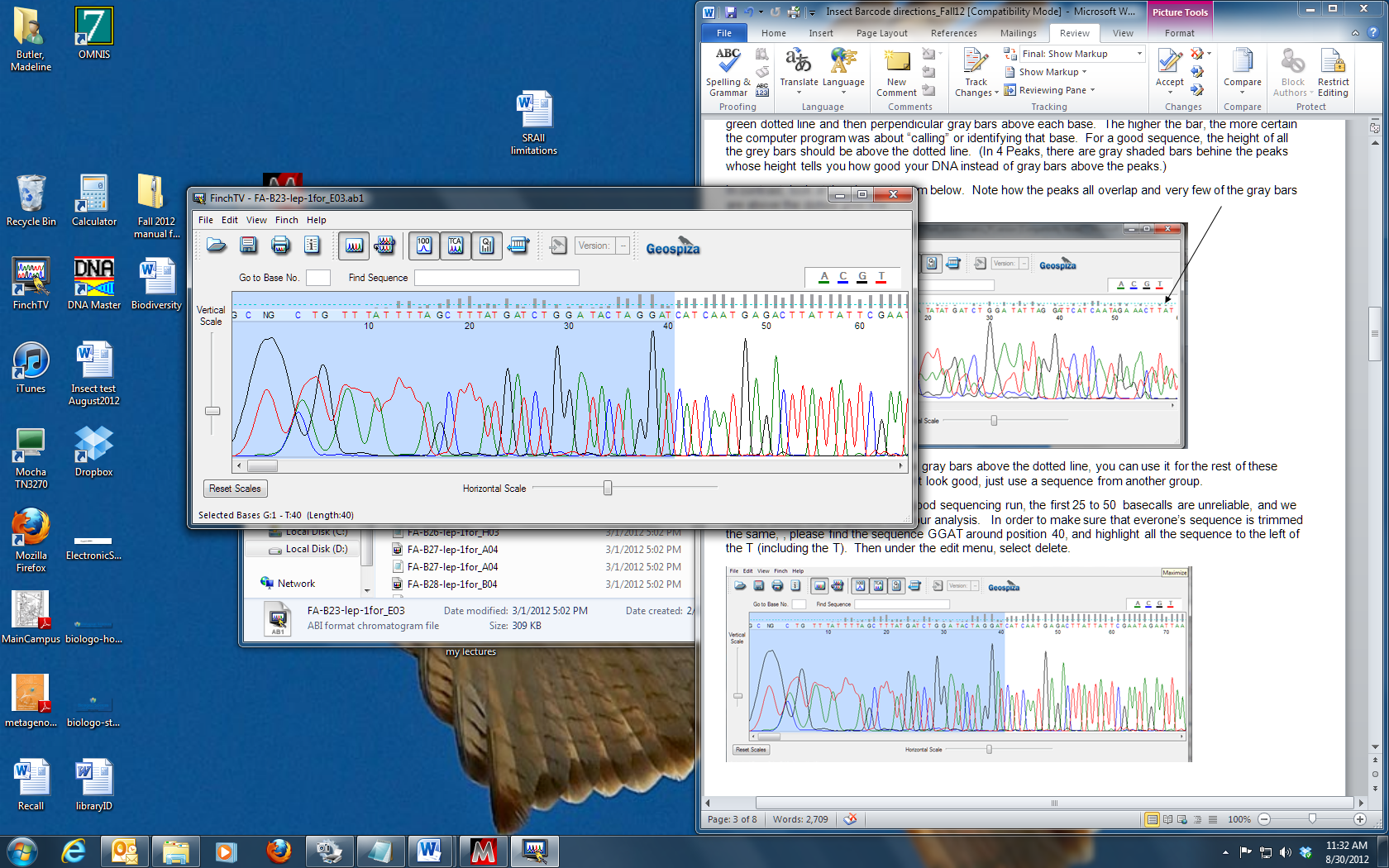
*DNA Barcoding Sequencing: Assessing the quality of the sequence and doing a BLAST**(Note: the sequences in this section are examples)*

1. The first thing you must do is look at your sequence chromatogram and determine if it is good enough to use in the subsequent analyses. Although most of the time we get PCR product, it may or may not have sequenced well.
2. Open your chromatogram (sequence) in the program Finch TV. The chromatogram should look like the one below in Finch TV. The peaks in the chromatogram represent the actual sequnce of the PCR product. Note that there are four colors, each representing a different base. Also note how the first 25 peaks or so do not look very sharp, but from about peak 25 on, the peaks are well resolved and there is no background. This is a good sequencing reaction. You can also use the gray bars above each base to tell how good the sequence is at that particular point. You can see there is a horizontal green dotted line and then perpendicular gray bars above each base. The higher the bar, the more certain the computer program was about “calling” or identifying that base. For a good sequence, the height of all the grey bars after the first 20 or so should be above the dotted line.



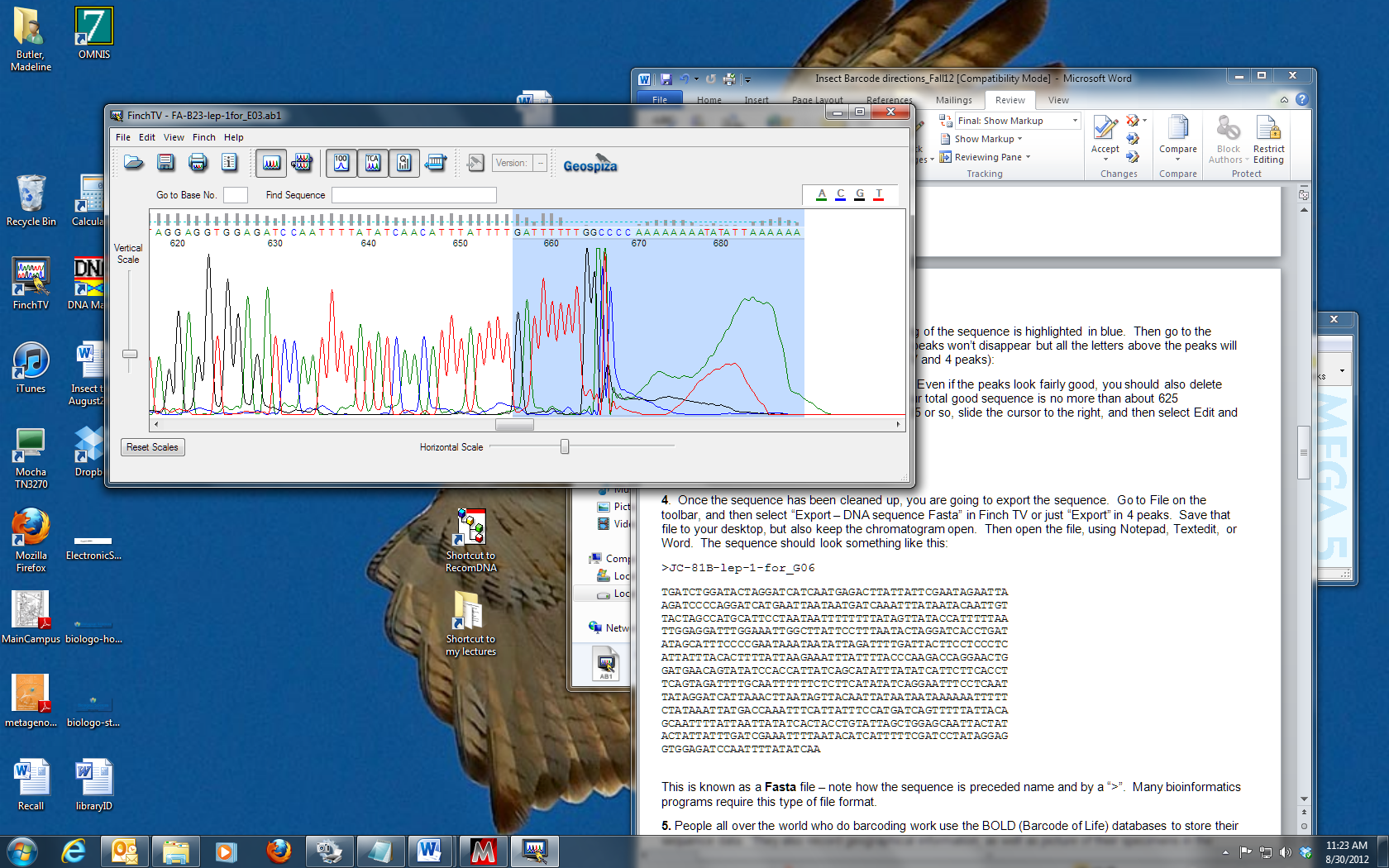
**Figure 2.** Example of a chromatogram from a good sequencing reaction

1. As mentioned above, even in a good sequencing run, the first 25 to 50 base-calls are unreliable and you need to delete that sequence from your analysis. Select the section of the sequence which does not have a gray bar above each letter. Then select Delete under the Edit menu. The actual peaks will not disappear but all the letters above the peaks will.



**Figure 4.** Trimming the beginning of the chromatogram

Now do the same for the end of the sequence. Delete this sequence.



**Figure 5.** Trimming the end of the chromatogram

1. Once the sequence has been cleaned up by trimming the ends, you are going to export the sequence. Go to File on the toolbar, and then select “Export – DNA sequence FASTA” in Finch TV. Save that file to your desktop, but also keep the chromatogram open. Then open the file, by right clicking on the file and selecting to open with Notepad. This is the text version of the chromatogram file and represents the sequence of the PCR product. This is known as a FASTA file – note how the sequence is preceded by a name and a “>”. Many bioinformatics programs require this type of file format. The exported sequence should look something like that shown in Figure 6.

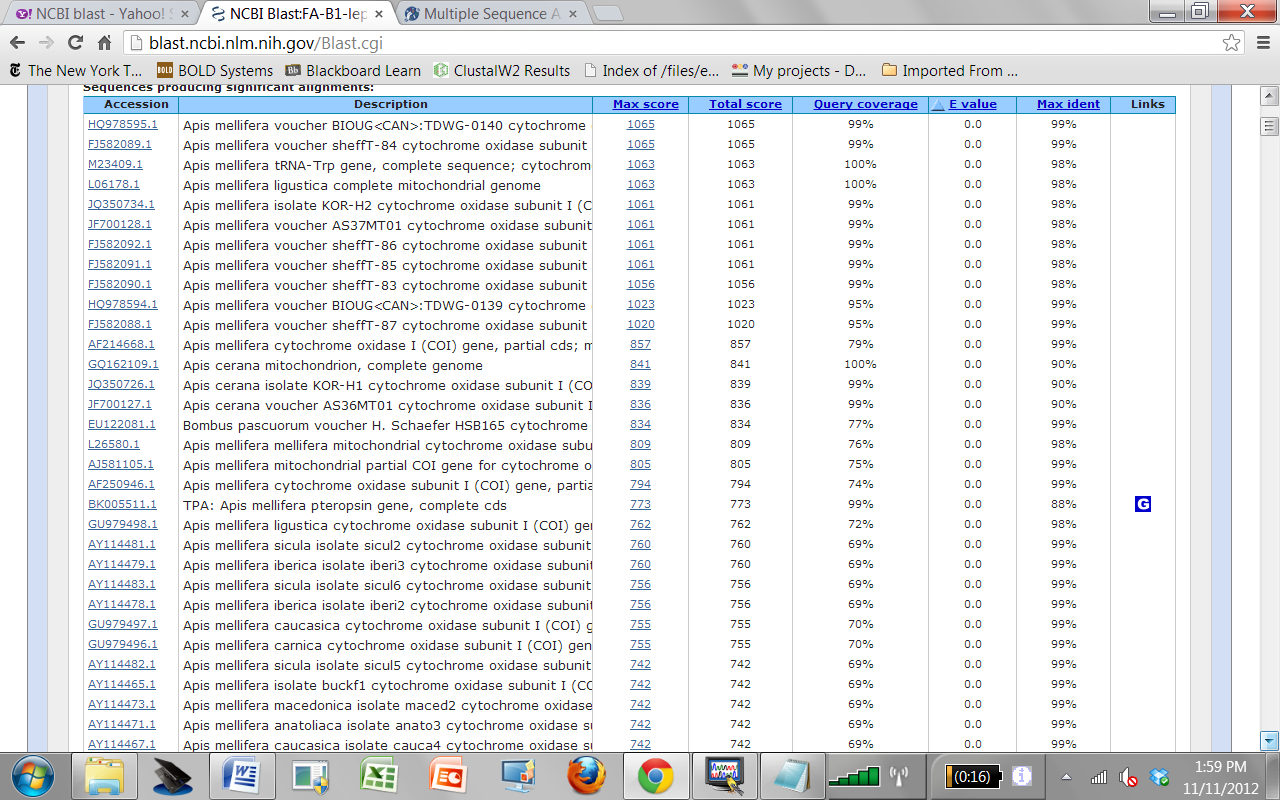


**Figure 6**. Sample of a FASTA DNA sequence file

1. Next you are going to see if there are any similar sequences to yours in the GenBank database. We will do this by using the NCBI BLAST tool.

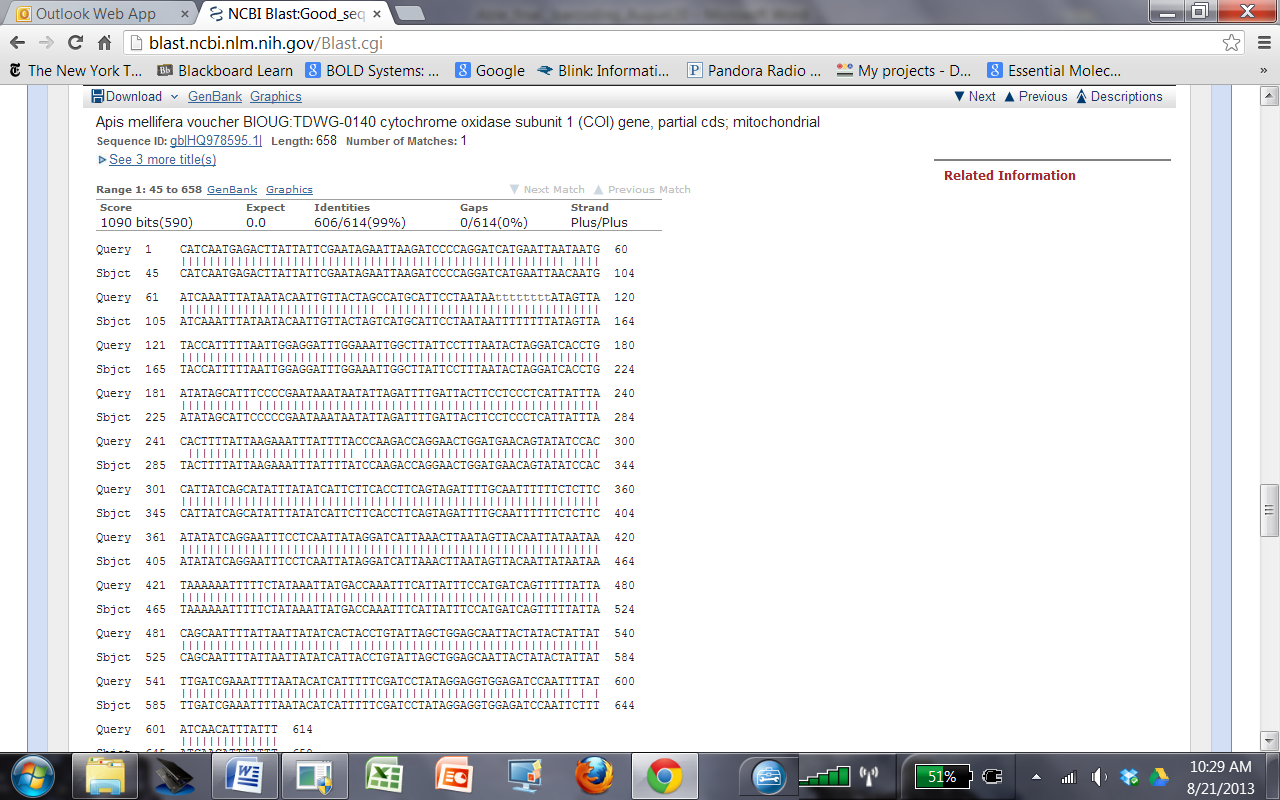
Go to NCBI BLAST (http://BLAST.ncbi.nlm.nih.gov/BLAST.cgi) and select “nucleotide BLAST”. Then under “Chose Search Set”, use the pull down menu to select “nucleotide collection” (note that the default is for Human sequence). Paste your FASTA sequence into the entry box, and then click on the BLAST button.

1. When you do the BLAST, you will get a list of entries in GenBank that are closest in match to your sequence, with the most similar match at the top of the list. Scroll through the list and look at the names – often there are several entries for the same Genus and species.



**Figure 7.** Example list of BLAST matches

Now scroll down below the list to see the actual alignment of your sequence to the top match in NCBI – it should look something like that shown in Figure 8.



**Figure 8.** Sequence alignment from BLAST

The Querysequence is the sequence you submitted to GenBank. The Subject sequence is the sequence of the closest match to your sequence in GenBank. Look first at the “Identities” value in Figure 8 – this tells you how similar your sequence is to the sequence in GenBank. In this case, the sequence was 99% identical to an entry for *Apis mellifera*. In the alignment shown in Figure 8, the numbers “604/614” tell you that 604 out of 614 bases were the same in the submitted sequence and the top match found in GenBank. The value in parentheses (99%) tells you how similar this is on a percentage basis. The Expect value (e-value) of the hit is the match you might expect by chance, given the size of the database. A smaller e-value indicates a more meaningful match.

1. For now, we are most interested in the Identities value. Although there is no hard and fast number, assume for this exercise that any match that is 97% or above is probably the correct match for your specimen – that is, assume your specimen is the same as the species identified in GenBank. If the closest match to your sequence is less than 97% to any entry in GenBank, the barcode sequence for your specimen is probably not in GenBank.
2. Copy the species name that is listed as being the closest match to your sequence and paste it into google. Record the species name below:

Latin name\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Common name\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Did the sequencing result match the order that you placed the bug into in week 1?

1. 2014 Madeline Butler, Heather Henter, and Stephanie Mel, University of California San Diego [↑](#footnote-ref-1)