Using DNA extraction, PCR and Genome Sequencing to Identify Unknown Organisms

Cameron Potts

BIO 2450 Genetics

11/19/2013

Andrea Villamizar

I certify that the following paper is all my own work, and that all of the information from outside sources is properly cited therein.

Abstract: An unknown sample of meat was identified using DNA extraction, PCR, and genome sequencing in this experiment. DNA extraction followed by PCR gives enough copies of the DNA sequence to be cross-referenced in a public database. The whole process was set up into six different laboratory sessions. One each for DNA extraction, PCR, PCR cleanup, cycle sequencing, cycle sequencing cleanup and electropherogram analysis using BLAST. The sequence of the unknown sample was found to have insufficient base pairs to run sequencing analysis. A backup sequence of the sample was searched and matched to that of the American Bison genome. The most suspected source of error was the PCR reaction. This is attributed to other studies having contamination problems with the PCR reaction and the fact that only 20 base pairs were present to analyze.

Introduction: This experiment involves use of laboratory techniques such as DNA extraction, polymerase chain reaction (PCR), and genome sequencing to identify an unknown specimen.

Analyzing the genome of an organism always begins with extracting the DNA from that organism (Huaqiang 2013). Genetic mapping, genetic fingerprinting, phylogenetic analysis, or any PCR based study requires DNA extraction (Huaqiang 2013). Diseases caused by microbes

can be diagnosed based on what type of sequence its DNA contains (Pan 2013). A good example of this is the diagnosis of pulmonary tuberculosis based on the identification of *Mycobacterium tuberculosis* DNA (Pan 2013). Markers in the genome associated with anti-malaria drug resistance can be identified by this method as well (Morris 2013). Different methods of extraction are used according to the organism being tested. In a study done in *The Journal of Thoracic Disease*, the efficiency of different methods was tested in detection of *M. tuberculosis* DNA in clinical specimens (Pan 2013). Some of the methods involved using phenolchloroform, magnetic beads, or extraction kits sold commercially for the sole purpose of DNA extraction (Pan 2013). Efficient extraction is important because DNA needs to be present in large enough quantities in the final product to be able to undergo a polymerase chain reaction (Morris 2013).

Polymerase chain reaction (PCR) makes numerous copies of a section of DNA so that it is easier to study in the laboratory (Russel 2010). It is a commonly used technique for genetics experiments (Russel 2010). An experiment in the *Journal of Thoracic Disease* used PCR along with DNA extraction to identify *M. tuberculosis* quickly and effectively (Pan 2013). Malaria was also identified using PCR in an experiment published in *Malaria Journal* (Morris 2013). Using PCR to identify pneumonia in a study conducted in *Indian Journal of Medical Research* was laborious to the researchers due to contamination, so following the protocol carefully is very important to obtain the correct sequence (Chaudry 2013).

Once a gene sequence is obtained by PCR, it can be searched in a public database to identify the correct organism that matches it (Russel 2010). Researchers from all around the world have come together to make a copy of the human genome available (Lander 2001). Information about the human genome can provide information regarding physiology and even disease prevention (Lander 2001). Another organism whose sequence has been published is the

honeybee (Sequencing Consortium 2006). Sequencing is just as important in this experiment as DNA extraction and the PCR reaction.

Once the DNA was extracted from the organism, a PCR reaction was able to happen followed by a search of the genetic sequence to correctly identify the unknown organism.

Methods: Six different major steps were needed to identify the unknown sample. Each one of these steps was completed in one lab session per week. These steps were DNA extraction, PCR, PCR cleanup, cycle sequencing, cycle sequencing clean up, and electropherogram sequence analysis using GenBank.

The first major step was the extraction of the DNA. A small piece of unknown animal tissue (sample C99) was obtained and placed in an eppendorf tube for DNA extraction. 180 micro-liters of Buffer ATL were added to the sample in the eppie tube to lyse the cells' membranes along with 20 micro-liters of proteinase K to digest the cell's protein tissues. The sample was then incubated for one hour at 56 degrees Celsius to allow time for the reactions to occur. After incubation was complete 200 micro-liters of Buffer AL was added to the unknown sample to stop the cell membrane lyses and protein digestion reactions. 200 micro-liters of ethanol (95%) was added to the sample and then vortexed to precipitate the DNA. 600 micro-liters of the sample (solution at this point) was pipetted into a spin column and centrifuged at 8000 rpm for one minute to bind the negatively charged DNA to the positively charged silica in the spin column. 500 micro-liters of Buffer AW1 was added to the DNA sample (now in the spin column) and centrifuged again at 8000 rpm for one minute to wash away any non-DNA proteins or extra cellular material that could cause contamination. The next step taken was the addition of 500 micro-liters of Buffer AW2 to the DNA sample (still in the spin column) to wash away any

salts that may contaminate the end product. After adding Buffer AW2, the sample was centrifuged again for 3 minutes at 13200 rpm. The spin column was then placed in a new eppendorf tube. 200 micro-liters of Buffer AE were then added to the membrane of the spin column to unbind the DNA. To allow time for the reaction with Buffer AE, the solution was allowed to sit for one minute at room temperature. The solution was then centrifuged again at 8000 rpm for one minute to separate the DNA from the spin column. Once the centrifuging was complete, the spin column was thrown away and the eppendorf tube containing the extracted/washed DNA was stored at 4 degrees Celsius until electorphoresis with 2% agarose gel and gel red as a stain. The electorphoresis was performed to determine the success of the extraction.

The second major step in the identification process was the PCR. A master mixture of 39.25 micro-liters of double distilled water (ddH_20), 3 micro-liters of magnesium chloride, 5 micro-liters of buffer (this included salts), 0.5 micro-liters of deoxyribonucleoside triphosphates (dNTPs), and 0.25 micro-liters of *Taq* polymerase was made and added to an empty tube. 0.5 micro-liters of a forward primer (ND4) and 0.5 micro-liters of a reverse primer (Leu) were then added to the tube with the master mix. The last addition to the mixture in the tube was 1 micro-liter of tDNA that was extracted from the unknown sample in the DNA extraction procedure detailed previously. The polymerase chain reaction was stored at 4 degrees Celsius until thermal cycling was conducted with the GeneAmp PCR system 9700. In order for the machine to begin cycling, the sample had to be incubated at 94 degrees Celsius for three minutes. Each cycle consisted of denaturation (1 minute at 94 degrees Celsius), primer annealing (1 minute at 54 degrees Celsius), and primer extention (4 minutes at 72 degrees Celsius). 40 cycles were completed. The polymerase chain reaction was ended when the sample underwent conditions of

72 degrees Celsius for 10 minutes. When thermal cycling was completed, the reactions were stored at 4 degrees Celsius until another electrophoresis was used to determine the success of the reaction.

An entire lab session was spent on cleaning up the PCR product so it would be ready for the cycle sequencing reaction. Polymerase, primers, dNTPs and salts were all components of the final PCR sample that needed to be removed in order for the cycle sequencing reaction to be successful. 40 micro-liters of membrane binding solution were added to the tube containing the PCR product. The entire solution was then pipetted into a spin column and allowed to incubate at room temperature for one minute. The spin column (inserted into a 2ml collection tube) was centrifuged at 13200 rpm for one minute. The collection tube was cleaned out before the spin column was reinserted into it. 700 micro-liters of membrane wash solution were added to the spin column before another centrifuge of 16000 rpm for 1 minute. To ensure the product was completely washed, one more 500 micro-liter dose of membrane wash solution was added to the spin column before it was placed in the centrifuge for 5 minutes at 16000 rpm. Anything that fell into the collection tube was discarded, and residual ethanol was removed by another centrifuging of 16000 rpm for 1 minute. The spin column was then transferred into a new 1.5ml collection tube before the addition of 50 micro-liters of nuclease free water. The new collection tube was incubated at room temperature for one minute before being put back in the centrifuge for 1 minute at 16000 rpm. At this point the cleaned up PCR product was in the collection tube so the spin column was discarded. The PCR product was stored in a refrigerator until it was time to do the last electrophoresis.

At the beginning of the cycle sequencing reaction two tubes were obtained. One was labeled ND4 and the other was labeled LEU. ddH_20 was added to each tube (the sought out

reaction volume was 10 micro-liters). 0.5 micro-liters of primer ND4 was added to the tube labeled ND4, and 0.5 micro-liters of primer LEU was added to the tube labeled LEU. 1 micro-liter of clean PCR product was added to each tube. The instructor was responsible for the addition of 2 micro-liters of Big Dye Terminator Cycle Sequencing Buffer to each tube. This buffer contains ddNTPs, *Taq* polymerase, salts, and dNTPs. Each cycle sequencing reaction was then stored at 4 degrees celsius until another thermal cycling procedure was performed (see above). The finished cycle sequencing reaction was stored at 4 degrees celsius until it was time to clean up the reaction.

The protocol for cleaning the cycle sequencing reaction was performed as follows. The ND4 and LEU tubes were centrifuged for a small amount of time. 2 micro-liters of 125mM EDTA was pipetted to each of the tubes labeled ND4 and LEU. Next 2 micro-liters of sodium acetate were added to the both tubes. Cold 95% ethanol in a volume of 50 micro-liters was then added to each tube before centrifuging at 3250 rpm for 45 minutes. The ethanol was then removed by putting it into paper towels. 70 micro-liters of 70% cold ethanol were again added to the tubes followed by another centrifuging of the tubes at 3000 rpm for 15 minutes. The ethanol was removed by placing it on paper towels, just as the first time. The samples were then dried in an incubator for 5 minutes at 37 degrees celsius. 10 micro-liters of formamide were added to the sample to re-suspend the DNA when the samples were dry. This completed the cycle sequencing reaction.

An electropherogram (picture of the sequence) was provided by the use of an ABI PRISM 3500 XL Genetic Analyzer. The electropherogram was uploaded into sequence analyzing software named GENIOUS to see the final sequence result and edit any extra bases that are not necessary for the search of the matching sequence. Once the sequence was ready, the

BLAST feature on the www.ncbi.nlm.nih.gov website was utilized to find the species with the matching DNA sequence.

Results: The results of the DNA electrophoresis show that the DNA extraction was successful (refer to figure 1). The strong bright bands in sample C99 of the PCR electrophoresis (figure 2) indicate a successful polymerase chain reaction. The cleanup of the PCR product went smoothly as well for the majority of the samples tested (figure 3). A backup sequence from a successful reaction using a C sample was used because the original C99 sample only provided a sequence that had 20 base pairs. The sequence that matched to unknown sample C99 belongs to Bison bison (97% match). The E-Value was 0.0, for the rest of the information regarding the American Bison (length of sequence, accession number, name of sequence) refer to figure 4.

Discussion: The fact that the original sample only came out to contain 20 base pairs in the final sequence analysis indicated that an error occurred somewhere in the procedure. One possible source of error was the PCR reaction. The PCR reaction should have multiplied the number of base pairs available to study. Researchers doing a PCR reaction in identifying pneumonia reported contamination of the final product, so this could have been the case in this experiment (Chaudry 2013). The 97% match between the American bison sequence and the backup sequence occurred due to the GENEIOUS software not allowing editing in the final copy of the sequence. If the GENEIOUS software would have allowed editing then the sequence would have most likely matched closer to 99% or 100%.

Literature Cited

Chaudry, R., Shama, S., Javed, S., Passi, K., Dey, A., Malhotra P., 2013 Molecular detection of *Mycoplasma pneumoniae* by quantitative real-time PCR in patients with community acquired pneumonia. *Indian Journal of Medical Research* **138:** 244-251

Huaqiang, T., Haito, H., Jianyao M., Huanxiu, L., 2013 Comparitive Analysis of Six DNA Extraction Methods in Cowpea (*Vigna unguiculata L. walp*). *Journal of Agricultural Sciences* **5:** 82-90

Lander, ES, 2001 Initial Sequencing of the human genome. *Nature* **409:**860-921

Morris,U., Aydin-Schmidt,B., Shakely, D., Martensson, A., Jornhagen, L., Ali, A., Msellu, M., Petzoid, M., Gil, J., Ferraria, P., Bjorkman, A., 2013 Rapid diagnostic tests for molecular surveillance of *Plasmodium falciparum* malaria-assessment of DNA extraction methods and field applicability. *Malaria Journal* **12:**1-6

Pan,S., Gu, B., Wang, H., Yan, Z., Wang, P., Pei, H., Xie, W., Chen, D., Liu, G., 2013 Comparison of four DNA extraction methods for detecting *Mycobacterium tuberculosis* by real-time PCR and its clinical application in pulmonary tuberculosis. *Journal of Thoracic Disease* **5:** 251-257

Russel, P.J., 2010 Recombinant DNA Technology. *IGenetics*: 263-265

Sequencing Consortium, The Honeybee Genome, 2006 Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **444:**512

Appendix:

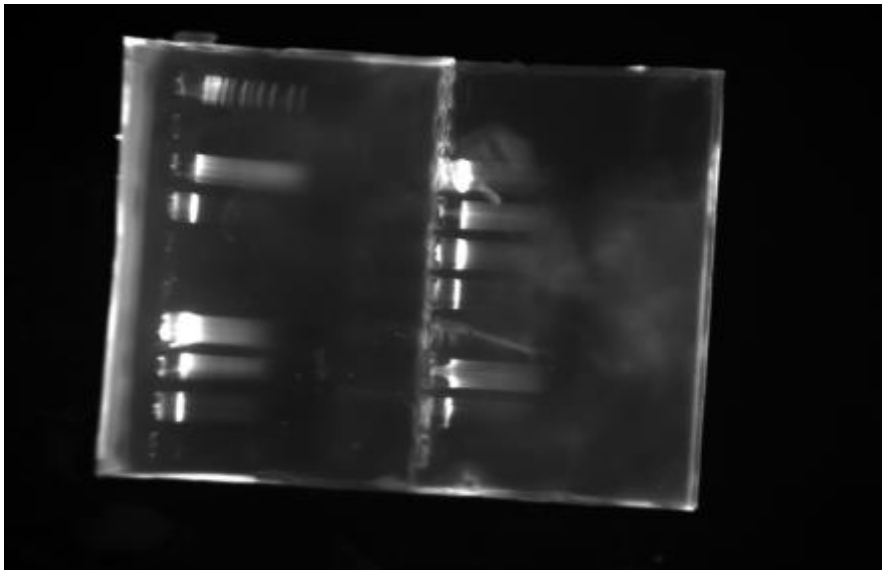

Figure 1- DNA Extraction Gel Electrophoresis

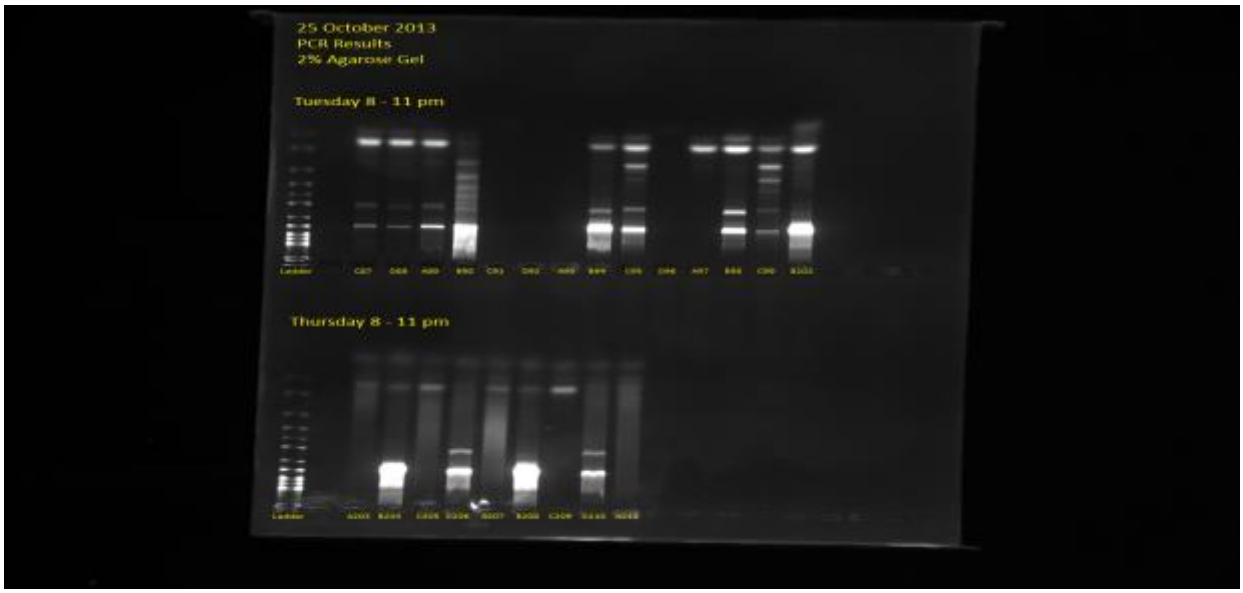

Figure 2- Electrophoresis following PCR, the sample referred to in this paper is C99

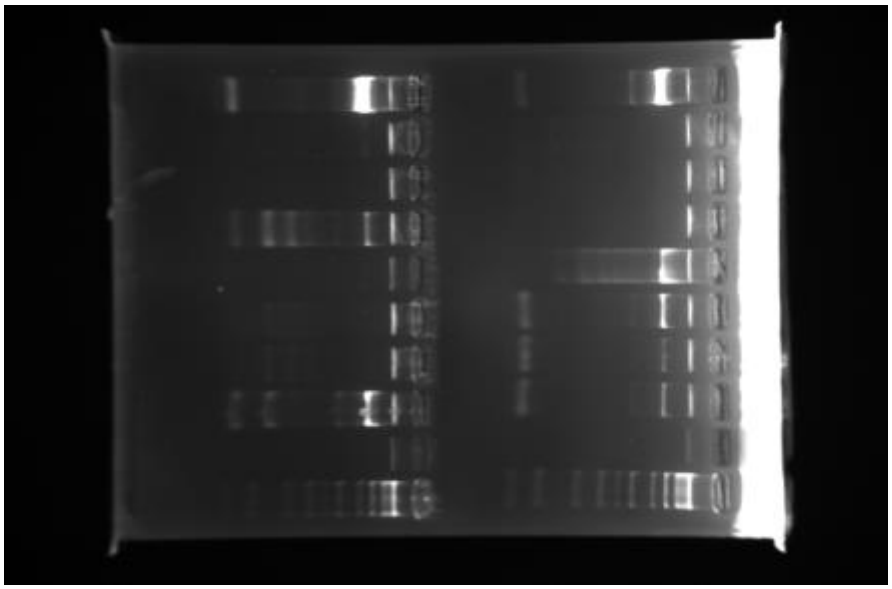

Figure 3- Gel electrophoresis following PCR cleanup

LOCUS GU947004 16323 bp DNA circular MAM 22-JUN-2012

DEFINITION Bison bison isolate BYNP1586 mitochondrion, complete genome.

ACCESSION GU947004

SOURCE mitochondrion Bison bison (American bison)

ATKAGGTGCAGMWCAWACGATGTAGTAGTAATGCGRTCCTACTAAAACTAGGAGG
GTACGGKRTGCTACGAATCACACTAAATCWAAATCCTATAACCGACTTTATAGCAT
ATCCATTCATTATWCTCTCCTTATGAGGCATAATCATAACCAGCTCAATCTGCCTCC
GTCAAACGGACCTAAAATCACTCATTGCATACTCTTCTGTAAGCCACATAGCGCTCG
TCATCGTAGCTATCCTTATCCAAACACCTTGAAGTTACATAGGAGCAACCRCTCTCA
TGATTGCCCATGGCCTCACATCCTCCATACTTTTCTGTCTAGCAAACTCAAACTACGA
ACGAATCCACAGCCGAACTATAATTCTAGCTCGAGGCCTACAAACGCTCCTTCCACT
AATAGCCACCTGATGACTACTAGCAAGTCTAACCAACCTAGCTTTACCCCCAACAAT
CAACCTAATTGGAGAGCTATTTGTAATAATGTCAACCTTTTCATGGTCTAACATTAC
AATTATTCTAATGGGAGTGAATATAGTAATCACCGCCCTATACTCTCTGTATATACT
AATTATAACCCAACGAGGAAAATACACCCACCACATTAATAATATCTCACCTTCATT
TACACGAGAAAATGCACTCATATCGYTGCACATCTTACCTTTACTACTCTTATCCCTA
AACCCAAAAATTATTCTAGGACCTCTATACTGTAAATATAGTTTAACAAAAACATTA
GATTGTGAATCTAACAATAGAAACTCATTACCTTCTTATTTACCGAAAAAGTATGCA
AGAACTGCTAATTCTATGCTCCCATACTTAACAGTACGGCTTTTTCGAACTTTTAAAG
GATAGTAGTCATCCGTTGGTCTTAGGAACCGAAAAATTGGTGCAATCCAAGTAAAA
GTAATGAA

Figure 4- Results from BLAST of genetic sequence used