Determination of road kill species based on Genetic sequencing of DNA sample

Jeremy Salinas

BIO 2450 Genetics

Nov 23, 2016

Edward Leone

ABSTRACT

DNA strands are the bases of growth in genetics, through research of DNA scientists have discovered the similarities that much species and genus' share with each other. Using previous knowledge, often first performed by hand, of how to replicate, clean, sequence and process DNA, an experiment was designed to take an unknown tissue sample and figure out its Genus and Species through a series of experiments. DNA extraction, Electrophoresis, PCR sequencing, PCR clean up, CSR protocol, CSR cleanup, and the software programs GENEIOUS and GenBanks "BLAST" were used to complete this process. Our findings were that our species was Bison Bison Bison, which is the American Bison, the E Value was 0.0 and it was a 97% match with 1591 bases matched, The Accession value is KX451357. This shows that even with a small amount of DNA, forensics scientists can replicate and find information leading to large animals. Further research could be done to find how shift Mutations would lead to differences in species that would lead to mistakes in BLAST's systems. The concept that before a genetic bank existed there was much guessing at what strands cross species had in common based on phenotypes is also an interesting way to look at how far we have come as an innovative culture.

INTRODUCTION

DNA is an essential building block to protein synthesis in microorganisms, both eukaryotic and prokaryotic. Bacteria and Humans contribute biological information through their DNA sequencing that allows for high scale data analysis (Kürekçi G, Dinçer P, 2014). There is a lot of ways science has moved innovatively forward, in an attempt to speed up processes known to be successful in DNA extraction based on sequencing and PCR (Saba et al., 2016). Through innovation we can become very accurate at identifying species based on DNA sequences, although there is always exceptions, such as the way genetic mutations to a sequence would

influence gene identification (Lee et al., 2015). DNA sequencing, as a way of identifying unknown species, has always been effective since the existence of a database. However things such as severe bottleneck affects can be known to reset the molecular clock (Sargsyan, 2012) leading to errors in identification or higher E values when using The GenBank Database, other errors are caused by "lack of expertise of the identifier"(Smith B, Lücking R, Johnston M., 2016). Although the room for error is there the benefits of these databases is far greater, and the techniques used are very effective. Because of Databases that exist, the clinical sequencing that is done in fields such as forensics or taxonomy, are more cost efficient and accurate than previously (Previte, 2015). The success of being able to identify an unknown tissue sample based on DNA sequencing depends on a lot of factors but mostly the ability to properly extract and replicate the right amount of DNA to allow accurate sequencing.

MATERIALS AND METHOD

DNA Extraction:

Four different types of unknown tissue were thawed and prepared for DNA Extraction, an assigned letter and number were issued to keep track of unknown samples throughout the entirety of the lab. Using sterile tools, a small piece (3mm) of The unknown sample was placed into the Eppendorf tube. After, by using the DNeasy Blood and Tissue Kit from Qiagen, often used in forensics, the DNA extraction protocol was performed.

Gel Electrophoresis:

Using pre-prepared 1% agarose gel made of TBE Buffer, made by the Lab IA, the Gel electrophoresis was performed by inserting samples of 9 microliters of isolated DNA onto Para film and 1 micro liter of GelRed Dye into wells. It was then run at 120 volts for 30-35 minutes until examination of results.

PCR:

For completion of PCR a new tube was acquired and labeled, in the tube 23 microliters of "master mix" was added, this was prepared by the IA to reduce contamination and expense, it contained 10.5 microliters of ddH20 and 12.5 microliters of Green Taq mix. Then into the same tube 0.5 microliters of forward primer, ND4, and Reverse primer Leu was added. 1 microliter of unknown DNA was then added into the tube as well. Thermal cycling was then completed by the lab IA with the GeneAmp PCR System 9700, consisting of incubation and cooling cycles. The tubes were then stored in -20 degree Celsius temperatures.

PCR Cleanup:

In the same tube were PCR was performed we added 4.0 Microliters of Shrimp alkaline phosphatase, 1.0 microliters of Exonuclease 1 and 15 microliters of H2O, the tube was then capped and placed in a thermal cycler following the ExoSap Protocol by Affymetrix. Success of PCR cleanup was determined by the lab IA and stored until CSR.

Cycle Sequencing Reaction:

Two sterile strip tubes were labeled using the same numbers and letters, into these tubes 5.5 microliters of ddH2O were added as well as 2 microliters of clean PCR product. The lab IA then added 2 microliters of Big Dye Terminator Cycle Sequencing Buffer. However the distinction between the strip tubes is that in one ND4 was added, while in the other LEU primer was added. They were then put through another process of thermal cycling conducted by the Lab IA with the GeneAmp PCR system 9700, it was again stored at -20 Celsius until CSR cleanup.

CSR cleanup:

Ice was prepared in order to complete this protocol on ice, staying cool is essential to the ND4 and Leu used. In two new tubes, one labeled ND4 and one labeled Leu, 2.0 microliters of

125mM EDTA, 2.0 microliters of 3M NaOAc, and 50 microliters of cold 95% ethanol were added to each tube. Both were centrifuged at 2500G for 30 minutes. The IA removed supernatant liquid and 70 microliters of 70% EtOH was added. Both tubes were centrifuged at 1650G for 15 minutes, after completion again The IA removed any supernatant liquid. The samples then air dried to allow evaporation of EtOH, after evaporation 10 microliters of formamide were added to the tubes.

Electropherograms and GenBank:

Using the software program "Geneious" alignment and analysis of our DNA Strands were completed, after editing and correcting mismatched sequences, a 50% strict consensus was selected and used to form a new sequence which would be representative of our findings to be used in the software program BLAST. After entering the determined DNA sequence in GenBanks BLAST program we found the species name, E-Value, Number of Bases and Common name for our unknown species, this of course was then recorded.

RESULTS

During electrophoresis we kept track of which DNA sample well was our own, as seen in figure 2 below (B11). However after most of these were tested they were found successful at containing valid DNA samples, as seen in figure 3 below. Because of this we were given stock solutions found successful by the Lab IA, leading to a reassignment of numbers and letters (B39). This led to success in PCR and PCR Clean up electrophoresis shown in figure 4 below. This new successful solution was continuously used throughout the remainder of the experiment, Using the DNA template strand created through Electropherograms in the program GENEIOUS, we produced the strand seen in figure 1. Applying this to BLAST led us to find that our species was

Bison Bison Bison, which is the American Bison, the E Value was 0.0 and it was a 97% match with 1591 bases matched, The Accession value is KX451357.

FIGURE 1

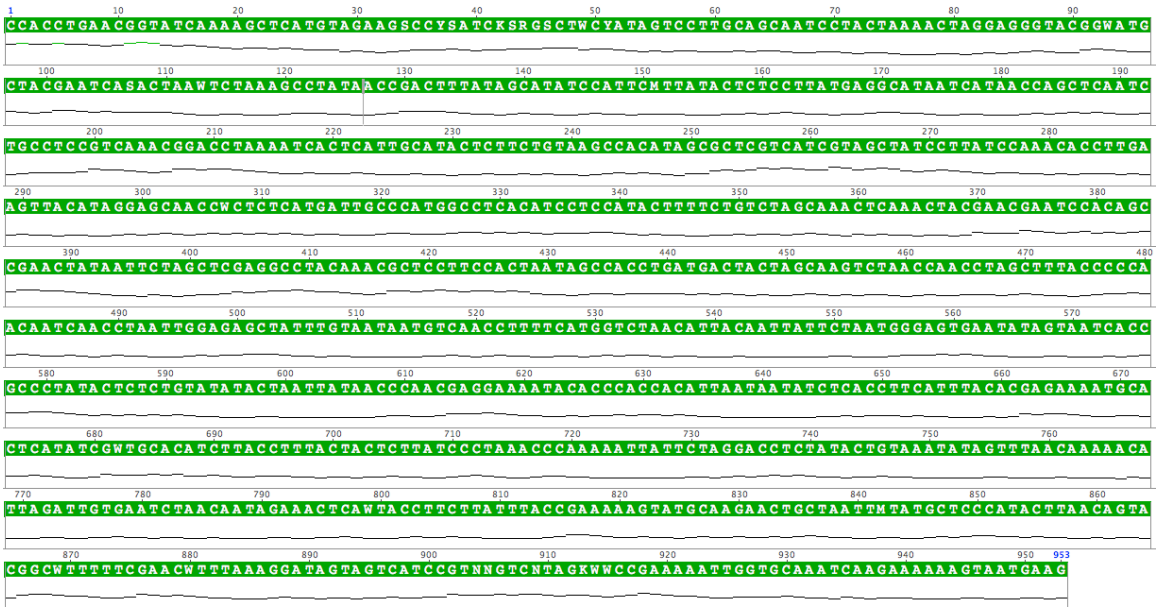


FIGURE 2

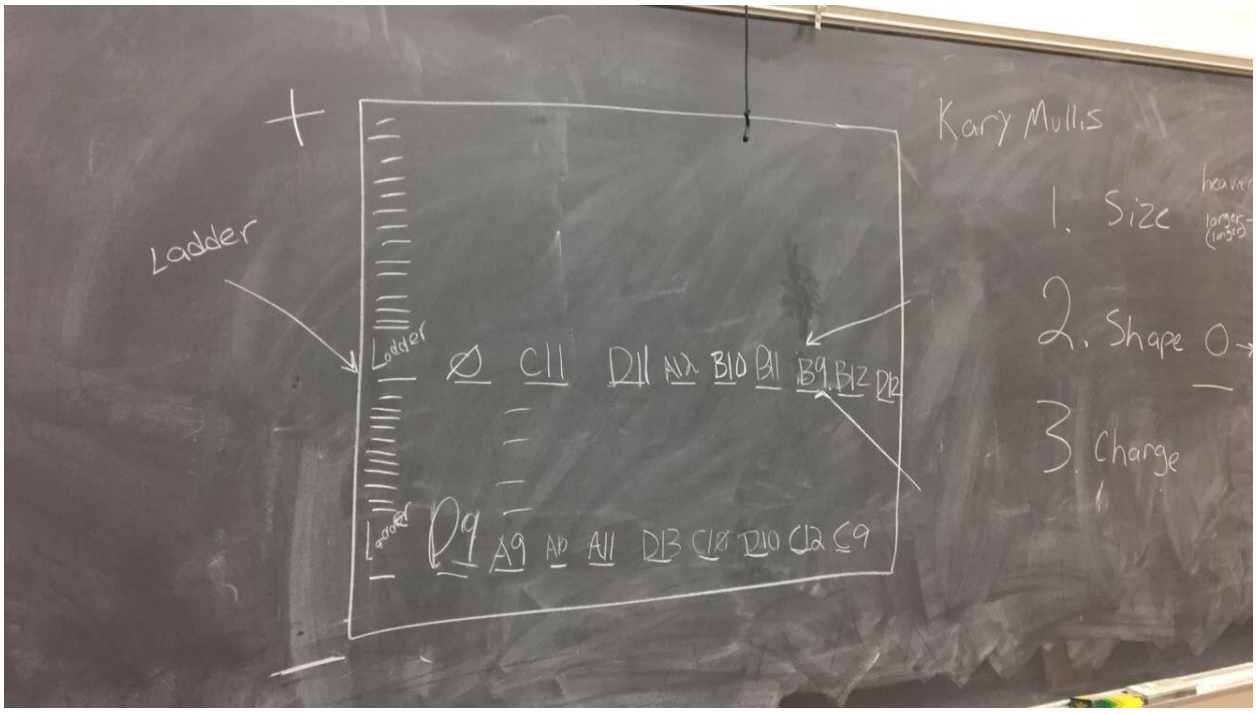


FIGURE 3

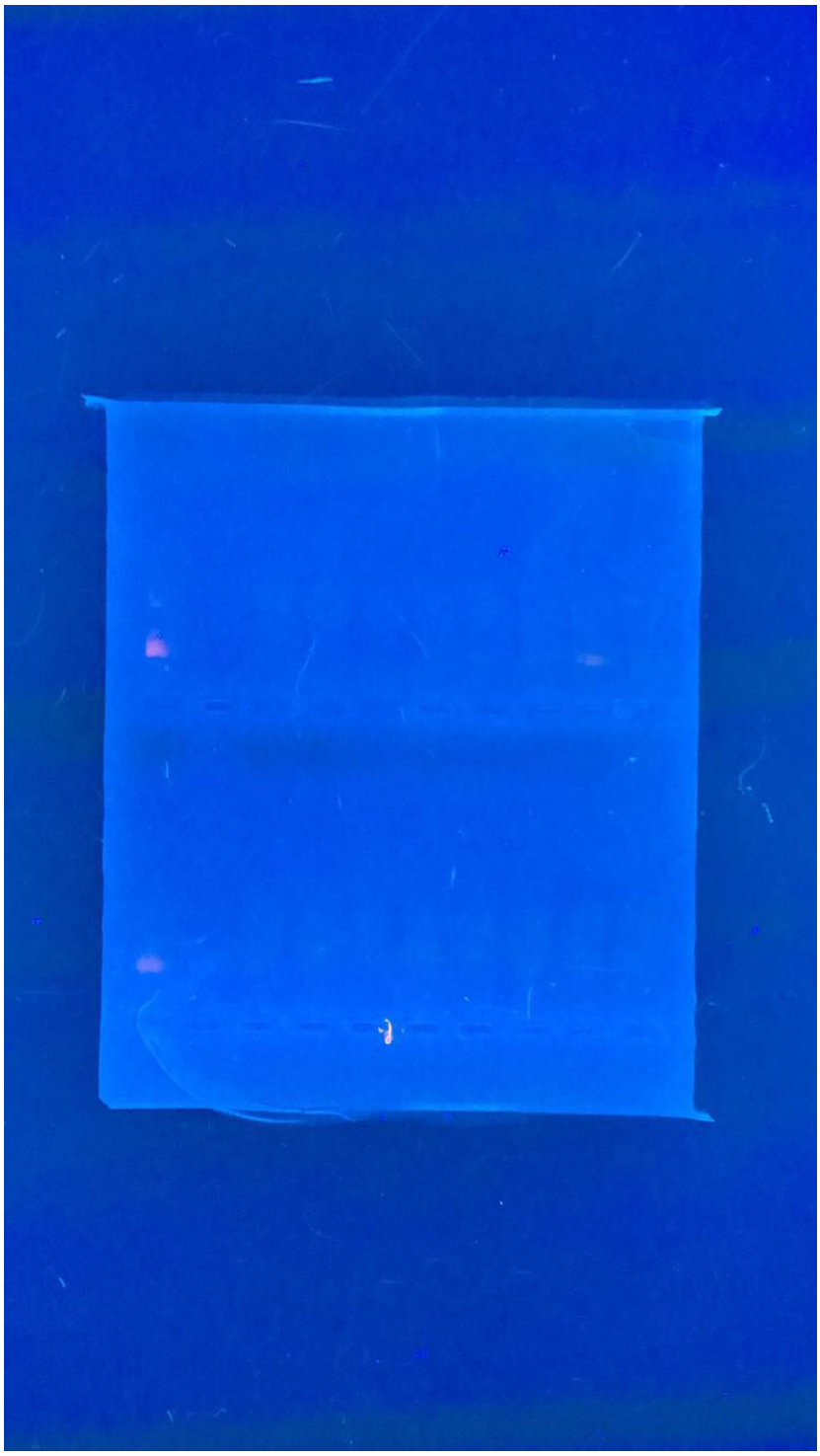
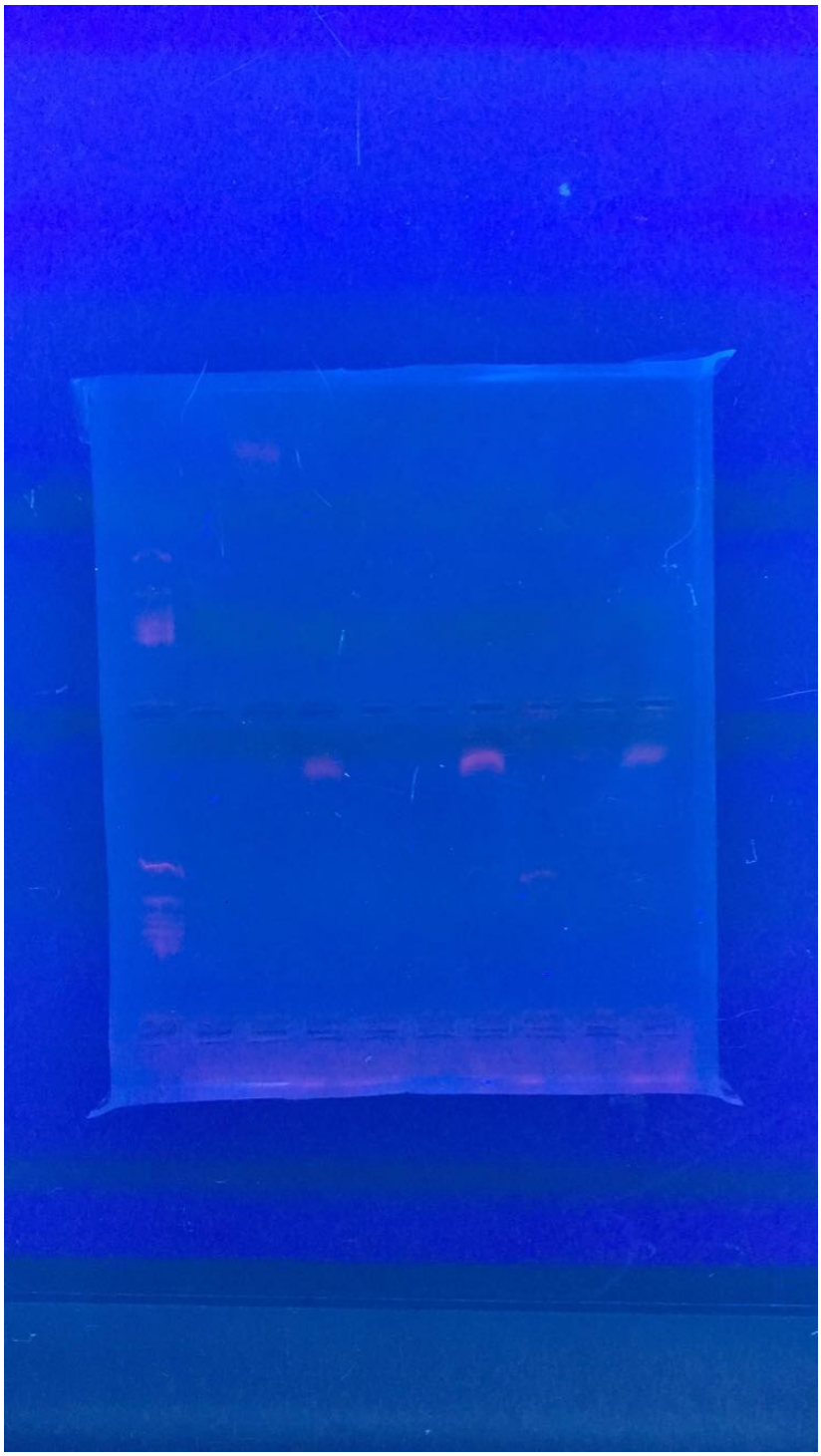


FIGURE 4



DISCUSSION

The experiment was a success in that the use of previous knowledge of DNA combined with modern technology was able to replicate a small sample of DNA into a sample that could be utilized to find a Genus and Species of our supposed "Roadkill". How far innovations have come since the time these processes were performed by hand are extremely vast, but quickly advancing forward. Our species, the American Bison, is extremely interesting because of the bottleneck almost leading to extinction in the 1880's, which was stopped by 5 private ranchers combined efforts (Hedrick, 2009). Even though the extinction was somewhat reversed through efforts, the ecological extinction still occurred within their original grazing range (Freese et al., 2007). This helps to demonstrate that even with species that are less common or have previously been on the verge of extinction, GenBank was preemptive enough to collect and store their DNA strands for over 280,000 species, in order to aide further research (Benson et al., 2014). Some reasons for error in this experiment, also addressed in the results, were the initial problem that very little of the classes samples in figure 3 yielded results, because of this, the initial number and letter combo assigned was changed to a stock solution prepared that was successful in DNA extraction. In order to allow completion of further steps of the lab, however if these errors in electrophoresis were not spotted the whole 6-week process preformed would have been unsuccessful in the database steps of the protocol. Even though a 97% match is high, the 3% could be due to mutations occurred in our sample or perhaps the accidental insertion/deletion in the editing process. It is interesting to consider what mutations could affect GenBank and trick the system into believing it was another species, maybe one that showed common ancestory such as cattle

(Kohl et al., 2013). Overall the success of this lab demonstrates the effectiveness of processes commonly used in practice today.

Works cited

Benson D, Clark K, Karsch-Mizrachi I, Lipman D, Ostell J, Sayers E. GenBank. *Nucleic Acids Research* [serial online]. January 2015;43(Database issue):D30-D35. Available from: MEDLINE Complete, Ipswich, MA. Accessed November 26, 2016.

Freese C, Aune K, Redford K, et al. Second chance for the plains bison. *Biological Conservation* [serial online]. n.d.;136(2):175-184. Available from: Science Citation Index, Ipswich, MA. Accessed November 26, 2016.

Hedrick P. Conservation genetics and North American bison (Bison bison). *The Journal Of Heredity* [serial online]. July 2009;100(4):411-420. Available from: MEDLINE Complete, Ipswich, MA. Accessed November 26, 2016.

Kohl M, Krausman P, Kunkel K, Williams D. Research Articles: Bison Versus Cattle: Are They Ecologically Synonymous?. *Rangeland Ecology & Management* [serial online]. November 1, 2013;66:721-731. Available from: ScienceDirect, Ipswich, MA. Accessed November 26, 2016.

Kürekçi G, Dinçer P. Next-Generation DNA Sequencing Technologies. *Erciyes Medical Journal / Erciyes Tip Dergisi* [serial online]. September 2014;36(3):99-103. Available from: Academic Search Complete, Ipswich, MA. Accessed November 26, 2016.

Lee D, Gorkin D, Beer M, et al. A method to predict the impact of regulatory variants from DNA sequence. *Nature Genetics* [serial online]. August 2015;47(8):955-961. Available from: Academic Search Complete, Ipswich, MA. Accessed November 26, 2016.

Previte M, Zhou C, He M, et al. DNA sequencing using polymerase substrate-binding kinetics. *Nature Communications* [serial online]. January 23, 2015;6:5936. Available from: MEDLINE Complete, Ipswich, MA. Accessed November 26, 2016.

Saba F, Papizadeh M, Fazeli S, et al. A Rapid and Reproducible Genomic DNA Extraction Protocol for Sequence-Based Identification of Archaea, Bacteria, Cyanobacteria, Diatoms, Fungi, and Green Algae. *Journal Of Medical Bacteriology* [serial online]. May 2016;5(3/4):22-28. Available from: Academic Search Ultimate, Ipswich, MA. Accessed November 26, 2016.

Sargsyan O. Analytical Framework for Identifying and Differentiating Recent Hitchhiking and Severe Bottleneck Effects from Multi-Locus DNA Sequence Data. *Plos ONE* [serial online]. May 2012;7(5):1-9. Available from: Food Science Source, Ipswich, MA. Accessed November 26, 2016.

Smith B, Lücking R, Johnston M. From GenBank to GBIF: Phylogeny-based predictive niche modeling tests accuracy of taxonomic identifications in large occurrence data repositories. *Plos ONE* [serial online]. March 1, 2016;11(3)Available from: Scopus®, Ipswich, MA. Accessed November 26, 2016.