Fall 12-13-2017

DNA barcoding of Quercus falcata, Quercus palustris, Quercus rubra, and their hybrids using rbcL, matK, and ycf1

Mckinzie Johnson
Southern Adventist Univeristy

Tim Trott
timtrott@southern.edu

DNA barcoding of *Quercus falcata, Quercus palustris, Quercus rubra,* and their hybrids using rbcL, matK, and ycf1

Mckinzie Johnson

BIOL 496

Fall Semester 2017

Dr. Timothy Trott

Introduction

As of 2010, the estimated number of different plant species is 400,000 (Botanic Gardens Conservation International). With so many species in existence, it is beneficial to have a way to identify and classify each one. Plant species can be identified using several techniques. Traditionally, plants have been identified by their morphological characteristics. More recently, the use of molecular techniques, including the comparison of DNA sequences, has facilitated the comparison of species and individuals at unprecedented levels of detail. DNA barcoding is one method of identifying plants using genetic markers. It involves the comparison of DNA sequences to identify plants down to the species level. DNA barcoding has been used to clarify the taxonomic relatedness between species. For example, this technique has aided in the identification of 490 species of Canadian arctic flora, in addition to various species of orchids and seagrasses (Hollingsworth et al., 2011; Saarela et al., 2013; Lahaye et al., 2008; Lucas et al., 2012). However, some plant species have presented unique challenges to the utilization of this identification method. Specifically, there have been challenges with species that are polyploid, employ various dispersal techniques, and hybridize (Hollingsworth et al., 2011; Piredda et al., 2011; Fazekas et al,. 2009). Additionally, the application of DNA barcoding is still limited because there is not yet a universal DNA barcode for the plant kingdom (Lahaye et al., 2008).

One species that has often presented challenges when using DNA barcoding is the *Quercus* genus. The difficulties identifying *Quercus sp.* via DNA barcoding are often caused by their tendency to hybridize (Piredda et al., 2011). *Quercus* hybrids have been distinguished by their morphology for many years (Gottlieb, 1972; Rushton, 1993). However, the question of whether the hybrids can be identified by their genetic markers in addition to, or in place of, their

morphology is under active investigation. Several studies have sought to find DNA barcodes that work on *Quercus sp.* and have had some success. However, the occurrences of hybridization often caused difficulties (Piredda et al., 2011; Borek, K. & Silvieus, S., 2009). When searching for regions of DNA that would be suitable for use as DNA barcodes, a common technique is to amplify the DNA segments and compare the sequences to others in a database or to other samples of the same species. Typically, Polymerase Chain Reaction (PCR) is used to amplify DNA regions that are then sequenced and analyzed using computer programs such as Clustal Omega to determine their degree of sequence identity and relatedness. Some of the most commonly used primer pairs for DNA barcoding in plants are trnH-psbA, rbcL, matK, and ycf1 (Hollingsworth et al., 2011; Yang et al., 2016). These four primers pairs amplify DNA sequences within the chloroplast genome (See appendix Figure 1). Specifically, trnH-psbA amplifies an intergenic spacer region of the DNA, whereas rbcL, matK, and ycf1 amplify portions of chloroplast genes (Bieniek, Mizianty, & Szklarczyk, 2014; Yang et al., 2016). These barcode markers have been used by themselves or in conjunction with a secondary barcode marker for increased levels of confidence in the accuracy of identification.

The two primers that have had most success with *Quercus sp.* are rbcL and matK (Borek & Silvieus, 2009; Hollingsworth et al., 2011). By using rbcL and matK barcoding markers, it was the goal of this research to determine whether the DNA barcoding could be used to identify and distinguish between the indigenous red oak species and their hybrids on White Oak Mountain in Southeastern Tennessee. We were able to successfully extract DNA from the species being studied. However, further research will need to be conducted to come to a

conclusion regarding the level of relatedness between non-hybrid and hybridized oaks based on sequence comparisons.

Methods and Materials

Sample Collection

Several oak species were collected. The three oaks of interest were the northern red oak, southern red oak, and pin oak. Leaves were also collected from a couple of black oak trees and a willow oak for future use and comparison. Northern red oak samples were identified primarily by leaf morphology (Figure 2) focussing on the following characteristics: a wide middle, shallow sinuses, and glabrous surface with the exception of slight hair in the axils of the veins on the underside of the leaf. Southern red oak samples were identified and collected based on their varying shapes with most of the leaves having five to seven narrow and deep lobes with bell-shaped bases (Figure 2). Pin oak samples were identified by their long, elliptical-oblong leaf shape, generally showing five to seven lobes that were smooth except for slight hairs at the axils of the veins (Kirkman et al., 2007).

Leaf material from three red oak species (northern red oak, southern red oak, and pin oak) was collected for this experiment from trees along the White Oak Mountain and Bauxite trail systems on Southern Adventist University campus. Leaf samples from three suspected oak hybrids were also collected. These hybrids were identified primarily by their indeterminate leaf morphologies showing characteristics intermediate between two of the main red oak species. Each tree was then photographed, its GPS coordinates were recorded, and the leaves collected were assigned a number. The data from the collections were compiled into an Airtable document

(see appendix Table 1). The collected leaves were then rinsed with reverse osmosis water and stored at -80°C in 50 ml centrifuge tubes. The following spring, some of the trees were revisited in order to harvest younger leaves in the hopes that they might provide greater yields of non-contaminated DNA. The leaves were cleaned and stored in the same manner as the previous collection. These samples were added to the Airtable document in an adjacent column under a different date for the same sample (see appendix Table 2).

DNA Extraction

DNA was extracted using a modified procedure based on the PowerPlant Pro® DNA Isolation Kit from MoBIO (13400-50). Leaf tissue was ground under N_2 (liq) and transferred to a 1.7 ml microcentrifuge tube. Tissue was further homogenized in a 2 ml PowerPlant® Bead Tube placed in a Disruptor Genie (Scientific Industries SI-DD38) for 10 minutes. The tube contained PD1 solution, phenolic separation solution, PD2 solution, and RNase A. Once the tissue was evenly homogenized, it was processed according to the PowerPlant Pro® DNA Isolation Kit instructions. The DNA samples were stored at -20°C until used.

PCR Amplification

PCR reactions were optimized for 0.5 µl of the template DNA as indicated by single PCR products after agarose electrophoresis. PCR reactions were performed for both rbcL and matK primer sets. The PCR amplification for both primers were performed in a mixture containing 2 µl of 0.2 µMoles of forward and reverse primers, 25 µl One-Taq Quick-Load 2X Master Mix

(BIOLABS - M0486S) in Standard Buffer, and 2 µl of DNA. The final volume was adjusted to 50 µl using distilled water.

The targeted gene for the DNA barcode was amplified using BIO-RAD C1000 Touch™ Thermal Cycler following the PCR profiles for each primer set. The rbcL primers PCR cycling profile began at 95°C for 4 minutes, followed by 35 cycles of 30 seconds at 94°C, 1 minute at 55°C, 1 minute at 68°C, ended with 10 minutes at 68°C (CBOL plant working group, 2009). The matK primers PCR cycling profile began at 94°C for 1 minute, followed by 35 cycles of 30 seconds at 94°C, 20 seconds at 52°C, 50 seconds at 68°C, and ended with 5 minutes at 68°C (CBOL plant working group, 2009). The number of cycles at 94°C was later increased to 45 cycles for both the rbcL and the matk profiles.

The PCR product was purified prior to sequencing using MonarchTM PCR & DNA Cleanup Kit. The final elution was adjusted to 20 µl with the DNA Elution Buffer.

PCR products were analyzed on 1% (w/v) agarose gels in TAE. Products were visualized by ethidium bromide staining or UViewTM 6x loading dye and imaged with an UVP EC3 imaging system. PCR reactions containing single clear bands were sent off for DNA sequencing by Macrogen (www.macrogenusa.com). The primers matK and rbcL were used for the sequencing reactions.

The successful DNA from the previous steps were amplified using the ycf1 primer. This primer followed the PCR cycling profile of 94°C for 4 minutes, followed by 45 cycles of 30 seconds at 94°C, 40 seconds at 52°C, 1 minute at 68°C, and ended with 10 minutes at 68°C (Dong et al., 2015).

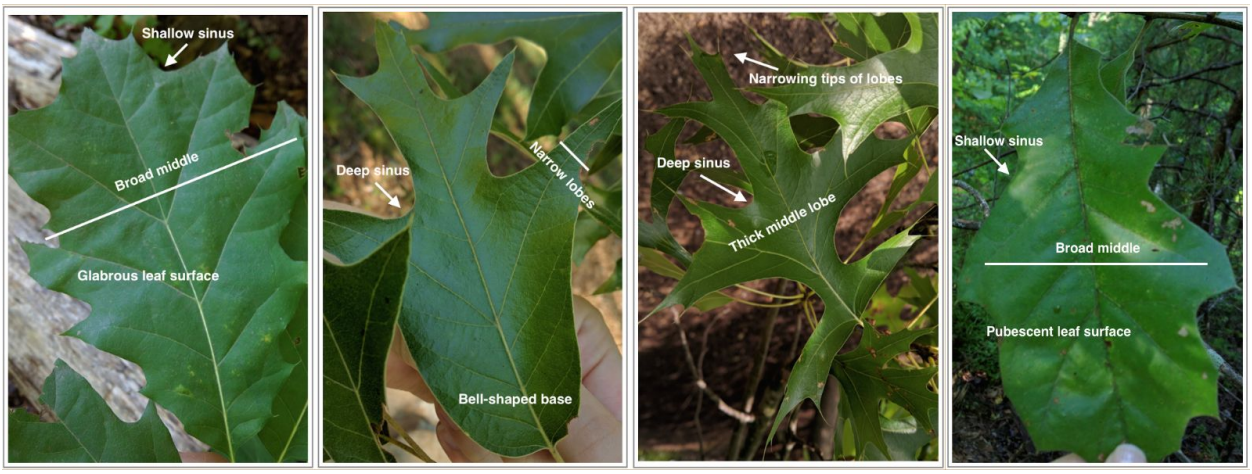
Sequence Analysis

The sequences were compared to reference sequences available at the NCBI (National Center for Biotechnology Information) (http://www.ncbi.nlm.nih.gov/genbank) and the sequence data were aligned and analyzed using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo) to test the reliability of the DNA barcodes. The sequences for *Q. rubra* were compared to reference sequences for rbcL, matK, and ycf1 (KP088816, AB125043, and KP088376.1 respectively) and to each other's sequences. Similarly, the sequences for *Q. falcata* oaks were compared to reference sequences for rbcL and matK (KJ773801 and KJ773058 respectively) and to each other's sequences. Lastly, the sequences for *Q. palustris* oaks were compared to reference sequences for ycf1 (KP088372.1) and to each other's sequences. The hybrids were also compared to each other, the non-hybrids, and the reference sequences to determine and quantitate the relatedness to their parental oak species.

Results and Discussion

Leaf identification:

Leaves were identified based on characteristics found in a field guide for plants native to Tennessee. The leaves identified as hybrids had a mixture of characteristics of two or more of the species of interest. Hybrids were generally found to have shallow sinuses and the broad middle of the northern red oak leaves, and also rusty colored hairs covering both the top and bottom surfaces as seen in southern red oak species. Based on these combined characteristics and the lack of a better morphologically matched species, it was concluded that trees with this leaf description were hybrids.



Northern red oak *(Quercus rubra)* | Southern red oak *(Quercus falcata)* | Pin oak *(Quercus palustris)* | Hybrid

Figure 2. Leaf Identification Photos. The top far left photo is of a northern red oak. It was described as being the broadest in the middle, having shallow sinuses, and a glabrous surface on both the top and bottom of the leaf, with the exception of small hairs in the axial of the veins on the underside of the leaf. The upper middle photo is of a southern red oak. It was identified based on the unique leaf shape with the bell-shaped base, narrowing lobes, deep sinuses, and the rust colored hairs on the underside of the leaf. The third photo from the left is of a pin oak. It was identified using the leaf characteristics of the extra thick middle lobe pair, deep sinuses, narrowing tips of the lobes, and smooth to glabrous surface except at the axials of the veins. Generally, there are five to seven lobes in a pin oak leaf. The far right photo is of a leaf identified as a hybrid due to the mixture of characteristics of two of the previously mentioned species. It was found to have shallow sinuses and a broad middle portion of the leaf similar to the northern red oak, and the rust colored hairs of the southern red oak.

DNA Extraction

To test if the DNA was extracted successfully, PCR products were visualized on agarose gels. Initially, some of the PCR reactions did not produce a visible band after staining and imaging. The success of the amplifications may have been hindered by the time of year that the leaf samples were gathered. In the summer, oaks increase the concentration of tannins in their leaves in an effort to ward off insects (Feeny, 1970). These tannins include many phenolic

compounds that inhibit the binding of the primer in PCR (Manoj et al., 2007). Since the leaves were collected in July and August, the concentration of these compounds would have been high. It was suspected that the strong phenolic compounds in the oak leaves might have been interfering with the PCR (Azmat et al., 2012). To account for this, the phenolic separation solution from the DNA extraction kit was added to the rest of the tissue samples. The first two samples were amplified before adding this extra step. In the future, it may be useful to add Polyvinylpyrrolidone (PVP) or diethyl ether (Simon, Gray, & Cook, 1996; Feeny, 1970) to our DNA extraction procedure to remove these PCR inhibitors from our extracted DNA. Seventeen of the twenty-five leaf samples were successfully extracted using this modified procedure.

PCR Amplification

After analyzing the PCR product on the agarose gels, it was clear that not all of the samples were producing a single strong PCR product. The rbcL primer set successfully amplified 17 out of 25 samples (six southern red oaks, six northern red oaks, two black oaks, two pin oaks, and one hybrid) (Figure 1). The matK primer set successfully amplified 11 out of 25 samples (six southern red oaks, three northern red oaks, one pin oak, and one hybrid) (Figure 2). The ycf1 primer set successfully amplified 10 out of 17 samples (three southern red oaks, two northern red oaks, two black oaks, one pin oak, one willow oak, and one hybrid) (Figures 3).

The PCR product run on the gel varied in length. The rbcL primers produced fragments between 633 and 500 base pairs (bp). The matK primers produced fragments of about 922 bp. The ycf1 primers produced fragments of about 883 bp. It might have been useful to store the

samples in smaller amounts and use them sooner to avoid repeatedly freezing and thawing the sample, which could shear the DNA (Shao et al., 2012).

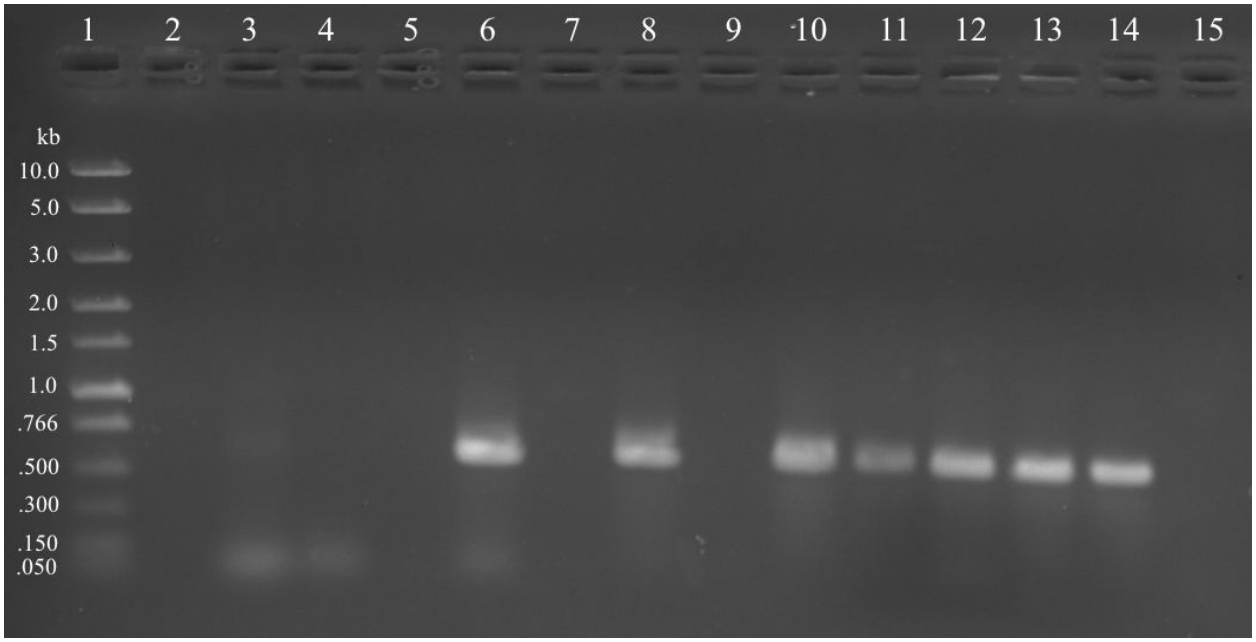


Figure 3. 1% Agarose gel of rbcL PCR fragments. The gel was run for an hour and forty minutes first at a voltage of 30V for ten minutes, then at 70V for an 1.5 hours. Lane one contains the DNA Fast Ladder used as a reference. Lane two is a no primer control, lane three is a no template control. No PCR product was detectable in lanes four, five, seven, nine, and fifteen after staining and imaging.

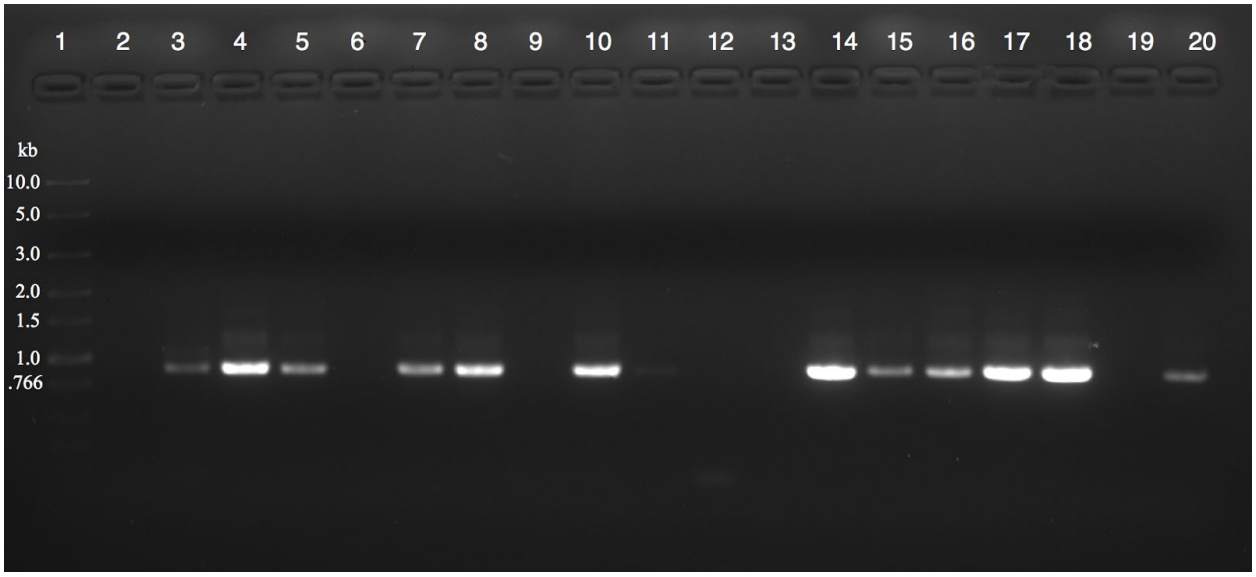


Figure 4. 1% Agarose gel of matK PCR fragments. The gel was run for an hour and forty minutes first at a voltage of 30V for ten minutes, then at 70V for an 1.5 hours. Lane one contains the DNA Fast Ladder used as a reference. Lane two is a no primer control, lane three is a no template control. No PCR product was detectable in lanes six, nine, twelve, thirteen, and nineteen after staining and imaging.

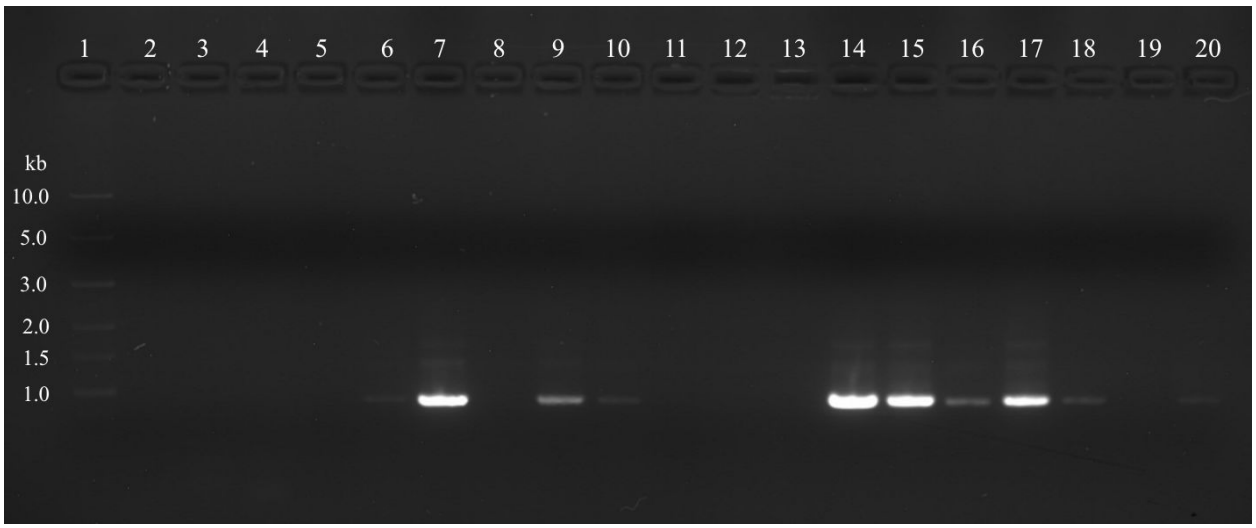


Figure 5. **1% Agarose gel of ycf1 PCR fragments**. The gel was run for an hour and forty minutes first at a voltage of 30V for ten minutes, then at 70V for an 1.5 hours. Lane one contains the DNA Fast Ladder used as a reference. Lane two is a no primer control, lane three is a no template control. No PCR product was detectable in lanes four, five, eight, eleven-thirteen, and nineteen after staining and imaging. Lanes five, eight, nine, twelve, nineteen, and twenty contain DNA from the younger leaves. The rest of the lanes contain DNA from the older leaf samples.

In total, 68% of the DNA samples processed produced a detectable PCR product when visualized on a 1% agarose gel. The different primers had different success rates of amplification. The rbcL primers had 68% successfully visualized PCR product while the matK had 57.9% and the ycf1 had 71.4%. We initially hypothesized that the increased success with ycf1 primers was due to our use of smaller, more immature leaves collected earlier in the spring as these trials were done after the second round of sample collections taken shortly after leaf emergence. Similar work by Bilbrey and Schiebout had previously shown increased success amplifying DNA from younger leaves (2017). However, further investigations concluded that in our project only two out of six samples taken from younger leaves produced a visualized fragment after electrophoresis of PCR products. In addition, we had greater success with tissue collected from older leaves. These results suggested that the issue was with the extraction and amplification process we were using and not with the age of the tissue samples at the time of collection.

Contamination issues and DNA sequencing

Analysis of the electropherograms from the initial sequence results indicated that the PCR products may be contaminated. Contamination was suspected due to the presence of overlapping signals on electropherograms (Figure 6). These signals seem to suggest that other DNA was in the sample contaminating the target sequence, or the PCR primers could have been amplifying sequences for multiple diverse species. The same problem with DNA amplification has occurred in other research with oak species, suggesting that there is an issue with the PCR process or the sample itself (Borek & Silvieus, 2009). In Borek and Silvieus' research, there

were issues of the PCR fragments not being long enough to analyze after sequencing (2009). Borek and Silvieus were obtaining fragments of around 400 bp. Our results ranged from 150 bp to nearly 900 bp. Borek and Silvieus also found that when attempting to amplify certain oak species again, there was little success. They hypothesized that there may have been an issue with the PCR process or that the oaks contained a compound that caused interference (2009). This issue was not resolved in their research. More research needed to be done in this area, considering Borek and Silvieus used a sample size of eight oak trees.

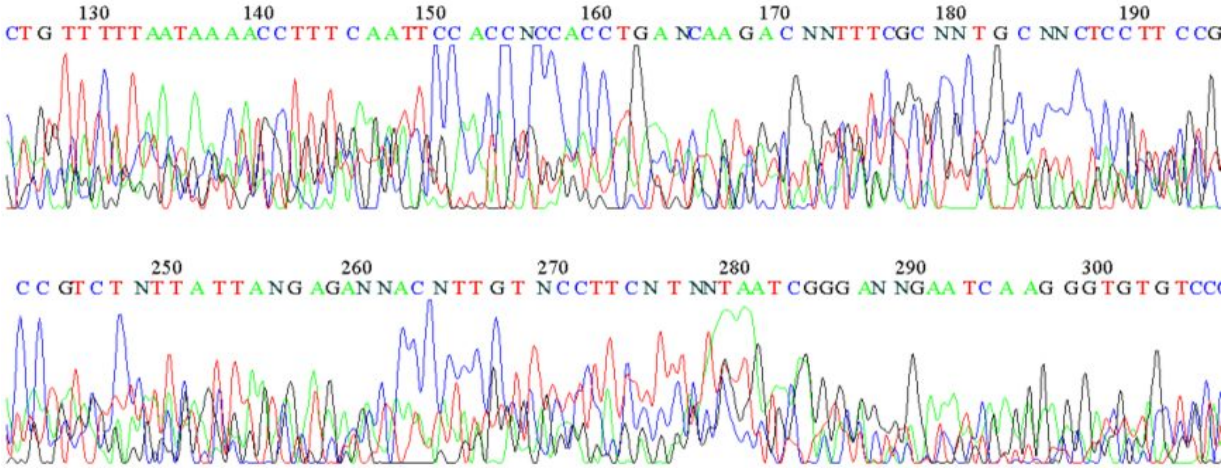


Figure 6. Electropherogram showing multiple overlapping signals. Typical segment produced from a sequence analysis of a southern red oak (sample 24) sequenced with rbcL forward primer.

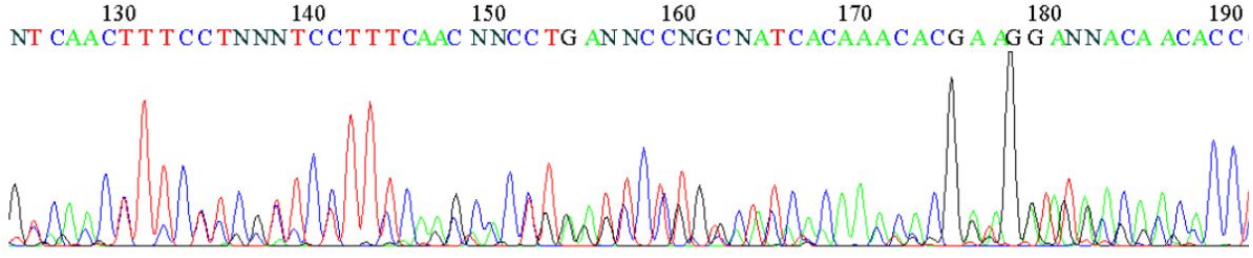


Figure 7. Electropherogram showing single, non-overlapping, signals for each nucleotide. Example of a segment of the PCR product sequence analysis results for a northern red oak (sample 3) sequenced with rbcL reverse primer.

In attempting to limit further contamination to the samples, the PCR preps were done under a hood with micropipettes that were sterilized via autoclaving and under UV light in addition to using filter pipette tips. However, the later sequence electropherograms showed only slight improvement. Emma Bilbrey, a student conducting a similar project at Union University in Tennessee, ran into the same issue and had some success after collecting younger leaves (2017). This success was in part due to the lower concentrations of phenols. After discussions with Emma Bilbrey, we decided to try the procedure on younger leaves. In the following weeks, younger leaves were collected from six of the previously sampled trees. Unfortunately, our results did not seem to have the same success that Ms. Bilbrey had. The PCR product from the younger leaves did not appear to be any less contaminated or overlapping than the previous DNA samples.

In an effort to clean up the sequences after amplification, a PCR cleanup kit was used. The kit's purpose was to separate the desired DNA fragments from short sequences, nucleotides, salts, and excess primers that would be in the PCR product ("Whatman," 2001). However, the sequences compared after the cleanup kit was used did not appear to be significantly improved. This may have been due to the fact that the extraneous material left in with the PCR product was not the source of the noisy electropherogram.

Sequence alignments and analysis

DNA sequences were compared to reference sequences obtained from GenBank to determine if they were valid for identification of these oaks to the species level. In previous research, a difference in the DNA sequences of less than 5% was seen as insignificant (Borek &

Silvieus, 2009). For our sequences, the alignments between sample oak sequences and the corresponding GenBank reference sequence for each species gave percent identities ranging from 29.14 - 53.61%. Alignments were performed on sequences that appeared to be useful, indicated by non-overlapping signals in the electropherograms. The differences between samples of the same species was much larger than expected. This result corresponds with the hypothesis that the sample was contaminated, giving low percent identity between samples of the same species.

It was predicted that the sequence comparisons between samples from members of the same species would have significantly higher percent identity to each other than to sequences of other species. After alignment, rbcL sequences for northern red oaks showed low percent identity to the northern red oak reference sequence and to sequences from other northern red oak samples, as seen in table 3. Additionally, when comparing the northern red oak samples to a southern red oak sample sequence, the percent identities were similar to the ones between northern red oak species (Table 3).

Table 3. Percent identity matrix showing low percent identity between same and different species using rbcL. DNA sequences of northern red oaks aligned with a southern red oak (sample 2) and a northern red oak reference sequence from GenBank (accession number: AB125026.1).

	Northern red oak (#6)	Northern red oak (#12)	Southern red oak (#2)
GenBank Reference	45.17	42.59	42.32
Northern red oak (#6)		47.95	47.89
Northern red oak (#12)			41.40

Similarly, the sequences from matK primers showed a similar pattern of low percent identity when comparing between samples. The percent identity of our northern red oak sample #12 compared to the reference sequence from GenBank was only 0.27%. Additionally, when

comparing between the northern red oak sample #6 and sample #12, the sequence was only 0.06% more identical than between the northern red oak sample #6 and the southern red oak sample #2 (Table 4). The low intraspecies percent identity was another indication that there was an issue with the procedure.

Table 4. Percent identity matrix showing low percent identity between same and different species using matK. A. DNA sequences of southern red oaks aligned with a northern red oak (sample 18) and a southern red oak reference sequence (accession number: KJ773058.1). B. DNA sequences of northern red oak aligned with a southern red oak (sample 9) and a northern red oak reference sequence (accession number: AB125043.1). Reference sequences obtained from GenBank.

A	Southern red oak (#21)	Southern red oak (#9)	Northern red oak (#18)
Southern red oak Reference	40.03	51.31	43.72
Southern red oak (#21)		42.29	34.62
Southern red oak (#9)			36.61
B	**Northern red oak (#18)**	**Northern red oak (#12)**	**Southern red oak (#9)**
Northern red oak Reference	41.16	42.69	46.32
Northern red oak (#18)		36.09	36.61
Northern red oak (#12)			46.79

The ycf1 had several electropherograms that were clearer than the ones for the rbcL and matK primers. However, the percent identity between the northern red oak species was lower than the similarity between the northern red oak and the southern red oak species. The comparison between the northern red oak sample #6 and the southern red oak sample #9 was 12.21% higher than the comparison between the northern red oak sample #6 with the northern red oak sample #18. The trend of higher percent identity continued when comparing the northern red oak sample #18 with the southern red oak sample #9. Additionally, the southern red oak

sample #9 showed more similarity to the GenBank reference sequence then either of the two northern red oak samples (Table 5).

Table 5. Percent identity matrix showing low percent identity between same and different species using ycf1. DNA sequences of northern red oaks aligned with a southern red oak (sample 2) and a northern red oak reference sequence from GenBank (accession number: KP088376.1).

	Northern red oak (#6)	Northern red oak (#18)	Southern red oak (#9)
GenBank Reference	40.94	42.59	43.97
Northern red oak (#6)		40.71	52.92
Northern red oak (#18)			41.29

The low similarity between the sample sequence and the reference sequence was most likely due to the contamination problem encountered previously. In an attempt to improve this low percentage, some of the electropherograms were manually analyzed to determine the identity of uncalled/unclear nucleotides. This manual process resulted in an increase of alignment identity by only 1.49 - 3.53%.

Conclusion and Future Research

For this project, DNA barcoding with the three primers (rbcL, matK, and ycf1) did not yield enough differences to provide accurate identification of *Quercus* species. Although, ycf1 had the most success for amplification and rbcL appeared to give the most consistent results for sequence alignments. However, the inaccuracies and minimal differences appeared to be due to issues with the process and methods that can be dealt with in future research. For example, the success during DNA extraction may be improved by using a different DNA extraction kit. The issue with PCR amplification may be resolved with younger leaves being collected and

immediately processed. Additional optimization of PCR conditions including modifying the primer annealing temperature and testing different polymerases from other vendors may also increase the specificity of the PCR products. Once these technical challenges are overcome the main questions regarding the ability of DNA barcoding to determine the paternity of Oak hybrids may be directly examined.

References

Azmat, M. A., Khan, I. A., Cheema, H. M. N., Rajwana, I. A., Khan, A. S., & Khan, A. A. (2012). Extraction of DNA suitable for PCR applications from mature leaves of *Mangifera indica* L. *Journal of Zhejiang University Science B*, *13*(4), 239–243. http://doi.org/10.1631/jzus.B1100194

Bieniek, W., Mizianty, M., & Szklarczyk, M. (2014). Sequence variation at the three chloroplast loci (matK, rbcL, trnH-psbA) in the Triticeae tribe (*Poaceae*): comments on the relationship and utility in DNA barcoding of selected species. *Plant Systematics and Evolution, 301*(4), 1275-1286.

Bilbrey, E., & Schiebout, M. DNA barcoding of *Quercus margaretta* from TN to validate morphological identification and to determine efficacy of different barcoding regions to determine *Quercus* species resolution. Poster presented at: 78th Association of Southeastern Biologists Annual Conference; 2017, Mar. 29 - Apr. 1; Montgomery, AL.

Borek, K., & Silvieus, S. (2009). DNA barcoding of quercus sp. at pierce cedar creek institute using the matK gene. Aquinas College.

Botanic Gardens Conservation International. (n.d.). Plant Species Numbers. Retrieved March 16, 2016, from https://www.bgci.org/policy/1521/

CBOL plant working group. (2009, November 16). Retrieved from http://barcoding.si.edu/plant_working_group.html.

Dong, W., Xu, C., Li, C., Sun, J., Zuo, Y., Shi, S., … Zhou, S. (2015). *ycf1*, the most promising plastid DNA barcode of land plants. *Scientific Reports, 5*(8348). doi: 10.1038/srep08348

Fazekas. A.J., Kesanakurti, P.R., Burgess, K.S., Percy, D.M., Graham, S.W., Barrett, S.C.H., … & Husband, B.C. (2009) Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Molecular Ecology Resources, 9*, 130–139.

Feeny, P. (1970) Seasonal changes in oak leaf tannins and nutrients as a cause of spring feeding by winter moth caterpillars. *Ecology, 51*(4). http://dx.doi.org/10.2307/1934037

Gottlieb, L.D. (1972). Levels of confidence in the analysis of hybridization in plants. *Annals of the Missouri Botanical Garden 59*, 435-446. doi: 10.2307/2395153

Hollingsworth, P. M., Graham, S. W., & Little, D. P. (2011). Choosing and using a plant DNA barcode. *Plos ONE*, *6*(5), 1-13. doi:10.1371/journal.pone.0019254

Kirkman, L. K., Brown, C. L., & Leopold, D. J. (2007). *Native trees of the Southeast: An identification guide*. Portland, OR: Timber Press.

Lahaye, R. van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., & Savolainen, V. (2008). DNA barcoding the floras of biodiversity hotspots. *PNAS, 105*(8), 2923-2928. www.pnas.org/cgi/doi/10.1073/pnas.0709936105

Lucas, C., Thangaradjou, T., & Papenbrock, J. (2012). Development of a DNA barcoding System for seagrasses: Successful but not simple. *Plos ONE, 7*(1), e29987. https://doi.org/10.1371/journal.pone.0029987

Manoj, K., Tushar, B., Sushama, C., 2007. Isolation and purification of genomic DNA from Black Plum (Eugenia jambolana Lam.) for analytical applications. Int. J. Biotechnol. Biochem., 3(1):49-55.

Piredda, R., Simeone, M. C., Attimonelli, M., Bellarosa, R., & Schirone, B. (2011). Prospects of barcoding the Italian wild dendroflora: Oaks reveal severe limitations to tracking species

identity. *Molecular Ecology Resources, 11,* 72-83. doi:10.1111/j.1755-0998.2010.02900.x

Rushton, B.S. (1993). Natural Hybridization within the genus *Quercus L. Annals of Forest Science 50.* doi:10.1051/forest:19930707

Saarela, J., Sokoloff, P., Gillespie, L., Consaul, L., & Bull, R. (2013). DNA barcoding the Canadian arctic flora: Core plastid barcodes (rbcL plus matK) for 490 vascular plant species. *Plos ONE*, *8*(10).

Shao, W., Khin, S., & Kopp, W. C. (2012). Characterization of effect of repeated freeze and thaw cycles on stability of genomic DNA using pulsed field gel electrophoresis. *Biopreservation and Biobanking, 10*(1), 4-11. doi: 10.1089/bio.2011.0016

Whatman, Inc.: PCR Clean-Up Kit. (2001). *Biomedical Products*, *26*(11), 34. Retrieved from http://link.galegroup.com.ezproxy.southern.edu/apps/doc/A81104327/ITOF?u=tel_a_sau&sid=ITOF&xid=efdaf945

Yang, Y., Zhou, T., Duan, D., Yang, J., Feng, L., & Zhao, G. (2016). Comparative Analysis of the Complete Chloroplast Genomes of Five Quercus Species. *Frontiers in Plant Science, 7*(959). doi: 10.3389/fpls.2016.00959

Appendix

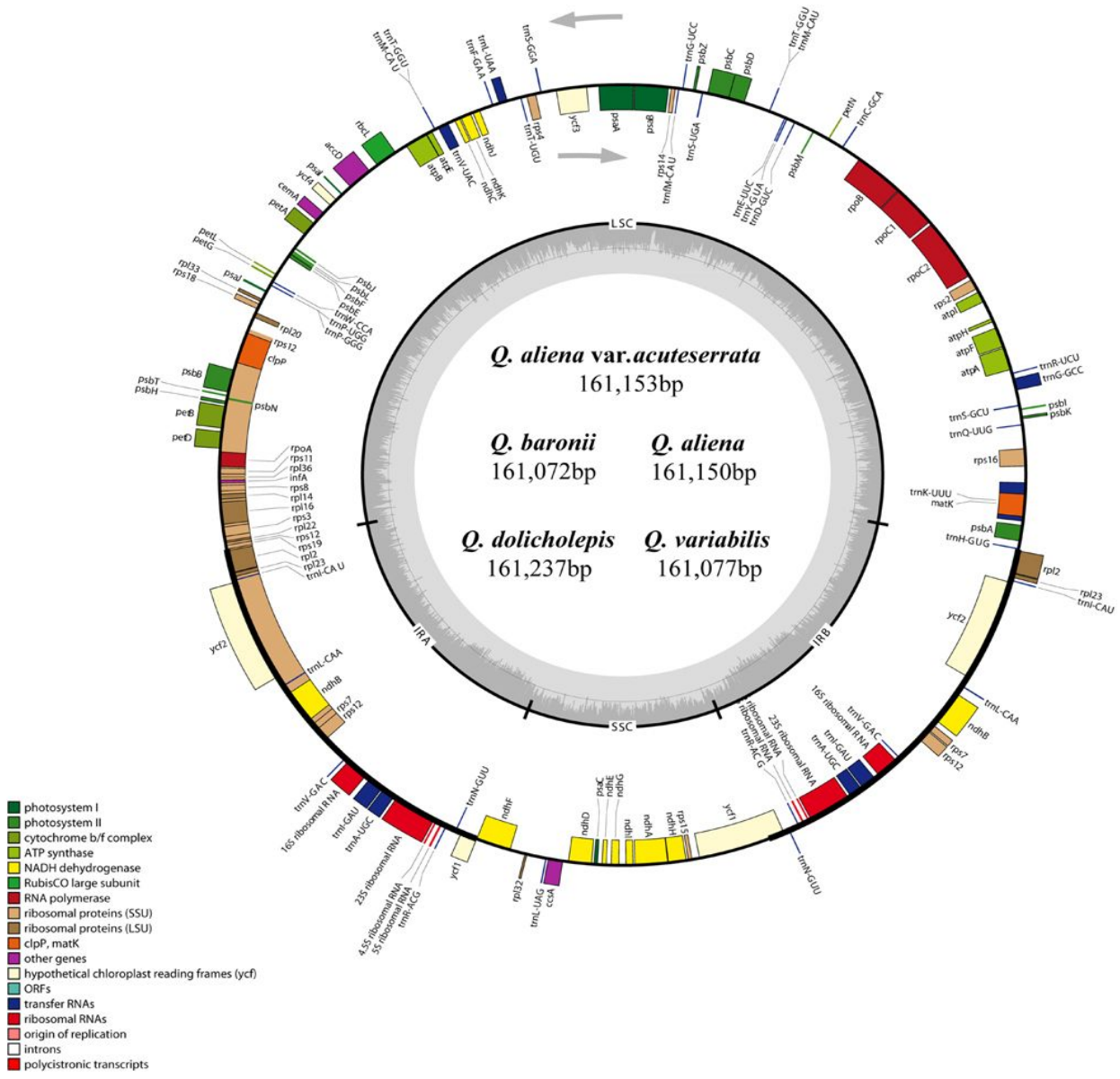


Figure 1. General chloroplast genome schematic for quercus species. (Yang et al., 2016).

Table 1. Collection information for all oak samples collected in the summer. Each tree was assigned a number and each leaf from the sample was assigned a sub-number (e.g. 186.1, 186.2, etc.). The GPS coordinates and altitude were obtained using the app Locus Map. The data was compiled on an Airtable spreadsheet (https://airtable.com/shrRbQiCb9WKLY0JR).

Sample Identification	Common Name	Scientific Name	GPS Coordinates	Date Collected	Amount Collected
186.1-5	Northern red oak	*Quercus rubra*	35.04499° N, 85.04249° W, Altitude: 249 m	7/18/16	5
01.1-5	Pin oak	*Quercus palustris*	35° 2' 25.800" N, 85° 3' 7.785" W, Altitude: 239 m	7/18/16	5
02.1-5	Southern red oak	*Quercus falcata*	35° 2' 25.089" N, 85° 3' 6.727" W, Altitude: 206 m	7/18/16	5
03.1-5	Northern red oak	*Quercus rubra*	35° 01.621' N, 85° 16.775' W	7/22/16	5
04.1-5	Willow oak	*Quercus phellos*	35° 01.621' N, 85° 16.775' W	7/22/16	5
05.1-5	Pin oak	*Quercus palustris*	35° 01.621' N, 85° 16.775' W	7/22/16	5
06.1-5	Northern red oak	*Quercus rubra*	35° 03.402' N, 85° 02.834' W, Altitude: 792 ft	7/26/16	5
07.1-5	Northern red oak	*Quercus rubra*	35° 03.402' N, 85° 02.834' W, Altitude: 792 ft	7/26/16	5
08.1-5	Southern red oak	*Quercus falcata*	35° 03.402' N, 85° 02.834' W, Altitude: 792 ft	7/26/16	5
09.1-5	Southern red oak	*Quercus falcata*	35° 03.402' N, 85° 02.834' W, Altitude: 792 ft	8/1/16	5
10.1-5	Black oak	*Quercus velutina*	35° 03.402' N, 85° 02.834' W, Altitude: 792 ft	8/1/16	5
11.1-5	Hybrid		35° 03.402' N, 85° 02.834' W, Altitude: 792 ft	8/1/16	5
12.1-5	Northern red oak	*Quercus rubra*	35° 03.402' N, 85° 02.834' W, Altitude: 792 ft	8/1/16	5

13.1-5	Northern red oak	*Quercus rubra*	35° 03.009' N, 85° 03.342' W, Altitude: ?	8/1/16	3
14.1-3	Southern red oak	*Quercus falcata*	35° 03.009' N, 85° 03.340' W, Altitude: 1019 ft	8/3/16	3
15.1-3	Black oak	*Quercus velutina*	35° 03.009' N, 85° 03.340' W, Altitude: 1019 ft	8/3/16	3
16.1-3	Southern red oak	*Quercus falcata*	35° 03.009' N, 85° 03.340' W, Altitude: 1019 ft	8/17/16	3
17.1-3	Southern red oak	*Quercus falcata*	35° 03.009' N, 85° 03.340' W, Altitude: 1019 ft	8/17/16	3
18.1-3	Northern red oak	*Quercus rubra*	35° 03.009' N, 85° 03.340' W, Altitude: 1019 ft	8/17/16	3
19.1-3	Hybrid		35° 03.009' N, 85° 03.340' W, Altitude: 1019 ft	8/17/16	3
20.1-3	Hybrid		35° 02.983' N, 85° 03.607' W, Altitude: 897 ft	8/17/16	3
21.1-3	Southern red oak	*Quercus falcata*	35° 02.983' N, 85° 03.607' W, Altitude: 897 ft	8/17/16	3
22.1-3	Pin oak	*Quercus palustris*	35° 02.983' N, 85° 03.607' W, Altitude: 897 ft	8/17/16	3
23.1-3	Pin oak	*Quercus palustris*	35° 02.983' N, 85° 03.607' W, Altitude: 897 ft	8/17/16	3
24.1-3	Pin oak	*Quercus palustris*	35° 02.983' N, 85° 03.607' W, Altitude: 897 ft	8/17/16	3

Table 2. Collection information for all samples collected in the spring. Each leaf was assigned a number using the tree's number and then a sample number 1-3. The GPS coordinates and altitude were obtained using the app Locus Map. The data was compiled on the same Airtable spreadsheet as the previous samples (https://airtable.com/shrRbQiCb9WKLY0JR).

Sample Identification	Common Name	Scientific Name	GPS Coordinates	Date Collected	Amount Collected
186.1-3	Northern red oak	*Quercus rubra*	35.04499° N, 85.04249° W, Altitude: 249 m	7/18/16	3
03.1-3	Northern red oak	*Quercus rubra*	35° 01.621' N, 85° 16.775' W	7/22/16	3
08.1-3	Southern red oak	*Quercus falcata*	35° 03.402' N, 85° 02.834' W, Altitude: 792 ft	7/26/16	3
09.1-3	Southern red oak	*Quercus falcata*	35° 03.402' N, 85° 02.834' W, Altitude: 792 ft	8/1/16	3
13.1-3	Northern red oak	*Quercus rubra*	35° 03.009' N, 85° 03.342' W, Altitude: ?	8/1/16	3
22.1-3	Pin oak	*Quercus palustris*	35° 02.983' N, 85° 03.607' W, Altitude: 897 ft	8/17/16	3
23.1-3	Pin oak	*Quercus palustris*	35° 02.983' N, 85° 03.607' W, Altitude: 897 ft	8/17/16	3
24.1-3	Pin oak	*Quercus palustris*	35° 02.983' N, 85° 03.607' W, Altitude: 897 ft	8/17/16	3