3-2017

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

Mckinzie Johnson
Southern Adventist University

Timothy Trott
Southern Adventist University

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

SOUTHERN ADVENTIST UNIVERSITY
Biology Department

Mckinzie Johnson and Timothy Trott, Ph.D.
Hickman Science Center, Southern Adventist University, Collegedale, TN

SOUTHERN ADVENTIST UNIVERSITY
Biology Department

Background

It is estimated that there are over 400,000 plant species on earth (Botanic Gardens Conservation International). Species are identified by their morphology as well as genetic identifiers such as DNA barcoding. DNA barcoding typically involves amplification of DNA sequences and subsequent comparison to reference sequences from a database or other samples of the same species.

These techniques help to clarify taxonomy and relatedness between species. However, certain species often present difficulties when utilizing DNA barcoding due to polyploidy, dispersal ability, and hybridization (Hollingsworth et al., 2011; Piredda et al., 2011; Fazekas et al,. 2009).

Oaks (*Quercus sp.*) are one genus that has presented difficulty in DNA barcoding due to hybridization (Piredda et al., 2011). Some studies have found a measure of success in barcoding oaks with primers such as rbcL, matK, and ycf1 (Hollingsworth et al., 2011; Dong et al., 2015).

Project Goal

By using the matK, rbcL, and ycf1 barcoding markers, it is the goal of this research to determine whether the DNA barcodes can be used to identifying red oak species indigenous to White Oak Mountain in Southeastern Tennessee. If this proves successful, it will be determined if the parental identity of a red oak hybrid can be clarified based upon its DNA barcode.

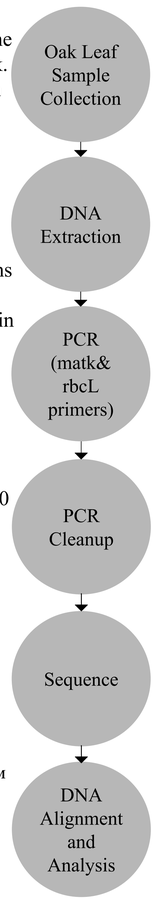
Methods and Materials (*Overall project workflow shown to the right*)

Oak Identification

Northern red oak samples were identified by the leaves being widest in the middle, their shallow lobes, and the 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 narrow and deep, 5 to 7 lobes with bell-shaped bases. Pin oak samples were identified by their long, elliptical-oblong leaf shape. There are generally 5 to 7 lobes that are smooth except for a slight hairs at the axils of the veins (Kirkman et al., 2007).

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

Figure 1. Unclear heredity. This suspected hybrid (rightmost) has the general shape of a northern red oak. However, it has rusty hairs on both the top and bottom of the leaf, which would be characteristic of a southern red oak.



Sample Collection

Leaf material from three red oak species (northern red oak, southern red oak, and pin oak) were collected for this experiment from White Oak Mountain and Bauxite trail systems on Southern Adventist University campus. Three samples of oaks suspected of being hybrids, based on their leaf morphology, were also collected. Leaf samples were collected in triplicate from each tree, documented in a spreadsheet, and photographed. The leaves were then rinsed with r.o. H_2O and stored at -80°C in 50 ml centrifuge tubes.

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 min. 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.

PCR Amplification

PCR reactions were optimized for 0.5 µl of template DNA as indicated by single PCR products after agarose electrophoresis. PCR reactions were performed for both rbcLa 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 as described previously. (CBOL plant working group, 2009)

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

PCR products was analyzed on 1% (w/v) agarose gels in TAE. Products was visualized by ethidium bromide staining or UView^{TM} 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).

Contamination Issues and DNA Sequencing

Analysis of the electropherograms from the initial sequence results indicated that the PCR products may be contaminated. The contamination was suspected due to the presence of overlapping signals (Figure 2). These signals seem to suggest that other DNA is in the sample.

In attempting to limit further contamination to the samples, the PCR preps were done under a hood with micropipettes that were sterilized under UV light and autoclaved and filter pipette tips. Additionally, a PCR cleanup kit was used on the samples. However, the latter sequence electropherograms showed only slight improvement. (Figure 2)

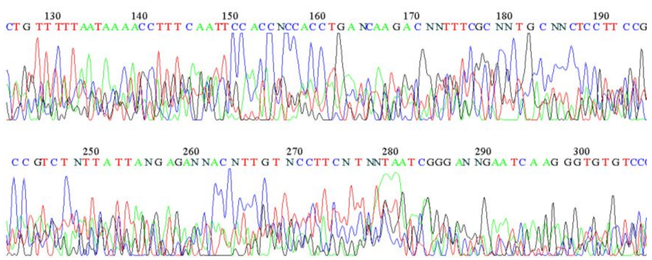


Figure 2. Electropherogram showing conflicting signals, indicating possible contamination. Example of a segment of contaminated sequence analysis results for a southern red oak (sample 24) sequenced with rbcL forward primer.

Sequence Alignments

The sequence results were compared to reference sequences obtained from GenBank to determine if they were valid for identification of the species. The alignments between sample oak sequences and the matching reference sequence for the species have given percent identities to each other in a range of 29.14 - 53.61% (Figure 3).

```
Reference       TACGAATCCATCTTCCTTTTTCTCCGTAACCAACCTTCTCATTTACGATTAACATCTTCT
SouthernRedOak  GNNNNANNCNNNTTCNTTTTTCCCCGNGATCNNNNNNCAGGNNTANNAGNANCANCTNNN
                  *  *   *** ****** ***  * *       *      **  *  * ** **

Reference       GGAGTCCT-TTTTGAACGACTCTGTTTATATAGAAAAATAGAACATTTTGCCGAAGTCTT
SouthernRedOak  NGNGGNNNNNNNNGNNNNANNNNNNNNAAAANNAAAAANANAAAATTTNGCNNAAGNCNT
                 * *         *    *         * *    **** * ** **** **  *** * *

Reference       TTCTAATGATTTTCCGGTCATCCCATGCTTTCTCAAGGATCCTTTCATGCATTATGTTAG
SouthernRedOak  TTCNAAGGATTTCNNGGNNNNCCGGNTTTCNCNG----GANCCTTNCNGGANNAGTNNNA
                *** ** *****   **     **     *  *         * **   * *  *

Reference       ATATCAAGGAAAATCAATTCTGGCTTCCAAAGACACACCTCTTCTAATGAATAAAT-GGA
SouthernRedOak  ANNTNGNGGAAAATCNATNNNGGNTNNCNANGANCCCCCNCNTCCAAGGGATAAATTGGA
                *  *   ******** **   ** *  * * **  * ** * ** ** * ****** ***

Reference       AATCTTACCTTGTCAATTTATGGCAATGTCATTTTGATGTATGGTCTCACGCGGCAAGTA
SouthernRedOak  NNNNNTANNNGGNAANTTANGGGAANGGNNNTTTNGNGGNGGGNCNCCNNNNNCCCGNN
                   **    *  ** *** ** ** *    ***   *   ** * *      *  *

Reference       TCCGTATAAACCAATTATCCAAGCATTCCCTCGATTTTTTGAGTTACTTTTCAAGTGTTC
SouthernRedOak  ACCGTTTNACACGANTNTTNNNNNNANTNCCCCCAATTTTTNGNNNNTTTTNAG-GGGTGN
                 **** * *  * * * *       * * ** * * *****       ***    * **
```

Figure 3. Typical segment of sequence alignment showing low percent identity. Alignment of southern red oak (sample 21) with a southern red oak reference sequence obtained from GenBank (accession number: KJ773058.1). Sequence obtained by using matK forward primer. Alignment done using Clustal Omega. 40.03% identity to the reference sequence.

The low similarity between the sample sequence and the reference sequence is 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%.

Analysis of Variation

We 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. Additionally, comparisons of sequences between southern red oaks with northern red oaks gave percent identities that were similar to the percentages between samples of the same species (Table 1). Similarly, the sequences from matK primers showed little difference in percent identity when comparing between samples of the same species, to a reference sequence of the same species, and between different species (Table 2).

Table 1. 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

Table 2. 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)
Norhtern red oak Reference	41.16	42.69	46.32
Northern red oak (#18)		36.09	36.61
Northern red oak (#12)			46.79

Summary of Findings

- Sequence alignments between the same species gave lower than expected percent identities for both matK and rbcL.
- Sequence alignments between sample species and a reference sequence of the same species also gave lower than expected percent identities.
- The comparison of sequences from different species showed similarly low percent identities as between members of the same species.

Work Remaining

- Continue to optimize PCR and sequencing conditions to improve specificity.
- Additional Barcoding trials with ycf1 primer pair to expand the data set.
- If the sequences prove to be reliable in differentiating between species, they will be used to identify parental identity of putative oak hybrids.

Acknowledgments

Southern Adventist University Academic Research Committee (ARC) Research Sustainability Grant

Biology / Allied Health Department of Southern Adventist University

ARC Academic Research Committee

References:

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., Tao, C., Guo, J., Zhou, S. (2015). Ycf1, the most promising plastid DNA barcode of land plants. *Scientific Reports, 5*. doi:10.1038/srep08348

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

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.

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