Developing a DNA barcode to distinguish between species of *Shorea*

Adam Essene
Department of Biological Sciences, Fordham University

ABSTRACT

Shorea is a genus of tropical hardwood trees found throughout Southeast Asia. The purpose of this study was to develop a DNA barcode for three species of *Shorea* found in Malaysian Borneo: *S. acuta, S. almon* and *S. inappendiculata.* In order to distinguish between these three species, primers were designed to target internal transcribed spacers (ITS) 1 and 2, two non-functional regions of ribosomal DNA. ITS1 and ITS2 were successfully amplified and sequenced for each species. Both loci contained variation that could identify *S. inappendiculata*, but neither were able to distinguish between *S. acuta* and *S. almon*. An additional barcoding locus is necessary in order to genetically identify these species.

Key Words: Shorea, internal transcribed spacer, ectomycorrhizal fungi, DNA barcoding

INTRODUCTION

Shorea is a genus comprising 196 species of hardwood trees found throughout the tropics of Southeast Asia. The majority of this diversity is found on the island of Borneo, which has over 138 documented species of *Shorea* (Ashton, 1982). Many of these species are economically significant timber trees (Tsumura et al. 2011); according to the International Union for Conservation of Nature (IUCN) Red List (1998), over 100 species of *Shorea* are classified as critically endangered. Like the all other members of the family Dipterocarpaceae, species of *Shorea* form obligate symbiotic root associations with a diverse community of ectomycorrhizal fungi (ECM) (Lee et al, 2002). As these fungal symbionts provide the physiological link between their host trees and nutrients in the soil (Smith & Read 2010), understanding the ecological factors that contribute to fine scale differences in ECM assembly has implications for the management of the increasingly fragmented forest reserves in this region (Edwards et al. 2011).

One poorly understood aspect of the ECM ecology in this system is the extent to which host specificity of different ECM taxa contributes to differences in assembly found associated with different species of *Shorea* (Peay et al, 2009). Studying the host specificity of these ECM provides a unique challenge; unlike other ectomycorrhizal tropical rainforests, which tend to be dominated by stands of one or two species of ECM-forming trees (McGuire et al. 2008), the lowland tropical rainforests of Southeast Asia typically contain a mix of many different species of *Shorea* with overlapping distributions. This means that root cores collected to assess ECM diversity will contain roots from many different potential hosts. Since these roots are morphologically indistinguishable, a DNA barcode that differentiates between species of *Shorea* is necessary in order to identify the host tree that each ECM is associated with.

In this study I assess the utility of the internal transcribed spacers (ITS) 1 and 2 as DNA barcodes to identify three ecologically distinct species of *Shorea*: *S. almon*, *S. acuta* and *S. inappendiculata*. ITS1 and ITS2 are non-functional regions of ribosomal DNA flanked by highly conserved sequences that code for the large and small subunits of the ribosome (figure 1). These regions have been proposed in the literature as potential barcodes for plant species identification because they are short, highly variable, and have many genomic copies (Stoeckle 2003, Chen et al 2010, Li et al. 2011). I conducted this study using leaf samples collected from a forest dynamics plot within Lambir Hills National Park, Malaysian Borneo, where over 50 different species of *Shorea* have been documented (Peay, 2009).

METHODS

All leaf samples were collected within the 52-hectare forest dynamics plot at Lambir Hills National Park, Malaysian Borneo. A total of nine leaf samples were collected from three saplings of each target species of *Shorea.* Species identification was based on the 2008 tree census data for the plot (provided by Dr. Stuart Davies). Leaf tissue was dried on silica gel for transport, and then DNA was extracted from each sample using the FastDNA Spin Kit (MP Biomedicals). Primers used to amplify

ITS1 and ITS2 (figure 1, table 1) were modified from Chen et al. (2010) to target the order Malvales (which contains *Shorea*) and prevent fungal amplification.

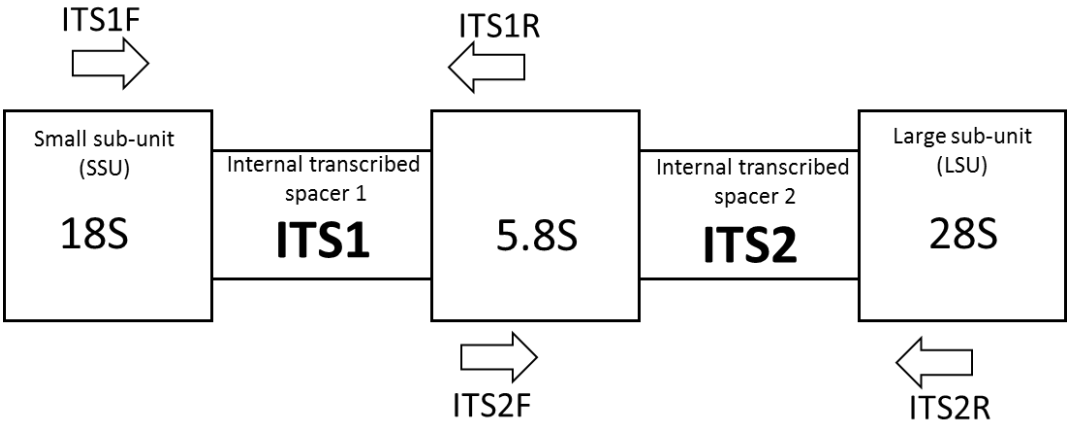

Figure 1. Schematic depiction of ITS1 & ITS2 showing approximate primer locations. ITS1R and ITS2F are reverse complements.

Primer	Sequence (5' - 3')
ITS1F	CACGAGAAGTCCACTGAACC
ITS1R	GCAATTCACACCAAGTATCGC
ITS2F	GCGATACTTGGTGTGAATTGC
ITS2R	GAGGACGCTTCTCCAGATAC

Table 1. Sequences of primers designed for ITS1 and ITS2.

Each PCR reaction contained 10 µL of GoTaq Green Master Mix (Promega, Madison, Wisconsin), 0.5µL of forward primer, 0.5µL of reverse primer, 1 ul of 0.5 ng/ul DNA, and dH20 to a final volume of 20 µL. PCR conditions were as follows: an initial denaturation step at 95°C for 3 minutes, then 50 cycles of 30s at 94°C, 30s at 61°C, and 30s at 72°C, followed by a final extension at 72°C for 2 minutes. PCR products were visualized on 1% agarose gels. All samples that were successfully amplified were PCR purified using a QIAQuick PCR Purification Kit (Qiagen, Hilden, Germany) and

sequenced by Genewiz, Inc (South Plainfield, NJ). Sequences were aligned with Clustal Omega (European Bioinformatics Institute, Cambridge, UK) and pairwise similarity comparisons were made using BLAST (National Center for Biotechnology Information, Bethesda, MD, USA).

RESULTS

Successful amplification for each species of *Shorea* was achieved using primers designed for ITS1 and ITS2 (figure 2). All PCR products were purified and sent for sequencing except for *S. inappendiculata* sample 1, ITS2. Sequencing was successful for all samples except for *S. almon* sample 2, ITS2.

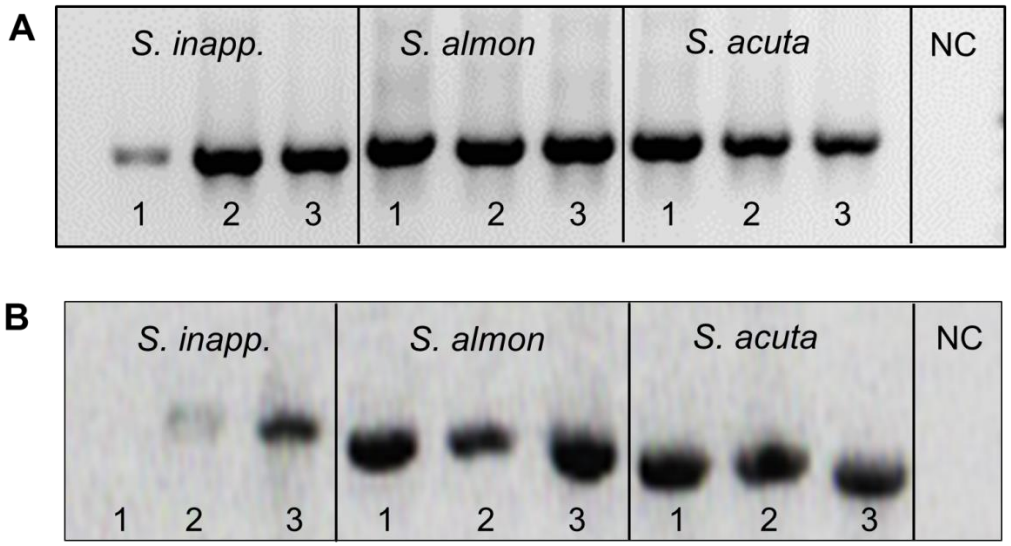

Figure 2. PCR amplification results for ITS1 (A) and ITS2 (B). Amplicons were loaded on a 1% agarose gel. Specimens of each species of Shorea are labeled 1 – 3, NC denotes no template, negative control.

Alignment of ITS1 sequences revealed 2 insertion/deletions and 8 base pair differences that distinguish *S. inappendiculata* from *S. almon* and *S. acuta* (Fig. 3). Differences between *S. almon* and *S. acuta* sequences were inconsistent with species designations. Pairwise similarity comparisons between

samples showed that the ITS1 sequences for *S. acuta* 1, *S. almon* 2 and *S. almon* 3 are nearly identical, while *S. acuta* 3 and *S. almon* 1 have very low inter- and intraspecific similarity (Table 2).

Figure 3. Two sections of ITS1 alignment that contain insertion/deletions and base pair changes (both highlighted in red) that distinguish *S. inappendiculata* from *S. almon* and *S. acuta*.

S. acuta 1					
S. acuta 2	98%				
S. acuta 3	95%	94%			
S. almon 1	96%	94%	92%		
S. almon 2	99%	98%	93%	95%	
S. almon 3	99%	98%	95%	96%	99%
	S. acuta 1	S. acuta 2	S. acuta 3	S. almon 1	S. almon 2

Table 2. Pairwise similarity (%) of S. acuta and S. almon ITS1 sequences

Alignment of ITS2 sequences revealed 5 insertion/deletions and 13 base pair differences that distinguish *S. inappendiculata* from *S. almon* and *S. acuta* (Fig. 4). There were no insertion/deletions or base pair differences that distinguish between *S. almon* and *S. acuta* at this locus.

Figure 4. Two sections of ITS2 alignment that contain insertion/deletions and base pair changes (both highlighted in red) that distinguish *S. inappendiculata* from *S. almon* and *S. acuta*.

DISCUSSION

The primer pairs that I designed for ITS1 and ITS2 were both successful at amplifying DNA from all three species of *Shorea*. Both of these regions enable the identification *S. inappendiculata*, but neither can adequately distinguish between *S. acuta* and *S. almon*. The pairwise similarity comparisons between *S. acuta* and *S. almon* ITS1 sequences illustrate that there is variation at this locus, but it is inconsistent with species identification. This unexpected result could be a product of one or more of the following: hybridization, which has been observed in other species of *Shorea* (Kamiya et al. 2011); species misidentification during the tree census, which is a real possibility given the morphological similarity of *S. almon* and *S. acuta* saplings; intraspecific divergence at this locus which does not reflect the phylogenetic relationship between my samples. The pairwise similarity comparisons most strongly support the latter explanation: the two most divergent samples, S. almon 1 and S. acuta 3, are also the two least similar. This kind of intraspecific divergence is a documented limitation of this locus for phylogenetic inference (Alvarez & Wendell, 2003, Kress et al. 2005).

In order to resolve the identity of my *S. almon* and *S. acuta* samples, an additional barcoding locus will be necessary. Studies comparing the relative ability of different loci to discriminate between species have found higher rates of success by combining ITS with other barcoding regions such as rbcL, matK and trnH-psbA (Kress et al. 2005, Chen et al 2010, Lia et al 2011). Additional work needs to be done to determine which of these loci in *Shorea* provide the greatest species discrimination power in combination with ITS. Using a two-locus barcode approach on the *S. almon* and *S. acuta* samples in this study as well as additional samples of each species from an herbarium should enable me to determine if the saplings I sampled were correctly identified. Finally, the utility of these primers for use with DNA from ectomycorrhizal roots needs to be assessed; if the primers developed for this study can amplify *Shorea* DNA with a low occurrence of fungal DNA contamination, then this is a valuable resource for future studies of the ectomycorrhizal ecology of this system.

Acknowledgements

Dr. Krista McGuire provided the impetus and direction for this project. Dr. Kabir Peay was extremely helpful negotiating my permission to conduct research at Lambir Hills National Park. I would like to thank Dr. Stuart Davies for providing tree distribution data for Lambir Hills, and Krystal Diaz and Amos Lim for giving me so much help in the field. Catharina Grubaugh and Kate Reid deserve a medal for their endless patience throughout this project and beyond. I would also like to thank Dr. Berish Rubin for his valuable guidance and constant support

Works cited

Ashton, P.S. 1982. Dipterocarpaceae. *In*: Van Steenis C.G.G.J. (ed.), Flora Malesiana 1(9): 237–552.

Chen, Shilin, H. Yao, J. Han, C. Liu, J. Song, L. Shi, Y. Zhu, X. Ma, T. Gao, X. Pang, K. Luo, Y. Li, X. Li, X. Jia, Y. Lin, C. Leon. 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plants. PLoS ONE 5(1): e8613.

Edwards, David, T. Larsen, T. Docherty, F. Ansell, W. Hsu, M. Derhe, K. Hamer, D. Wilcove. 2011. Degraded lands worth protecting: the biological importance of Southeast Asia's repeatedly logged forests. Proceedings of the Royal Society Biological Sciences 278(1702): 89-90.

H. S. Lee, S. J. Davies, J. V. LaFrankie, S. Tan, T. Yamakura, A. Itoh, T. Ohkubo, and P. S. Ashton. 2002. Floristic and structural diversity of mixed dipterocarp forest in Lambir Hills National Park, Sarawak, Malaysia. Journal of Tropical Forest Science 14: 379-400.

Kamiya, Koichi, Y. Gan, S. Lum, M. S. Khoo, S. C. Chua, N. Faizu. 2011. Morphological and molecular evidence of natural hybridization in Shorea. Tree Genetics & Genomes, 7(2): 297-306.

Kress, John, K. Wurdack, E. Zimmer, L. Weigt, D. Janzen. 2005. Use of DNA barcodes to identify flowering plants. Proceedings of the National Academy of Science of the United States of America 102(23): 8369–8374.

Lia, De-Zhu, L-M Gaoa, , H-T. Lia, H. Wanga, X-J. Gea, J-Q. Liua, Z-D. Chena, S-L. Zhoua, S-L. Chena, J-B. Yanga, C-X. Fua, C-X. Zenga, H-F. Yana, Y-J. Zhua, Y-S. Suna, S-Y. Chena, L. Zhaoa, K. Wanga, T. Yanga, and G-W. Duana. 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proceedings of the National Academy of Science of the United States of America 108(49): 19641–19646.

McGuire K., T. Henkel, I. Granzow de la Cerda, G. Villa, F. Edmund and C. Andrew. 2008. Dual mycorrhizal colonization of forest-dominating tropical trees and the mycorrhizal status of non-dominant tree and liana species. Mycorrhiza 18:217-222.

Peay, Kabir, P. Kennedy, S. Davies, S. Tan, T. Bruns. 2009. Potential link between plant and fungal distributions in a dipterocarp rainforest: community and phylogenetic structure of tropical ectomycorrhizal fungi across a plant and soil ecotone. New Phytologist 185(2): 529-542.

Smith, S. E., & Read, D. J. (2010). *Mycorrhizal symbiosis*. Academic Press, San Diego, CA.

Stoeckle, Mark. 2003. Taxonomy, DNA, and the bar code of life. BioScience 53(9): 796-797.

Tsumura, Yoshihiko, T. Kado, K. Yoshida, H. Abe, M. Ohtani, Y. Taguchi, Y. Fukue, N. Tani, S. Ueno, K. Yoshimura, K. Kamiya, K. Harada, Y. Takeuchi, B. Diway, R. Finkeldey, M. Na'iem, S. Indrioko, K. Kit, S. Ng, N. Muhammad, S. Lee. 2011. Molecular database for classifying *Shorea* species (Dipterocarpaceae) and techniques for checking the legitimacy of timber and wood products. Jounal of Plant Research 124(1): 35-48.