# Identification of Burseraceae trees from Peru: a comparison of the nuclear DNA marker ITS and the plastid DNA marker rbcL for DNA barcoding

**Elias Elbogen**

*Dept. of Environmental Sciences, University of California, Berkeley, California 94720 USA*

**Keywords: plant identification, Protium, phylogeny, Bayesian, sequence divergence**

## ABSTRACT

The immense plant diversity that is characteristic of tropical rain forests often makes it difficult for ecological and conservation studies to identify individual plant species and measure biodiversity. DNA barcoding is a species identification technique that utilizes standard, short DNA sequences to distinguish between species when traditional taxonomic identification is not practical. Accurate identification of animals with DNA barcoding has been well established, but a universally accepted DNA barcode for plants still does not exist. The use of nuclear DNA markers and plastid DNA markers from the chloroplast are the two contending approaches to DNA barcoding. This study compares the utility of the nuclear DNA marker ITS and the plastid DNA marker rbcL as DNA barcodes among 35 Burseraceae tree species from the Peruvian Amazon. I found that the proposed DNA barcode rbcL greatly underperformed the nuclear marker ITS as a DNA barcode. While both markers exhibited greater than 90% amplification success ITS demonstrated a mean pairwise percentage sequence divergence of 5.4% while rbcL demonstrated 0.83%. Additionally, at 1% sequence divergence resolution ITS discriminated between 99% of species-pairs while rbcL only discriminated between 26%. The results of my study suggest that ITS should not be completely discounted from the plant DNA barcode debate and rbcL be reevaluated as a proposed universal barcode.

## INTRODUCTION

Tropical rain forests are the most diverse terrestrial ecosystems (Fine and Ree 2006) and Amazonian rain forests contain the greatest diversity of tree species on earth (Gentry 1988). Balslev et al. (1998) found that a single hectare of Amazonian rain forest can contain up to 900 vascular plant species. As a result of this incredible species diversity, tropical forests comprise countless rare, and often endemic, species rather than large populations of more common species (Condit et al. 2000). To better understand the composition of tropical forest ecosystems it is essential to understand the forces that drive and maintain their remarkable biodiversity. Scientists have explained these forces in terms of the niche partitioning theory (Grinnell 1917), the stochastic niche theory (Tilman 2004), and the neutral theory of biodiversity and biogeography (Hubbell 2001), to state a few. Moreover, because tropical forests house such a vast amount of rare and diverse species, conservation efforts are often concentrated on these biodiversity hotspots (Myers et al. 2000) or directly tied to biodiversity estimates (Hubbell et al. 2008; Balmford & Long 1995). However, both ecological and conservation efforts rely on accurate measurements of biodiversity.

In order to measure biodiversity (e.g. the number of distinct species) individual species must be identified. Though, in such diverse and species-rich plant communities as tropical forests it is often not feasible to identify plants and measure plant diversity using traditional taxonomy (Dick & Kress 2009; Condit 1998; Sheil 1995). Traditionally, plants are most reliably identified by their reproductive characters (e.g. fruits and flowers) yet countless tropical forest species are not collected when reproductive because of short flowering periods or their reproductive characters are out of reach (e.g. large trees, epiphytes, and lianas) (Dick & Kress 2009). As a result, many plants must be identified by their more indeterminate, vegetative characters (e.g. leaves and bark). It is also very common for tropical plant species to be completely unidentifiable even by experts, either because the plant in question is very similar to related species or it lacks a scientific name altogether (Dick & Kress 2009; Condit 1998; Sheil 1995). For example, in the western Amazon, Ruokolainen et al. (2005) were unable to identify about 20% of the trees in their forest inventory plots.

DNA barcoding is a species identification technique

that utilizes standard, short DNA sequences to distinguish between species when traditional taxonomic identification is not practical. Accurate identification of animals with DNA sequences has been well established using the mitochondrial gene cytochrome c oxidase I (COI) as the standard, universal DNA barcode (Hebert et al. 2004). DNA barcoding for plants may also become an alternative, or adjunct, to traditional taxonomy and identification. But, there has yet to be found a universally suitable DNA region for all plants. Plastid DNA markers (usually from the chloroplast; maternally inherited) and nuclear DNA markers are the two most widely accepted DNA identification approaches (Dick & Kress 2009). Both methods use short, standardized regions of DNA to differentiate between taxa, making it possible to identify otherwise unidentifiable plants or plant parts and to quantify genetic diversity within plant groups or communities (Ausubel 2009; Kress & Erickson 2008). In order to accurately discriminate between species, genetic markers must exhibit high interspecific variation and low intraspecific variation (Kress & Erickson 2007). The majority of proposed plant DNA barcodes are single locus or multiple loci plastid markers from the chloroplast (Chase et al. 2005; Kress & Erickson 2007, 2008; Fazekas et al. 2008; Lahaye et al. 2008; CBoL 2009; Gonzalez et al. 2009; Newmaster & Ragupathy 2009) and nuclear markers are more often used for phylogenetic analyses of specific taxa groups (Fine et al. 2005; Sun et al. 1994).

Both chloroplast and nuclear DNA markers have advantages and disadvantages as identification tools for plants. Chloroplast DNA markers have been shown to more precisely produce single gene genealogies (Edwards 2009) and nuclear DNA markers may more accurately produce species phylogenies (Dick & Kress 2009). The chloroplast genome evolves more slowly than the nuclear genome, which means it is less variable (Lahaye et al. 2008; Kress & Erickson 2007). And multiple copies of nuclear markers can exist within the nuclear genome, making differentiation problematic (Chase et al. 2005). But, the effectiveness of each genetic approach depends on the specific group of plants being identified. The Plant Working Group of the Consortium for the Barcode of Life (CBoL) has chosen the combination of the two protein coding plastid genes matK and rbcL (both located in the chloroplast) as the standard DNA barcode for seed plants (CBoL 2009). The most commonly used nuclear marker is the internal transcribed spacer region (ITS) of nuclear ribosomal DNA (Dick & Kress 2009; Kress et al. 2005; Chase et al. 2005).

I will compare the efficacy of the two molecular identification approaches – the chloroplast DNA marker approach and the nuclear DNA marker approach – and compare both to traditional taxonomic classification. Fine et al. (2005) studied 35 western Amazonian tree species from the genera Protium, Crepidospermum, and Tetragastris in the tropical family Burseraceae. The 35 species from this study were accurately identified using traditional taxonomy based on reproductive and vegetative morphological characters and analyzed phylogenetically using two nuclear DNA markers, ITS and ETS (Fine et al. 2005). This provides an ideal sample group for inquiring about molecular identification approaches, allowing me to compare the performance of the proposed DNA barcode rbcL (matK failed to amplify in my pilot study) to differentiate between the 35 previously identified tree species from Fine et al. (2005) to the species differentiation performance of the nuclear DNA marker ITS. I will investigate the ability of each DNA marker to discriminate among said species and measure species diversity (e.g. number of species discriminated) by constructing phylogenies based on each DNA marker and comparing pairwise sequence divergence values produced from ITS and rbcL sequences respectively. I will also examine how changing the criteria for discriminating between species (1%, 5%, and 10% sequence divergence) affects measurements of species diversity. If barcoding approaches to measuring diversity become standard, it is important to consider the ramifications of how changing the criteria for determining species changes conclusions about diversity. One hypothesis is that the proposed DNA barcode rbcL will discriminate among species more accurately than the nuclear DNA marker ITS. The alternative hypothesis is that ITS will discriminate among species more accurately than rbcL.

## MATERIALS AND METHODS

### Taxon sampling

All 35 samples used in this study were previously collected from the Allpahuayo-Mishana National Reserve, southwest of Iquitos, Peru, and extracted for use by Fine et al. (2005) which mapped habitat association onto a phylogeny of the Amazonian trees. Of the 35 Burseraceae species collected, 31 species represent the genus Protium and 4 species represent the closely related genera Crepidospermum and Tetragastris (Table 2).

### Molecular Methods

DNA was extracted from dried leaf specimens using a DNeasy Plant Mini Kit (Qiagen, Valencia, CA). PCR amplification was performed for the coding chloroplast region rbcLa (first part of the rbcL gene, ~725 base pairs) in 20 µL reactions using TLA PCR PreMix (BIOneer, Alameda, CA) with 1 µL of 10 µM forward and reverse primers, and 50-100ng of DNA template. PCR products were cleaned using 1.0 µL of ExoSAP (USB Corporation, Cleveland, OH) followed by additional thermocycling. For primer information, PCR conditions, and references, see Table 1.

Sequencing was performed in the Museum of Vertebrate Zoology Laboratory at UC Berkeley on an ABI 377XL DNA sequencer (Applied Biosystems, Foster City, CA) in 10 µL reactions using BIGDYE sequencing reagents and protocols (Applied Biosystems, Foster City, CA). Sequences were edited in Geneious (Biomatters Ltd., Auckland, New Zealand) and aligned automatically with MUSCLE (Edgar 2004) and manually with MacClade v4.06 (Sinauer Associates, Sunderland, MA).

| Marker | Source | Primer | Sequence 5'-3' | PCR conditions |
|---|---|---|---|---|
| *rbcL.a* | Cowan et al. 2006 | 1f | ATGTCACCACAAACAGAAAC | 94°C 1min<br>35 cycles:<br>[94°C 30s 50°C 40s 72°C 40s]<br>72°C 5min |
| | | r724 | TCGCATATGTACCTGCAGTAGC | |
| ITS 1 & 2 (both include the 5.8s spacer) | Fine et al. 2005 | ny183 (f1) | CCTTATCATTTAGAGGAAGGAG | 97°C 1min<br>35 cycles:<br>[97°C 50s 55°C 50s 72°C 90s]<br>72°C 7min |
| | | ny45 (f1) | GCATCGATGAAGAACGTAGC | |
| | | ny43 (r2) | TATGCTTAAATTCAGCGGCT | |
| | | ny109 (r2) | GTGACGCCCAGGCAGACGT | |

Table 1. Primers and PCR conditions for the two DNA markers tested in this study. For ITS primers, (r) indicates a reverse primer, (f) indicates a forward primer, and (1) and (2) indicate amplification of either ITS 1 or ITS 2.

DNA sequences for ITS (~705 base pairs) were obtained from GenBank (accession numbers: AY375490-AY375527) as published by Fine et al. (2005).

Genetic and Phylogenetic Analyses

The uncorrected (P) pairwise distances between sequences were calculated and organized into matrices (Table 3, 4) using PAUP (PAUP* 4.0). P-distance is the number of uncorrected base pair changes (base pairs where multiple changes have occurred are not taken into consideration) between two DNA sequences.

The percentage sequence divergence criteria for distinguishing between species-pairs were set at 1%, 5%, and 10% sequence divergence and the number of species-pairs discriminated at each resolution were calculated for both ITS and rbcL (Table 5).

ITS and rbcL gene sequences were phylogenetically analyzed by using Bayesian statistics to construct phylogenetic trees following the MrBayes Manual Quick-Start protocol, using a GTR (general time reversible) evolutionary model (MrBayes 3.1). Bayesian posterior probabilities and branch lengths (base pair substitutions per site) were calculated and mapped onto ITS and rbcL based consensus trees (Fig. 1, 2).

## RESULTS

Amplification and Sequencing

Of the 35 Burseraceae species ITS was successfully PCR amplified for 33 species and rbcL was successfully amplified for 32 species (Table 2). ITS sequences were easily sequenced and aligned, however only 22 rbcL sequences were successfully sequenced and aligned (Table 2). Species that were not successfully amplified or sequenced were excluded from subsequent analyses.

Genetic and Phylogenetic Analyses

The pairwise percentage sequence divergence values for ITS sequences ranged from 0.59% to 9.0% with a mean of 5.4% (Table 3). The values for rbcL sequences ranged from 0 to 4.1% with a mean of 0.83% (Table 4).

The number of species-pairs discriminated by each marker decreased drastically as the percentage sequence divergence criteria for discriminating species increased (1%, 5%, and 10%) and ITS demonstrated a much higher rate of species discrimination than rbcL (Table 5). For example, ITS was able to discriminate between species-pairs at 1% and 5% sequence divergence resolutions though rbcL was only able to discriminate species-pairs at 1% resolution.

Bayesian consensus trees were constructed successfully based on ITS and rbcL sequences respectively (Figs. 1, 2). The topologies of each tree are quite different from one another. The phylogeny based on ITS is better resolved with three small, distal polytomies, the largest of which comprises six branches (Fig. 1); Crepidospermum and Tetragastris species are grouped into the same clade. The phylogeny based on rbcL is primarily composed of a large 16 branch polytomy comprising only Protium species (Fig. 2); Crepidospermum and Tetragastris species were resolved but not grouped into the same clade. The resolved topology (the lower portion) of the rbcL phylogeny has some similarities to the ITS phylogeny. For example, both phylogenies represent Protium ferrugineum and Protium Subserratum as sister species and Crepidospermum goudotianum as most closely related to Crepidospermum pranceii.

## DISCUSSION AND CONCLUSIONS

I investigated the utility of the nuclear DNA marker ITS and the plastid DNA marker rbcL as DNA barcodes by assessing the ability of each gene region to distinguish between species and match taxonomic classifications of 35 tree species from the Peruvian Amazon. Although many have reported the potential of rbcL as a candidate plant barcode (CBoL 2009; Newmaster & Ragupathy 2009; Kress & Erickson 2007; Newmaster et al. 2006), my findings, along with past findings (Kress et al. 2005; Salazar et al. 2003; Renner 1999; Gielly & Taberlet 1994), contend that rbcL is not variable enough at the species level to be considered a good DNA barcode. On the other hand, my results confirmed the reported variability of ITS (Chase et al. 2005; Kress et al. 2005; Sun et al. 1994). The high variability of ITS in my study suggest potential for ITS as a DNA barcode for Burseraceae species, but because ITS sequences were amplified with modified primers (Fine

Table 2. PCR amplification success. Asterisks indicate species that were amplified but unsuccessfully sequenced.

| Species | *rbcL* amplification | ITS amplification |
|---|---|---|
| *Crepidospermum goudotianum* | yes | yes |
| *C. prancei* | yes | yes |
| *C. rhoifolium* | no | yes |
| *Protium aldenianum* | yes* | yes |
| *P. altsonii* | yes* | yes |
| *P. amazonicum* | no | yes |
| *P. apiculatum* | yes | yes |
| *P. aracouchini* | yes | yes |
| *P. calanense* | yes | yes |
| *P. crassipetalum* | yes | yes |
| *P. decandrum* | yes* | yes |
| *P. divaricatum* ssp. *divaricatum* | yes | yes |
| *P. divaricatum* ssp. *krukoffi* | yes | yes |
| *P. elegans* | yes | no |
| *P. ferrugineum* | yes | yes |
| *P. gallosum* | yes | yes |
| *P. glabrescens* | yes* | yes |
| *P. grandifolium* | yes* | yes |
| *P. guacayanum* | yes | yes |
| *P. hebetatum* | yes* | yes |
| *P. heptaphyllum* ssp. *ulei* | yes | yes |
| *P. klugii* | yes | yes |
| *P. krukoffi* | yes* | yes |
| *P. laxiflorum* | yes | yes |
| *P. nodulosum* | yes* | yes |
| *P. opacum* | yes | yes |
| *P. pallidum* | yes | yes |
| *P. paniculatum* | yes | yes |
| *P. rubrum* | yes* | no |
| *P. sagotianum* | yes* | yes |
| *P. subserratum* ssp. *subserratum* | yes | yes |
| *P. tenuifolium* | no | yes |
| *P. trifoliolatum* | yes | yes |
| *P. urophyllidium* | yes | yes |
| *Tetragastris panamensis* | yes | yes |

et al. 2005) and ITS has been discredited as a potential barcode for not reliably sequencing across diverse taxa (CBoL 2009; Gonzalez et al. 2009; Kress & Erickson 2007) I do not propose ITS as a potential universal DNA barcode. I must also note that two other proposed DNA barcodes, matK (CBoL 2009; Lahaye et al. 2008) and trnH-psbA (Kress & Erickson 2007; Kress et al. 2005), were not evaluated in this study because they did not successfully amplify in my pilot study, representing a subset of five Protium species: P. altsonii, P. hebetatum, P. laxiflorum, P. ferrugineum, and P. subserratum. The conflicting results from past studies and my study suggest that determining a single universal DNA barcode for plants may be an unrealistic goal.

An ideal DNA barcode should be present in large groups of taxa, consistently PCR amplified, reliably sequenced with little manual editing, and demonstrate a high rate of species discrimination. In this study, ITS was successfully amplified and sequenced for 94% of the 35 Burseraceae species and demonstrated high sequence variability with a mean pairwise sequence divergence value of 5.4%. The high variability of ITS has been well documented: 1.73% intraspecific divergence among tropical tree species from French Guiana (Gonzalez et al. 2009), 5.7% mean sequence divergence for 48 species-pairs from diverse plant lineages including Fabales and Rosales (Kress & Erickson 2007), and 13.6% sequence divergence between Atropa and Nicotiana (Kress et al. 2005). And, ITS has even been proposed as a potential DNA barcode for angiosperms (Kress et al. 2005). At the same time, ITS has been discounted as a potential barcode due do its poor sequencing success across diverse taxa (Gonzalez et al. 2009; Kress & Erickson 2007). Contrary to my results, Gonzalez et al. (2009) studied 285 tropical tree species from French Guiana, encompassing 143 genera including Burseraceae, and found that ITS demonstrated a much lower sequencing success of only 41%. The greater performance of ITS in my study is probably a result of Fine et al. (2005) using modified primers as opposed to the proposed universal primers used by Gonzalez et al. (2009). This poses a serious drawback to the utility of ITS as a potential DNA barcode. The basis for DNA barcoding is to be able to identify plants more efficiently and accurately with a single DNA marker and single set of primers. If specific primers need to be designed in order to amplify ITS for diverse taxa then it is not universally applicable.

I found the DNA barcoding performance of rbcL to be inferior to ITS for Burseraceae. rbcL was successfully amplified for 91% of the Burseraceae species but only successfully sequenced for 63%. As demonstrated by my study, the high amplification success of rbcL has also been reported to be similar to 90% by Gonzalez et al. (2009), CBoL (2009), and Kress & Erickson (2007). On the other hand, rbcL has been discounted as a potential DNA barcode for being too long (~1,400 base pairs) to reliably amplify (Kress et al. 2005). The ease of amplification for rbcL in my study can most likely be attributed to the fact that I amplified only the first portion of the gene, rbcLa (~720 base pairs) (Gonzalez et al. 2009; Kress & Erickson 2007). The discrepancy between rbcL amplification and sequencing success in my study was probably a result of either poor quality DNA or a low quantity of DNA obtained during the DNA extraction process. With further troubleshooting, requiring additional time and money, usable rbcL sequences could most likely be attained for the species that did not have quality sequences. However, because of the low variability exhibited by the 22 rbcL sequences I obtained it would likely have little to no effect on my results to do so.

I found that rbcL demonstrated low sequence variability with an average pairwise sequence divergence value of 0.83%. As seen in Table 4, the rbcL sequences for some species-pairs were almost identical, demonstrating a sequence divergence value of zero. Kress et al. (2005) also reported that rbcL exhibited the same value of 0.83% mean sequence divergence between Atropa and Nicotiana species and other studies have reported similar results (Salazar et al. 2003; Renner 1999; Gielly & Taberlet 1994), disregarding it as a potential universal barcode. On the other hand, Newmaster et al. (2006) found that rbcL could discriminate among approximately 85% of congeneric species-pairs across diverse taxa and other studies have reported high rates of species discrimination for rbcL: 67% of Acacia species from 56 populations with a mean

**Table 3. Pairwise distances between species of Burseraceae based on the nuclear DNA marker ITS.** Numbers in bold indicate a sequence divergence value less than 1%. Numbers in italics indicate a value less than 5%.

| Species | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 *P. crassipetalum* | - | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 *T. panamensis* | 0.065 | - | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3 *C. rhoifolium* | 0.090 | 0.061 | - | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4 *C. goudotianum* | 0.074 | *0.044* | *0.032* | - | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 5 *C. prancei* | 0.071 | *0.043* | *0.025* | *0.019* | - | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 6 *P. sagotianum* | 0.072 | 0.069 | 0.083 | 0.071 | 0.065 | - | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 7 *P. tenuifolium* | 0.065 | 0.068 | 0.081 | 0.065 | 0.059 | *0.030* | - | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 8 *P. ferrugineum* | 0.074 | 0.058 | 0.075 | 0.056 | 0.061 | 0.074 | 0.071 | - | | | | | | | | | | | | | | | | | | | | | | | | | |
| 9 *P. subserratum* | 0.070 | 0.061 | 0.080 | 0.065 | 0.065 | 0.075 | 0.077 | *0.049* | - | | | | | | | | | | | | | | | | | | | | | | | | |
| 10 *P. gallosum* | 0.062 | 0.056 | 0.073 | 0.061 | 0.058 | 0.062 | 0.065 | 0.056 | 0.056 | - | | | | | | | | | | | | | | | | | | | | | | | |
| 11 *P. urophyllidium* | 0.065 | 0.058 | 0.074 | 0.062 | 0.059 | 0.062 | 0.064 | 0.055 | 0.058 | *0.010* | - | | | | | | | | | | | | | | | | | | | | | | |
| 12 *P. amazonicum* | 0.073 | 0.064 | 0.085 | 0.073 | 0.070 | 0.074 | 0.076 | 0.065 | 0.065 | *0.021* | *0.022* | - | | | | | | | | | | | | | | | | | | | | | |
| 13 *P. apiculatum* | 0.068 | 0.062 | 0.079 | 0.067 | 0.064 | 0.070 | 0.071 | 0.059 | 0.062 | *0.015* | *0.016* | **0.009** | - | | | | | | | | | | | | | | | | | | | | |
| 14 *P. glabrescens* | 0.073 | 0.072 | 0.083 | 0.063 | 0.063 | 0.073 | 0.064 | 0.071 | 0.079 | 0.069 | 0.066 | 0.078 | 0.072 | - | | | | | | | | | | | | | | | | | | | |
| 15 *P. guacayanum* | 0.071 | 0.077 | 0.083 | 0.070 | 0.065 | 0.074 | 0.065 | 0.077 | 0.084 | 0.073 | 0.068 | 0.082 | 0.076 | *0.046* | - | | | | | | | | | | | | | | | | | | |
| 16 *P. laxiflorum* | 0.063 | 0.063 | 0.077 | 0.064 | 0.060 | 0.064 | 0.061 | 0.059 | 0.065 | 0.059 | 0.058 | 0.067 | 0.062 | *0.040* | *0.043* | - | | | | | | | | | | | | | | | | | |
| 17 *P. altsonii* | 0.065 | 0.064 | 0.083 | 0.070 | 0.065 | 0.068 | 0.059 | 0.070 | 0.072 | 0.062 | 0.061 | 0.073 | 0.068 | *0.043* | 0.050 | *0.018* | - | | | | | | | | | | | | | | | | |
| 18 *P. hebetatum* | 0.065 | 0.062 | 0.081 | 0.068 | 0.064 | 0.066 | 0.064 | 0.069 | 0.072 | 0.061 | 0.059 | 0.071 | 0.066 | *0.043* | 0.050 | *0.018* | **0.006** | - | | | | | | | | | | | | | | | |
| 19 *P. heptaphyllum* | 0.061 | 0.061 | 0.077 | 0.061 | 0.058 | 0.064 | 0.061 | 0.068 | 0.070 | 0.059 | 0.055 | 0.072 | 0.065 | *0.034* | *0.047* | *0.034* | *0.037* | *0.034* | - | | | | | | | | | | | | | | |
| 20 *P. krukoffii* | 0.064 | 0.061 | 0.077 | 0.061 | 0.058 | 0.069 | 0.062 | 0.069 | 0.069 | 0.058 | 0.053 | 0.070 | 0.064 | *0.039* | *0.049* | *0.039* | *0.038* | *0.038* | *0.024* | - | | | | | | | | | | | | | |
| 21 *P. trifoliolatum* | 0.068 | 0.067 | 0.080 | 0.064 | 0.061 | 0.074 | 0.068 | 0.077 | 0.075 | 0.065 | 0.061 | 0.077 | 0.071 | *0.039* | 0.052 | *0.042* | *0.037* | *0.038* | *0.024* | *0.018* | - | | | | | | | | | | | | |
| 22 *P. aidanianum* | 0.065 | 0.065 | 0.083 | 0.067 | 0.062 | 0.068 | 0.062 | 0.074 | 0.080 | 0.068 | 0.064 | 0.076 | 0.071 | *0.033* | *0.044* | *0.044* | *0.046* | *0.049* | *0.038* | *0.043* | *0.041* | - | | | | | | | | | | | |
| 23 *P. div. ssp. divaricatum* | 0.065 | 0.061 | 0.076 | 0.062 | 0.058 | 0.065 | 0.062 | 0.065 | 0.072 | 0.061 | 0.056 | 0.070 | 0.064 | *0.036* | *0.039* | *0.028* | *0.036* | *0.033* | *0.028* | *0.036* | *0.037* | *0.033* | - | | | | | | | | | | |
| 24 *P. klugii* | 0.070 | 0.073 | 0.077 | 0.062 | 0.059 | 0.068 | 0.065 | 0.065 | 0.077 | 0.068 | 0.064 | 0.077 | 0.071 | *0.039* | *0.037* | *0.041* | 0.050 | 0.050 | *0.040* | *0.046* | *0.046* | *0.038* | *0.037* | - | | | | | | | | | |
| 25 *P. div. ssp. krukoffii* | 0.066 | 0.064 | 0.076 | 0.061 | 0.055 | 0.065 | 0.059 | 0.061 | 0.072 | 0.061 | 0.059 | 0.070 | 0.064 | *0.034* | *0.039* | *0.028* | *0.036* | *0.037* | *0.036* | *0.042* | *0.037* | *0.037* | *0.030* | *0.030* | - | | | | | | | | |
| 26 *P. paniculatum* | 0.067 | 0.065 | 0.077 | 0.064 | 0.056 | 0.070 | 0.068 | 0.067 | 0.078 | 0.064 | 0.059 | 0.073 | 0.067 | *0.043* | *0.046* | *0.036* | *0.042* | *0.039* | *0.037* | *0.044* | *0.039* | *0.046* | *0.034* | *0.036* | *0.021* | - | | | | | | | |
| 27 *P. aracouchini* | 0.071 | 0.064 | 0.072 | 0.062 | 0.059 | 0.069 | 0.067 | 0.069 | 0.075 | 0.067 | 0.063 | 0.077 | 0.071 | *0.030* | *0.043* | *0.037* | *0.039* | *0.041* | *0.032* | *0.040* | *0.035* | *0.034* | *0.031* | *0.033* | *0.033* | *0.039* | - | | | | | | |
| 28 *P. calanense* | 0.071 | 0.065 | 0.080 | 0.064 | 0.061 | 0.071 | 0.068 | 0.071 | 0.077 | 0.068 | 0.068 | 0.079 | 0.072 | *0.036* | *0.049* | *0.037* | *0.039* | *0.041* | *0.035* | *0.046* | *0.041* | *0.037* | *0.034* | *0.036* | *0.033* | *0.042* | *0.016* | - | | | | | |
| 29 *P. nodulosum* | 0.070 | 0.071 | 0.079 | 0.065 | 0.061 | 0.068 | 0.064 | 0.070 | 0.078 | 0.067 | 0.062 | 0.076 | 0.070 | *0.037* | *0.033* | *0.038* | *0.047* | *0.044* | *0.036* | *0.046* | *0.044* | *0.038* | *0.031* | *0.031* | *0.033* | *0.037* | *0.031* | *0.037* | - | | | | |
| 30 *P. pallidum* | 0.066 | 0.064 | 0.074 | 0.061 | 0.056 | 0.062 | 0.062 | 0.067 | 0.074 | 0.062 | 0.058 | 0.071 | 0.065 | *0.040* | *0.039* | *0.037* | *0.045* | *0.042* | *0.036* | *0.044* | *0.043* | *0.036* | *0.027* | *0.028* | *0.030* | *0.033* | *0.028* | *0.031* | *0.027* | - | | | |
| 31 *P. opacum* | 0.058 | 0.057 | 0.067 | 0.054 | *0.049* | 0.055 | 0.052 | 0.059 | 0.065 | 0.053 | 0.049 | 0.063 | 0.056 | *0.033* | *0.033* | *0.027* | *0.034* | *0.033* | *0.025* | *0.036* | *0.031* | *0.027* | *0.022* | *0.025* | *0.021* | *0.025* | *0.018* | *0.024* | *0.018* | *0.015* | - | | |
| 32 *P. decandrum* | 0.061 | 0.067 | 0.074 | 0.061 | 0.056 | 0.064 | 0.060 | 0.059 | 0.071 | 0.058 | 0.053 | 0.067 | 0.061 | *0.031* | *0.033* | *0.034* | *0.043* | *0.043* | *0.034* | *0.039* | *0.040* | *0.028* | *0.030* | *0.024* | *0.025* | *0.030* | *0.025* | *0.031* | *0.025* | *0.022* | *0.013* | - | |
| 33 *P. grandifolium* | 0.063 | 0.065 | 0.073 | 0.059 | 0.055 | 0.062 | 0.056 | 0.056 | 0.071 | 0.059 | 0.055 | 0.068 | 0.062 | *0.030* | *0.034* | *0.031* | *0.040* | *0.040* | *0.031* | *0.036* | *0.037* | *0.030* | *0.028* | *0.024* | *0.024* | *0.030* | *0.022* | *0.028* | *0.024* | *0.024* | *0.013* | **0.007** | - |

**Table 4. Pairwise distances between species of Burseraceae based on the plastid DNA marker *rbcL*.** Numbers in bold indicate a sequence divergence value less than 1%. There is no value greater than 4.1%.

| Species | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 *P. calanense* | - | | | | | | | | | | | | | | | | | | | | | |
| 2 *P. trifoliolatum* | 0.034 | - | | | | | | | | | | | | | | | | | | | | |
| 3 *P. klugii* | 0.028 | **0.003** | - | | | | | | | | | | | | | | | | | | | |
| 4 *P. aracouchini* | 0.034 | **0.007** | **0.003** | - | | | | | | | | | | | | | | | | | | |
| 5 *P. div. ssp. krukoffi* | 0.024 | 0.010 | **0.002** | **0.009** | - | | | | | | | | | | | | | | | | | |
| 6 *P. ferrugineum* | 0.033 | 0.012 | **0.008** | 0.011 | 0.012 | - | | | | | | | | | | | | | | | | |
| 7 *P. guacayanum* | 0.029 | **0.004** | **0.000** | **0.002** | **0.005** | **0.007** | - | | | | | | | | | | | | | | | |
| 8 *P. div. ssp. divaricatum* | 0.029 | **0.004** | **0.000** | **0.002** | **0.005** | **0.008** | **0.000** | - | | | | | | | | | | | | | | |
| 9 *P. laxiflorum* | 0.034 | **0.007** | **0.003** | **0.002** | **0.007** | 0.013 | **0.002** | **0.002** | - | | | | | | | | | | | | | |
| 10 *P. apiculatum* | 0.041 | 0.012 | **0.009** | **0.009** | 0.015 | 0.015 | **0.009** | **0.009** | 0.012 | - | | | | | | | | | | | | |
| 11 *P. gallosum* | 0.037 | 0.010 | **0.006** | **0.009** | 0.012 | 0.013 | **0.007** | **0.007** | 0.010 | **0.001** | - | | | | | | | | | | | |
| 12 *P. elegans* | 0.029 | **0.006** | **0.000** | **0.004** | **0.007** | 0.010 | **0.002** | **0.002** | **0.004** | 0.011 | **0.008** | - | | | | | | | | | | |
| 13 *P. heptaphyllum* | 0.029 | **0.004** | **0.002** | **0.002** | **0.007** | 0.010 | **0.002** | **0.002** | **0.002** | **0.009** | **0.007** | **0.003** | - | | | | | | | | | |
| 14 *P. crassipetalum* | 0.032 | **0.006** | **0.002** | **0.004** | **0.007** | 0.010 | **0.002** | **0.002** | **0.004** | **0.007** | **0.005** | **0.003** | **0.003** | - | | | | | | | | |
| 15 *P. paniculatum* | 0.029 | **0.004** | **0.000** | **0.002** | **0.005** | **0.008** | **0.000** | **0.000** | **0.002** | **0.009** | **0.006** | **0.002** | **0.002** | **0.002** | - | | | | | | | |
| 16 *P. pallidum* | 0.029 | **0.004** | **0.000** | **0.002** | **0.005** | **0.008** | **0.000** | **0.000** | **0.002** | **0.009** | **0.006** | **0.002** | **0.001** | **0.001** | **0.000** | - | | | | | | |
| 17 *P. opacum* | 0.033 | **0.006** | **0.002** | **0.005** | **0.007** | 0.010 | **0.002** | **0.002** | **0.004** | 0.011 | **0.008** | **0.003** | **0.003** | **0.003** | **0.002** | **0.002** | - | | | | | |
| 18 *P. subserratum* | 0.033 | **0.008** | **0.008** | **0.007** | 0.012 | **0.008** | **0.007** | **0.006** | **0.008** | 0.011 | 0.011 | **0.008** | **0.009** | **0.009** | **0.006** | **0.007** | **0.008** | - | | | | |
| 19 *C. goudotianum* | 0.029 | **0.008** | **0.006** | **0.006** | 0.010 | 0.011 | **0.005** | **0.005** | 0.014 | **0.009** | 0.016 | **0.006** | 0.012 | 0.012 | **0.005** | 0.013 | **0.006** | 0.013 | - | | | |
| 20 *C. pranceii* | 0.029 | **0.008** | **0.004** | **0.006** | **0.008** | **0.009** | **0.003** | **0.003** | 0.014 | **0.009** | 0.014 | **0.005** | 0.012 | 0.011 | **0.003** | 0.011 | **0.005** | 0.014 | **0.004** | - | | |
| 21 *P. urophyllidium* | 0.033 | **0.006** | **0.002** | **0.004** | **0.008** | 0.011 | **0.003** | **0.003** | **0.006** | **0.005** | **0.003** | **0.005** | **0.005** | **0.002** | **0.003** | **0.003** | **0.005** | **0.009** | **0.008** | **0.006** | - | |
| 22 *T. panamensis* | 0.036 | **0.009** | **0.004** | **0.007** | **0.005** | 0.011 | **0.004** | **0.004** | **0.008** | **0.004** | **0.002** | **0.005** | **0.005** | **0.002** | **0.004** | **0.004** | **0.006** | 0.011 | **0.009** | **0.007** | **0.002** | - |

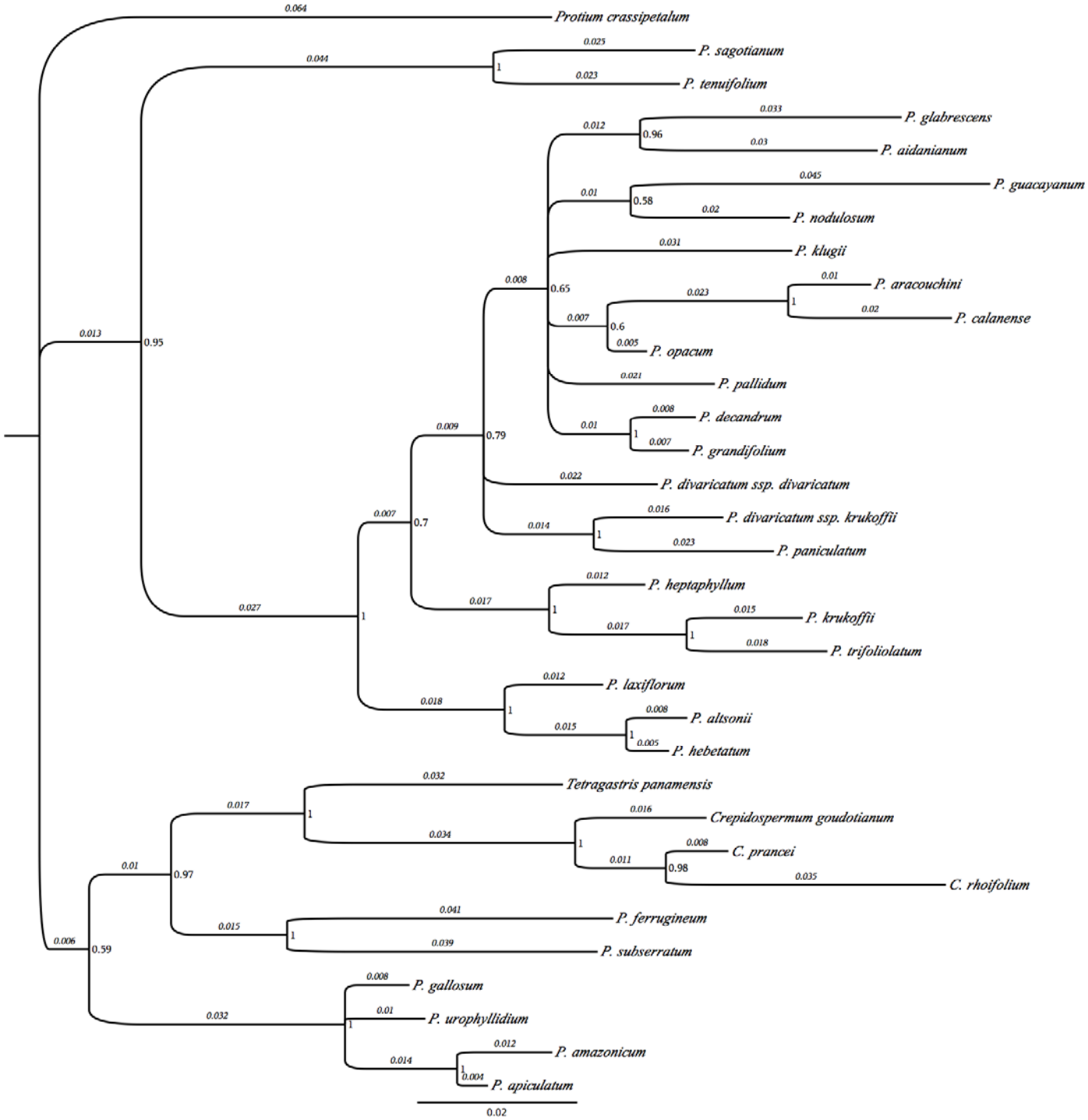


**Figure 1. Unrooted Consensus Bayesian tree constructed from the nuclear DNA marker ITS.** Nodes are labeled with their respective posterior probabilities and branch length values represent the number of substitutions per site.

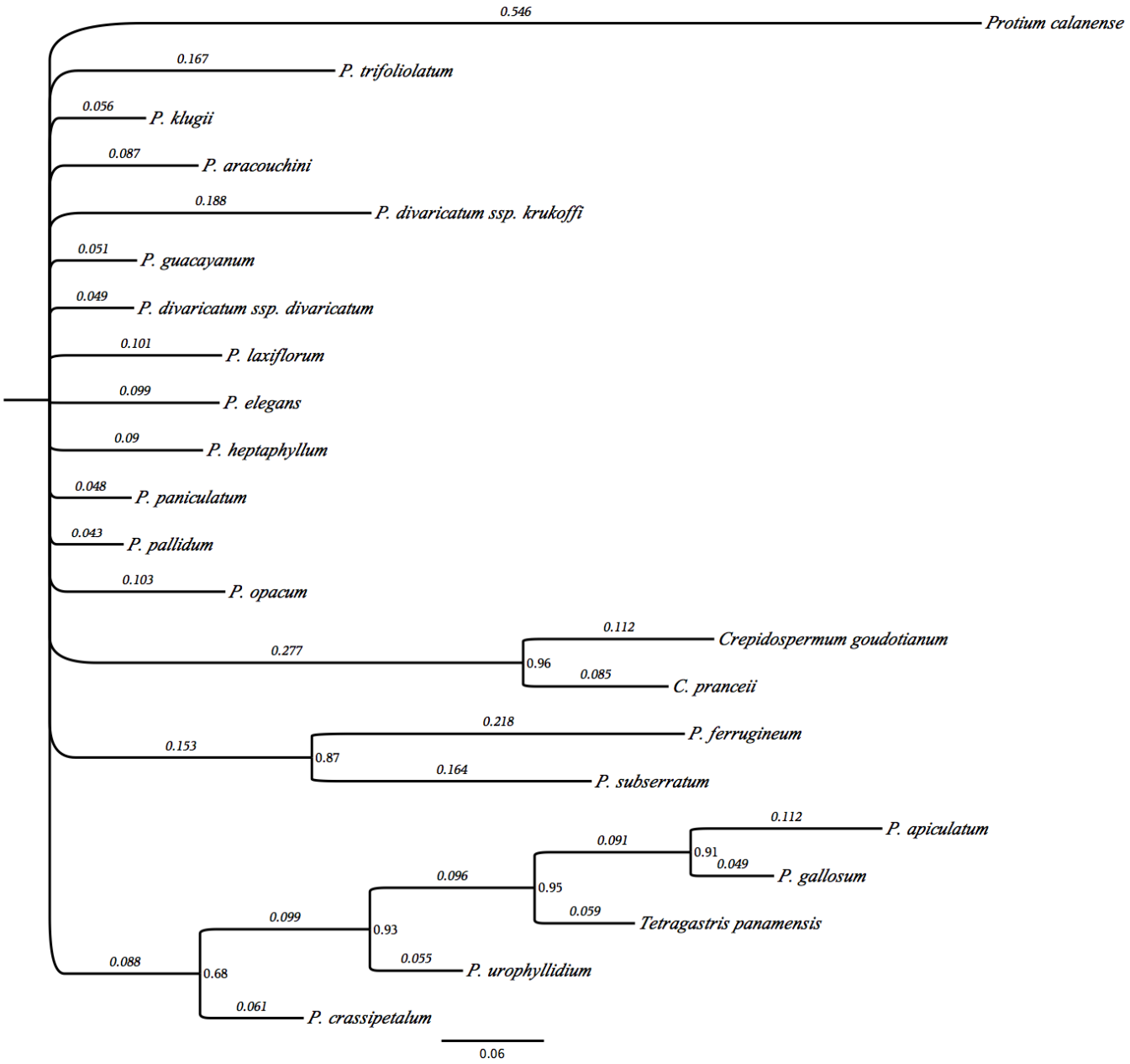


**Figure 2. Unrooted Consensus Bayesian tree constructed from the plastid DNA marker *rbcL*.** Nodes are labeled with their respective posterior probabilities and branch length values represent the number of base pair substitutions per site.

sequence divergence of 1.4% (Newmaster & Ragupathy 2009), 69.8% of 48 species-pairs representing major plant lineages including mosses, laurales, and brassicales (Kress & Erickson 2007), and 61% of 397 samples representing major lineages of angiosperms, gymnosperms, and cryptograms (CBoL 2009). Additionally, rbcL has also been suggested as a candidate universal locus for proposed two-locus barcodes (CBoL 2009; Kress & Erickson 2007). The contradictory performance of rbcL across studies confirms that the utility of rbcL is dependent upon taxa.

The phylogenies constructed from ITS and rbcL substantiate my findings that ITS is much more variable than rbcL in the Burseraceae species studied. Phylogenies infer evolutionary relationships based on similarities and differences between DNA sequences. When relationships between taxa cannot be determined from nucleotide differences in the DNA sequences the branching is displayed as a polytomy, where each node has two or more immediately descending branches. The large, 16 branch polytomy in the rbcL phylogeny indicates that the rbcL sequences were not divergent enough for MrBayes to determine which species were more closely related to each other. In contrast, the ITS phylogeny was better resolved, with fewer polytomies, because the ITS sequences were much more divergent. Though rbcL did not resolve species relationships very well, Crepidospermum sequences were divergent enough to be separately grouped in the phylogeny, suggesting that rbcL may be useful at differentiating between genera, as reported by CBoL (2009) and Kress & Erickson (2007). At the same time, however, the ITS phylogeny grouped Tetragastris and Crepidospermum species in a monophyletic group and the rbcL phylogeny did not, implying that rbcL is not divergent enough in the Burseraceae to infer accurate evolutionary relationships.

As a result of changing the sequence divergence criteria for discriminating between species to 1%, 5%, and 10% sequence divergence resolutions it was very evident that increasing the resolution decreases the number of species identified. At 1% resolution ITS discriminated 99% of species-pairs and rbcL only discriminated 26%. In addition, at 5% resolution ITS discriminated 62% of species-pairs and rbcL discriminated 0% and at 10% resolution both ITS and rbcL discriminated 0 species-pairs. These results reiterate once again that ITS is a much more variable gene region than rbcL in Burseraceae trees. Moreover, these findings are important because they demonstrate that measurements of species diversity based on genetic information are directly related to the source of that genetic information (e.g. DNA marker) and the species discrimination parameters by which that information is analyzed. If DNA barcoding becomes standard for measuring biodiversity considerations must be made to avoid vast over- or underestimations of the number of distinct species.

With a universally accepted DNA barcode plant species could be identified relatively quickly and easily for biodiversity studies and related conservation efforts. My study determined that the nuclear DNA marker ITS discriminates among Burseraceae species with a considerably higher variability than the plastid DNA marker rbcL, suggesting that ITS may perform well as a DNA barcode for Burseraceae species. However, a very important consideration in the plant barcode debate is universal application. That is, an ideal barcode should be easily sequenced and exhibit high interspecific variability for all land plants. Despite the effectiveness of ITS in my study, I cannot propose it as a universal DNA barcode because it was amplified with modified primers (Fine et al. 2005) and previous DNA barcode studies have found ITS to perform poorly as a DNA barcode (CBoL 2009; Gonzalez et al. 2009; Kress & Erickson 2007). Additionally, despite the poor performance of rbcL in my study previous studies have proposed rbcL as a candidate universal DNA barcode (CBoL 2009; Newmaster & Ragupathy 2009; Kress & Erickson 2007; Newmaster et al. 2006). These contradicting findings suggest that a single, universal DNA barcode for plants is not practical (Gonzalez et al. 2009; Chase et al. 2007; Newmaster et al. 2006). Though it may be more costly and timely, it may be more realistic to decide on multiple DNA barcodes based on taxa-specific utility. Rather than focus barcoding research on creating a centralized plant barcode database from a single, universal barcode I suggest that it would be more useful to direct barcoding research towards a poly-barcode approach that defines the most effective DNA barcodes for different groups of taxa, like families or genera

Table 5. Percentage of species-pairs discriminated using three different pairwise percentage sequence divergence criteria. The number of species-pairs discriminated at each resolution is in parenthesis.

| Sequence Divergence Criterion | ITS | *rbcL* |
|---|---|---|
| 1% | 99% (525) | 26% (60) |
| 5% | 62% (328) | 0 |
| 10% | 0 | 0 |

## AKNOWLEDGEMENTS

I first thank the ES 196 instructors Patina Mendez and Kurt Spreyer and graduate student instructors Lucy Diekmann and Gabrielle Wong-Parodi for their continual dedication and support to all of the Environmental Sciences students. I especially thank Patina Mendez for her time and personal assistance reading and rereading my thesis drafts.

Secondly, I thank my faculty mentor Paul Fine (Integrative Biology). Without his direction and guidance my project could not have been completed. Despite his extremely busy schedule, including the birth of his second child, Paul was available for one-on-one meetings throughout the duration of my project, for which I am especially grateful.

Thirdly, I express the most wholehearted thanks to Tracy Misiewicz. As a PhD student in the Fine Lab and

informal advisor for my project, Tracy and I interacted on almost a daily basis. My project seemed doomed a number of times throughout the year but it was always Tracy's invaluable advice and patience that kept me and my project going.

## LITERATURE CITED


1. Fazekas, A. J., K. S. Burgess, and P.R. Kesanakurti. 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS ONE 3: e2802.
2. Fine, P. V. A., D. C. Daly, G. V. Munoz, I. Mesones, K. M. Cameron. 2005. The Contribution of Edaphic Heterogeneity to the Evolution and Diversity of Burseraceae Trees in the Western Amazon. Evolution 59: 1464-1478.
3. Fine, P. V. A. and R. H. Ree. 2006. Evidence for a Time-Integrated Species-Area Effect on the Latitudinal Gradient in Tree Diversity. The American Naturalist 168: 796-804.
4. Gentry, A.H. 1988. Tree species richness of upper Amazonian forests. Proceedings of the National Academy of Sciences 85: 156-159.
5. Gielly, L. and P. Taberlet. 1994. The Use of Chloroplast DNA to Resolve Plant Phylogenies - noncoding versus rbcL sequences. Molecular Biology and Evolution 11: 769-777.
6. Gonzalez, M. A., C. Baraloto, J. Engel, S. A. Mori, P. Pétronelli, B. Riéra, A. Roger, C. Thébaud, and J. Chave. 2009. Identification of Amazonian Trees with DNA Barcodes. PLoS ONE 4: e7483.
7. Grinnell J. 1917. The Niche-relationships of the California thrasher. Auk 34: 427–433.
8. Hebert, P. D. N., E. H. Penton, J. M. Burns, D. H. Janzen, and W. Hallwachs. 2004. Ten species in one: DNA barcoding reveals cryptic species in the Neo tropical skipper butterfly Astraptes fulgerator. Proceedings of the National Academy of Sciences 101: 14812-14817.
9. Hubbell S. P. 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press.
10. Hubbell S. P., F. He, R. Condit, L. Borda-de-Água, J. Kellner, H. ter Steege. 2008. How many tree species are there in the Amazon and how many of them will go extinct? Proceedings of the National Academy of Sciences 105: 11498–11504.
11. Kress, W. J. and D. L. Erickson. 2008. DNA barcodes: Genes, genomics, and bioinformatics. Proceedings of the National Academy of Sciences 105: 2761-2762.
12. Kress, W.J. nd L. Erickson. 2007. A Two-Locus Global DNA Barcode for Land Plants: The Coding rbcL Gene Complements the Non-Coding trnH-psbA Spacer Region. PLoS ONE 2: e508.
13. Kress, W.J., K. J. Wurdack, E. A. Zimmer, L. A. Weight and D. H. Janzen. 2005. Use of DNA barcodes to identify flowering plants. Proceedings of the National Academy of Sciences 102: 8369-8374.
14. Lahaye, R., M. van der Bank, D. Bogarin, J. Warner, F. Pupulin, G. Gigot, O. Maurin, S. Duthoit, T. G. Barraclough and V. Savolainen. 2008. DNA barcoding the floras of biodiversity hotspots. Proceedings of the National Academy of Sciences 105: 2923-2928.
15. Newmaster, S. G. and S. Ragupathy. 2009. Testing plant barcoding in a sister species complex of pantropical Acacia (Mimosoideae, Fabaceae). Molecular Ecology Resources 9: 172-180.
16. Newmaster, S. G., A. J. Fazekas, and S. Ragupathy. 2006. DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach. Canadian Journal of Botany 84: 335-341.
17. Newmaster, S. G., A. J. Fazekas, R. Steeves, and J. Janovec. 2008. Testing candidate plant barcode regions in the Myristicaceae. Molecular Ecology Resources 8: 480-490.
18. Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca and J. Kent. 2000. Biodiversity hotspots for conservation priorities. Nature 403: 853-858.
19. Renner, S. S. 1999. Circumscription and phylogeny of the Laurales: Evidence from molecular and morphological data. American Journal of Botany 86: 1301-1315.
20. Ruokolainen, K., H. Tuomisto, R. Kalliola. 2005. Landscape heterogeneity and species diversity in Amazonia. Pages 251–270 in E. Bermingham, C.W. Dick, and C. Moritz, editors. Tropical Rainforests: Past, Present, and Future. University of Chicago Press, Chicago, IL, USA.
21. Salazar, G. A., M. W. Chase, M. A. S. Arenas, and M. Ingrouille. 2003. Phylogenetics of Cranichideae with emphasis on Spiranthinae (Orchidaceae, Orchidoideae): Evidence from plastid and nuclear DNA sequences. American Journal of Botany 90: 777-795.
22. Sheil, D. 1995. A critique of permanent plot methods and analysis with examples from Budongo Forest, Uganda. Forest Ecology and Management 77: 11-34.
23. Sun, Y., D. Z. Skinner, G. H. Liang, and S. H. Hulbert. 1994. Phylogenetic analysis of Sorghum and related taxa using internal transcribed spacers of nuclear ribosomal DNA. Theoretical and Applied Genetics 89: 26-32.
24. Tilman, D. 2004. Niche tradeoffs, neutrality, and community structure: A stochastic theory of resource competition, invasion, and community assembly. Proceedings of the National Academy of Sciences 101: 10854-10861.