Testing the effectiveness of *rbcLa* DNA-barcoding for species discrimination in tropical montane cloud forest vascular plants (Oaxaca, Mexico) using BLAST, genetic distance, and tree-based methods

Sonia Trujillo-Argueta [1], **Rafael F. del Castillo** [Corresp., 2], **Abril Velasco-Murguía** [1]

[1] CIIDIR Oaxaca, Instituto Politécnico Nacional, Xoxocotlán, Oaxaca, Mexico

[2] CIIDIR Oaxaca, Instituto Politécnico Nacional, Xoxocotlán, Oaxaca, México

Corresponding Author: Rafael F. del Castillo
Email address: fsanchez@ipn.mx

DNA-barcoding is a species identification tool that uses a short section of the genome that provides a genetic signature of the species. The main advantage of this novel technique is that it requires a small sample of tissue from the tested organism. In most animal groups, this technique is very effective. In plants, however, the recommended standard markers, such as*rbcLa,*may not always work, and their effects remain to be tested in many plant groups, particularly from the Neotropical region. We examined the discriminating power of*rbcLa*in 55 tropical cloud forest vascular plant species from 38 families (Oaxaca, Mexico). We followed the CBOL criteria using BLASTn, genetic distance, and monophyly tree-based analyses (neighbor-joining, NJ, maximum likelihood, ML, and Bayesian inference, BI).*rbcLa*universal primers amplified 69% of the samples and yielded 91.30% bi-directional sequences. Sixty-three new*rbcLa*sequences were established. BLAST discriminates 80.8% of the genus but only 15.4% of the species. Genetic distances among*Quercus, Oreopanax*, and*Daphnopsis*species were nil. Contrastingly, Ericaceae (5.6%), Euphorbiaceae (4.6%), and Asteraceae (3.3%) species displayed the highest within-family genetic distances. According to the most recent angiosperm classification, NJ and ML trees successfully resolved (100%) monophyletic species. ML trees showed the highest mean branch support value (87.3%). Only NJ and ML trees could successfully discriminate*Quercus*species belonging to different subsections:*Quercus martinezii*(white oaks) from*Q. callophylla*and*Q. laurina*(red oaks). The ML topology could distinguish species from the Solanaceae clade that the best BLAST match could not. Also, the BI topology showed a polytomy in this clade, and the NJ tree displayed low-support branch values. We do not recommend genetic-distance approaches for species discrimination. More published *rbcLa* sequences are necessary for BLAST to be more effective. Instead, the ML tree-based analysis displays the highest species discrimination among the tree-based analyses. With the ML topology in

selected genera,*rbcLa*helped distinguish infrageneric taxonomic categories, such as subsections, grouping affine species within the same genus, and discriminating species in most cases. Since the ML phylogenetic tree could discriminate 48 species out of our 55 studied species, we recommend this approach to resolve tropical montane cloud forest species using*rbcLa*, as an initial step and improving DNA amplification methods.

"Testing the effectiveness of *rbcLa* DNA-barcoding for species discrimination in tropical montane cloud forest vascular plants (Oaxaca, Mexico) using BLAST, genetic distance, and tree-based methods"

Sonia Trujillo-Argueta[1], Rafael F. del Castillo[1+] and Abril Velasco-Murguía[1].

[1] Instituto Politécnico Nacional, CIIDIR, Oaxaca de Juárez, Oaxaca, México.

Corresponding Author:

Rafael F. del Castillo

Hornos 1003, Sta. Cruz Xoxocotlán, Oaxaca, 71230, México

Email address: fsanchez@ipn.mx

Abstract

DNA-barcoding is a species identification tool that uses a short section of the genome that provides a genetic signature of the species. The main advantage of this novel technique is that it requires a small sample of tissue from the tested organism. In most animal groups, this technique is very effective. In plants, however, the recommended standard markers, such as *rbcLa,* may not always work, and their effects remain to be tested in many plant groups, particularly from the Neotropical region. We examined the discriminating power of *rbcLa* in 55 tropical cloud forest vascular plant species from 38 families (Oaxaca, Mexico). We followed the CBOL criteria using BLASTn, genetic distance, and monophyly tree-based analyses (neighbor-joining, NJ, maximum likelihood, ML, and Bayesian inference, BI). *rbcLa* universal primers amplified 69% of the samples and yielded 91.30% bi-directional sequences. Sixty-three new *rbcLa* sequences were established. BLAST discriminates 80.8% of the genus but only 15.4% of the species. Genetic distances among *Quercus, Oreopanax*, and *Daphnopsis* species were nil. Contrastingly, Ericaceae (5.6%), Euphorbiaceae (4.6%), and Asteraceae (3.3%) species displayed the highest within-family genetic distances. According to the most recent angiosperm classification, NJ and ML trees successfully resolved (100%) monophyletic species. ML trees showed the highest mean branch support value (87.3%). Only NJ and ML trees could successfully discriminate *Quercus* species belonging to different subsections: *Quercus martinezii* (white oaks) from *Q. callophylla* and *Q. laurina* (red oaks). The ML topology could distinguish species from the Solanaceae clade that the best BLAST match could not. Also, the BI topology showed a polytomy in this clade, and the NJ tree displayed low-support branch values. We do not recommend genetic-distance approaches for species discrimination. More published *rbcLa* sequences are necessary for BLAST to be more effective. Instead, the ML tree-based analysis displays the highest species discrimination among the tree-based analyses. With the ML topology in selected genera, *rbcLa* helped distinguish infrageneric taxonomic categories, such as subsections, grouping affine species within the same genus, and discriminating species in most cases. Since the ML phylogenetic tree could discriminate 48 species out of our 55 studied species, we recommend this approach to resolve tropical montane cloud forest species using *rbcLa*, as an initial step and improving DNA amplification methods.

Introduction

A biodiversity inventory is crucial as a first step to protecting species and ecosystems. A significant portion of global biodiversity remains unnamed. Recent estimations indicate that 8.7 million species of multicellular organisms occur on Earth, but about 20% of those species have been described using morphological approaches since 1750 (Centre for Biodiversity Genomics, CBG, 2021). Thus, it is urgent to speed up the species identification process (Hvistendahl, 2021). DNA barcoding was recently proposed to identify species using short-standardized sequences and only requires a small sample of tissue (Hebert et al., 2003). Cytochrome oxidase 1 (*CO1*) successfully discriminates against many animal species but does not resolve plant species. The Consortium for the Barcode of Life's (CBOL) plant working group evaluated several plastid DNA regions based on universality, sequence quality, and species discrimination, recommending

using a core of a 2-locus combination of partial genes *rbcLa* + *matK* as the plant barcode (Group C Hollingsworth et al., 2009). Such a universality has not been found in all plant groups, and other studies suggest using additional loci (Kress & Erickson, 2007; Fasekas et al., 2008; Pang et al., 2012; China plant B Li et al., 2011). Moreover, *matK* may work very well for species of orchids (Lahaye et al., 2008) but not for certain fern groups (Trujillo-Argueta et al., 2021). Furthermore, in some angiosperm genera, such as *Salix* (Percy *et al*., 2014) and *Quercus* (Piredda et al., 2011), plastid markers might not work at all.

On average, the resolution of the tested DNA barcoding markers for plants is not as high as barcode markers used for many animal groups (CBOL Plant Working Group, 2009; Fazekas et al., 2008). Of the possible plant markers, *rbcLa* appears to be one of the best available. Although far from perfect, the resolution of *rbcLa* was shown to be better than those tested in arid plants in the United Arab Emirates (Maloukh et al., 2017) and Saudi Arabia (Bafeel et al., 2012). Also, *rbcLa* can be a valuable tool to identify species in conditions in which other methods are impractical. For instance, this marker was successfully used for studying the patterns of root diversity in old-field communities in Ontario, Canada (Kesanakurti et al., 2011). This kind of research is encouraging, but more studies are needed to explore the resolution potential of this marker for species in ecosystems other than those of temperate regions. The Neotropics are considered the richest region in biodiversity (Gaston & Williams 1996; Thomas 1999). However, some ecosystems in the neotropics have been little explored regarding DNA barcode studies. The available studies are often limited to a few plant groups such as orchids (Lahaye et al., 2008) or ferns (Nitta, 2020; Trujillo-Argueta et al., 2021).

In Mexico, the tropical montane cloud forest (TMCFs) is a top priority ecosystem for conservation due to its high diversity, endemism richness, and anthropogenic threats (Villaseñor, 2010; Toledo-Aceves et al., 2011). However, local DNA barcoding needs to be developed. This study aims to evaluate the performance of the plant core DNA barcode *rbcLa* using universal primers for vascular plants and without using other markers in a tropical montane cloud forest of the Mixteca Baja, Oaxaca, Mexico. We followed the three above-mentioned CBOL criteria and built a barcode library of native plant species for this region.

Methods

Species this study and study site

One hundred samples of plants belonging to different families were collected in a tropical montane cloud forest at San Miguel Cuevas, Santiago Juxtlahuaca Municipality, Oaxaca, Mexico (17°15’00.96” N, 98°02’57.34”, centroid coordinates). The climate in this area is semi-humid, temperate to semi-warm (1382 mm and 16.8 C, mean annual precipitation and temperature, Fernandez-Eguiarte et al., 2020), with soils rich in organic matter (Instituto Nacional de Estadística y Geografía, 2005), and a mean altitude of 2187m. The municipal council of San Miguel Cuevas granted permission to conduct our field studies on their lands. Two commissioners of the communal property of this municipality, Mr. Pedro Gil (2017- July 2018) and Mr. Damián Domínguez (July 2018 - June 2019), were directly responsible for such

permissions. Mr. Heladio Luna Rodríguez, a San Miguel Cuevas Community local authority member, supervised, guided, and helped us throughout the field trips. In no case was the entire plant collected. Collecting the samples did not kill the plants, which were left alive in their original places. Based on The International Union for Conservation of Nature (IUCN) Red List (accessed December 19th, 2021) from all 55 species in this study, more than half (57.7%) were not previously registered (Data Deficit DD); 43.6% belong to the Least Concern (LC) category. Also, *Daphnopsis tuerckheimiana* holds the status of Near Threatened (NT); and *Oreopanax sanderianus*, that of Vulnerable species (VU).

Plant vouchers were determined by the following specialists: Daniel Tejero-Díez, UNAM FES Iztacala, México, lycopod and ferns; Sergio Zamudio, Institute of Ecology, Veracruz, México, Berberidaceae; Rafael F. del Castillo, IPN CIIDIR Oaxaca, Mexico, Pinaceae; Jesús Guadalupe González Gallegos, University of Guadalajara, Mexico, Lamiaceae; Socorro González Elizondo, IPN CIIDIR Durango, Mexico, Cyperaceae and Ericaceae; Susana Valencia Avalos, UNAM Facultad de Ciencias, Mexico, Fagaceae; J.R. Kuethe, University of Auckland, New Zealand, Passifloraceae; and Rufina García, Abril Velasco-Murguía and Rafael F. del Castillo, IPN CIIDIR Oaxaca, Mexico, the rest of the specimens. The herbarium vouchers were deposited at the herbarium of CIIDIR Oaxaca, Instituto Politécnico Nacional (OAX), pending for registration numbers due to the pandemic crisis. The species and their IUCN Red List Status are shown on Table 1.

DNA amplification and sequencing

Several fresh leaves from each sampled plant were collected and placed in a Ziplock® bag. The samples were kept at -20°C in a freezer until processed. The number of samples collected per taxon was one and occasionally two.

Genomic DNA was extracted from 2mg leaf tissue with FastDNA SPIN kit and FastPrep® (MP Biomedicals, USA) equipment.

DNA concentration (ng/µl) and purity (260/280A) from the genomic DNA extracted were measured with a Biophotometer (Eppendorf®). Plant core barcoding partial gene rbcLa was used for amplification. We used standard primers from the Canadian Center for DNA Barcoding (CCDB) (Kuzmina, 2011), rbcLa-F ATGTCACCACAAACAGAGACTAAAGC (Tate & Simpson, 2003) and rbcLa-R GTAAAATCAAGTCCACCRCG (Kress & Erickson, 2007). rbcLa was amplified using a 25µL volume of reaction mixture: 15.8 µL of nuclease-free water, 5 µl MyTaq Buffer reaction (kit MyTaqDNA Polymerase Bioline), 1µL of forward primer, 1µL of reverse primer, 0.2 µL of MyTaq Polymerase and 2µL of isolated genomic DNA template. PCR reaction was carried out using an Applied Biosystems Veriti® thermocycler. We followed Fasekas et al. (2012) protocols for rbcLa amplification. The PCR temperature cycling program was: 94°C for 4 min; 35 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 1 min; final extension of 72°C for 10 min. Amplified PCR products were detected using agarose gel electrophoresis (1.2 % agarose gel TBE) under UV light by staining with GelRed Nucleic Acid (Biotium). PCR products were purified using the EZ-10 Spin Column PCR Products Purification Kit (Biobasic).

All PCR products were sequenced by Capillary Electrophoresis Sequencing (CES) in an ABI 3130xl Genetic Analyser at the Laboratorio Bioquímica Molecular UBIPRO FES Iztacala UNAM and with an AB3730 at the Laboratorio de Servicios Genómicos LANGEBIO-CINVESTAV.

DNA Alignment

rbcLa sequence chromatograms were manually edited and assembled into contigs using CodonCode Aligner v.9.0.1 http://www.codoncode.com/aligner. Consensus sequences were generated and aligned using MUSCLE (Edgar, 2004). These alignments were examined by eye and corrected when necessary.

BOLD and Genebank

Our study was registered under the name "Diversity of a humid temperate forest in Oaxaca, Mexico" project code DVHTF at The Barcode of Life Data System (BOLD, http://www.boldsystems.org). BOLD is a bioinformatics workbench devoted to acquiring, storing, analyzing, and publishing DNA barcode records (Ratnasingham & Hebert, 2007). Three files were included in the metadata submitted to BOLD: 1) Specimen data file including detailed voucher information, scientific names of the taxa sampled, collection dates, geographical coordinates, elevation, collectors, identifiers, and habitat. 2) An image file was submitted with high-quality specimen images from each plant. 3) A trace file was submitted along with primers and the direction of sequences. Sequences uploaded to BOLD were edited and aligned in FASTA format and referenced by Sample IDs. Sequences were also submitted to the GenBank.

Species differentiation

To evaluate species discrimination using rbcLa sequences, we used three approaches:

a) The Basic Local Alignment Search Tool for nucleotide (BLASTn) method (Altschul et al. 1990), which searches against the sequence database available online by the National Center for Biotechnology Information (NCBI) https://www.ncbi.nlm.nih.gov. Identification at the genus level was considered successful when all hits with the maximum percent identity scores >99% involved a single genus. Species identification was considered successful only when the highest maximal percent identity included a single species and scored >99% (Abdullah, 2017; Bafeel et al., 2012).

b) Genetic divergence. Interspecific and intraspecific distances were analyzed in MEGAX (Kumar et al., 2018). Genetic distance was inferred from 1000 replicates, and the evolutionary distances were computed using the Kimura 2-parameter method with gaps/missing data treatment adjusted using pairwise deletion. The genetic distances (%) of families, genera, and

species were analyzed in the Barcode of Life Data Systems (BOLD, www.boldsystems.org) (Ratnasingham & Hebert 2007).

c) Monophyly tree-based analyses using Neighbor-Joining (NJ), Maximum Likelihood (ML), and Bayesian Inference (BI) analysis.

NJ was analyzed in MEGAX (Kumar et al., 2018) inferred from 1000 replicates, and the evolutionary distances were computed using the Kimura 2-parameter method with gaps/missing data treatment adjusted using pairwise deletion. ML analyses were run on the IQ-TREE web server (http://iqtree.cibiv.univie.ac.at). Internal node support and bootstrap analyses were calculated using 1000 iterations. Tree inference using Bayesian analysis was run on MrBayes 3.2.2 on XSEDE via the CIPRES supercomputer cluster (www.phylo.org) for 10 million generations. The resultant ML and BI trees were visualized in the interactive Tree of Life (iTOL) (Letunic & Bork 2019). We evaluated which of the tree-based methods (NJ, ML, and MB) recovered more monophyletic species with a bootstrap/posterior probabilities support of >70% (de Groot et al., 2011).

Results

DNA Amplification and sequencing success

We could successfully amplify 69% of the botanical samples collected. We studied 38 families, of which 27 had one species and 11 families 2 to 5 species (Table 1). Of the 55 studied species, 29.1% were herbs, and 70.9% were trees and shrubs. From this subset, we could obtain high-quality bidirectional sequences (>250bp) in 91.3% of the species, using the standard primers of the CCDB for the *rbcLa* barcode.

BLAST

Using BLASTn, we obtained 100% resolution in all the 38 families studied and 80.8% in 48 genera. Only 47.3% of our 55 studied species were previously registered in the *rbcLa* sequences of the GenBank database (Figure 1). We also contributed to 13 new species in the GenBank Taxonomy Database. These species were not previously registered for any other gene sequence. With the available accessions at the GenBank, we found that *rbcLa* can unambiguously discriminate only 15.4% of the studied species at the species level (Figure 1). Just four species, *Monnina xalapensis, Cnidoscolus aconitifolius, Iresine diffusa,* and *Lophosoria quadripinnata*, were found to best BLAST match to a single species with more than 99% identity. Most of our *rbcL* sequences matched from 2-12 species with ≥99% maximal percent identity; and seven species, *Alnus acuminata, Solanum hispidum, Quercus laurina, Quercus callophylla, Pinus montezumae, Osmanthus americanus,* and *Physalis phyladelphica*, matched the *rbcL* sequences in the GenBank with >30 different species. The best BLAST match identifications per species for the *rbcLa* plastid barcode are shown in Table 2.

A specimen data file, image file, and trace file(s) were submitted to BOLD along with edited and aligned sequences for each of our 63 botanical samples (55 species and eight different duplicates) and can be accessed through the BOLD DNA database (http://www.boldsystems.org) under the 'DVHTF' project. Sixty-three sequences were obtained in this study for *rbcLa*, BOLD Process ID, and GenBank Accession numbers (Table 1).

Genetic divergence

The distribution of intra- and interspecific K2P distances across all taxon pairs of our 55 species of plants of The Mixteca Baja, Oaxaca, tropical montane cloud forest, obtained from partial gen *rbcLa* are shown in Figure 2. Mean pairwise genetic distance within species was 0, within genus 0.65 $\pm$ 0.07, and 1.76 $\pm$ 0.03 within families. Congeneric species of *Quercus*, *Daphnopsis,* and *Oreopanax* did not show genetic divergence. Contrastingly, *Solanum*, *Deppea,* and *Pinus* displayed intergeneric differences (Table 3) The mean genetic divergence observed in the studied families with two or more genera is shown in Table 4. The highest mean divergence values were observed in the Ericaceae, Euphorbiaceae, and Asteraceae families.

Monophyly tree-based analyses

Phylogenetic tree-based analysis using Neighbor-Joining (Supplementary Fig.S1), Maximum Likelihood (Figure 3), and Bayesian Inference tree (Supplementary Fig.S2) were reconstructed to evaluate our 55 species discrimination using the *rbcLa* barcode region. In all cases, ferns and lycopodium were used as outgroups. These tree-based methods evaluated which tree rendered the greatest species resolution and whether the barcode sequences generated monophyletic species (Table 5). NJ and ML phylogenetic trees resolved 100% of monophyletic species using *rbcLa.* Nevertheless, the clade support value > 70% with a bootstrap of 1000 replicates yielded the most robust phylogeny in the ML tree (87.3%) than the one obtained in the NJ tree (70.9%). Therefore, we present the ML phylogenetic tree (Figure 3). Although the BI tree showed the highest clade support value (92.7%), this tree did not resolve all 55 species as monophyletic species. Two polytomies were observed in the clade of the *Quercus* species and the Solanaceae clade (Supplementary Fig. S2).

Discussion

Our study reveals the advantages and limitations of the *rbcLa* barcode region for species identification of vascular plant species of a neotropical montane cloud forest. First, the amplification success was not universal, but bi-directional sequencing was highly successful when feasible. BLAST identification at the genus level is accurate in most cases but usually not for species identification. Finally, in selected genera, this marker helped distinguish infrageneric taxonomic categories, such as subsections, and helps to group affine species within the same genus. Below, we discuss in detail these issues.

Multiple factors can cause the absence of DNA amplification in some samples. Since we could amplify *rbcLa* in several species, the possibilities of methodological failures or problems with

the reactants or the lab equipment used are unlikely. One possible cause of the amplification failure is DNA degradation in some samples, as those were collected in the field and brought to the lab. During this time, the tissues may become degraded in some species. This appears to be a plausible explanation for cases in which DNA from tissue samples was successfully amplified in one individual but not in another of the same species. This is the case of *Solanum nigricans* (this study) and *Dryopteris wallichiana*, which could not be amplified in this study but were successfully amplified in a previous study using samples from different plants (Trujillo-Argueta et al., 2021). Another possibility is that the pair of *rbcLa* universal premiers used may not work for certain species. Our *rbcLa* amplification success (69%) could be increased using the alternative set of universal primers proposed by CCDB for gene barcode *rbcL.*

Our sequencing success (91.30%) was high and similar to those reported in other works. In a study of root diversity patterns using plastid gene *rbcL*, Kesanakurti et al. (2011) registered 96% amplification success with 85% sequencing success. In another study that identified Sicily's most threatened plant taxa, the amplification and sequencing successes were 96% and 95%, respectively (Giovino et al., 2016). In a study of the temperate flora of Canada, the use of *rbcLa* gave a 91.4% sequencing success (Burgess et al., 2011).
Our BLAST results were higher for genus discrimination (80.77%) than the values obtained for species differentiation (15.38%). Results from other regions and species are variable. For example, in wild, arid plants, discrimination at genus and species levels were lower than ours: 50% and 8%, respectively (Bafeel et al., 2012), but higher in a comprehensive study of the local flora of Canada (91% and 44%, Braukmann et al., 2017). In a study of threatened species of Sicily, the discrimination at the genus level was lower (52%) but higher at the species level (48%) than our results (Giovino et al., 2016). The peculiarities of the biology of the studied species may also account for the observed discrimination variability. Part of our low percent species discrimination results using BLASTn can be explained by low marker resolution, as was noticed in those species that matched their *rbcLa* sequence with more than 30 different species in the GenBank database (*Alnus acuminate, Solanum hispidum, Quercus laurina, Quercus callophylla, Pinus montezumae, Osmanthus americanus and Physalis phyladelphica*). Another explanation is misidentified voucher specimens in public DNA databases, an issue that several authors have acknowledged (e.g., Abdullah, 2017; Burguess et al., 2011). Since it is customarily to describe species based on morphological characteristics, it is possible that hybridization and polyploidy, which are common in plants, may contribute to decreasing barcoding species discrimination (Hollinsworht, 2011; Fasekas et al., 2008). Since more than half of the species in this study (52.72%) lacked comparative data in the GenBank database, it is necessary to increase the DNA barcode database, particularly for tropical wild plant species. Indeed, we contributed to new 63 *rbcLa* sequences to BOLD, its metadata, and the GenBank database. Although 42 of our species already had a *rbcLa* sequence on the GenBank database, new records on these species might help discover new haplotypes or geographical variants (Hajibabaei et al., 2007). Even if *rbcLa* does not have high species discrimination, it does for genus discrimination, which for some ecological studies might be enough (e.g., Kesanakurti et al., 2011).

Our distribution of intra- and interspecific genetic divergence (Figure 2) agrees with the premise that a DNA barcode must exhibit high interspecific but low intraspecific divergence (Lahaye et al., 2008). The percent interspecific divergence of this study (0.65) is similar to those reported in other hotspot diversity areas such as the Mediterranean Basin (0.89) (Giovino et al., 2016) and

Southern Africa (0.82) (Lahaye et al., 2008). The lack of genetic divergence observed in the three genera of trees: *Quercus* (*Q. martinezii, Q. laurina, and Q. callophylla*); *Oreopanax* (*O. sanderianus and O. xalapensis*), and *Daphnopsis* (*D. selerorum and D. tuerckheimiana*) concurs with Smith & Donoghue (2008). These authors found that the rates of molecular evolution are low in woody plants with long generation times compared to herbs. In the case of oaks (*Quercus),* several attempts have been made to identify species in Italy, using different plastid barcodes without success since hybridization and polyploidy are expected to be high in this group (Piredda et al., 2011). Null genetic divergence obtained in *Oreopanax* and *Daphnopsis* (Table 3) is of concern since *Oreopanax sanderianus* and *Daphnopsis tuerckheimiana* are on the red list of IUCN. The highest values of genetic distance found in the Ericaceae (5.57%), Euphorbiaceae (4.59%), and Asteraceae (3.3%) families that hold many herbs and shrubs species agree with the assumption that *the rbcLa* barcode has a better species differentiation for non-tree species. Moreover, a study conducted in a subalpine forest in Southwest China found a better DNA barcode resolution for herbs than for tree species (Tan et al., 2018). However, more studies are needed to confirm this trend in other species and localities.

The phylogenetic arrangements found in our study using barcode *rbcLa* concur with the recent Angiosperm Phylogeny Group classification (APG IV) of flowering plants (The Catalog of Life Partnership, 2017). The percent monophyletic species resolution obtained in this study using NJ (100%), ML (100%), and BI (85.45%) phylogenetic trees, was higher compared to 17% of species resolution found in arid wild plants using ML trees (Bafeel et al., 2012), barcoding the biodiversity of Kuwait (58%) using NJ trees (Abdullah, 2017) and the 71.8% registered in two biodiversity hotspots of Mesoamerica and Southern Africa, using ML and BI trees (Lahaye et al., 2008). Our ML phylogenetic tree showed the most robust phylogeny (87.27%), *Ocotea helicterifolia, Quercus callophylla, Quercus laurina, Iresine difussa, Berberis lanceolata, Moussonia deppeana,* and *Osmanthus americanus*, could not be resolved as monophyletic species with a clade bootstrap support value $\geq$ 70%.
Most of these species are trees in agreement with the assumption that rates of molecular evolution are low in woody plants compared to herbs (Smith & Donoghue, 2008). For those species that could not be differentiated with the ML tree, we suggest the addition of a second barcode.

Species discrimination can be improved by using tree-based phylogenetic methods rather than BLAST analysis and genetic distance approaches. For instance, using NJ and ML phylogenetic trees, it was possible to differentiate *Quercus martinezii* from *Q. laurina* and *Q. callophylla* (Supplementary Fig.S1, Figure 3) despite unsuccessful best BLAST matches and the null genetic divergence observed in *Quercus*. Based on an updated infrageneric classification of the oaks (Denk et al., 2017), *Q. martinezii* belongs to the white oaks (subsection *Quercus*), while *Q. callophylla* and *Q. laurina* belong to the red oaks (subsection *Lobatae*). In the Solanaceae family, three out of the five studied species (*Physalis philadelphica, Solanum hispidum*, and *Solandra maxima*) share high similitude with at least 30 species using the best BLAST match results. Furthermore, using our best BI tree, we observed a polytomy in the *Solanaceae* clade (Supplementary Fig.S2), and a low discrimination value in the NJ tree. However, these species could be resolved with our ML phylogenetic tree. Taxonomic species are usually described based on morphological characteristics that can easily be altered by local adaptation, phenotypic plasticity, or neutral morphological polymorphism, which may cause a

single variable species to be classified as many species (e.g., Gemeinholzer & Bachmann, 2005). On the other hand, very recent divergence and little differentiation might contribute to the inability of barcoding to separate species in some cases (Birch et al., 2017).

Conclusions

DNA barcoding using *rbcLa* can be a promising identification tool primarily at the family and genus level for vascular plant species of the neotropical montane cloud forest. We identify three major problems with the use of this technique. First, the lack of a universal amplification capability is probably associated with DNA degradation in some cases, but without ruling out other factors requiring further study. Second, the inability to detect certain morphological species is probably not related to *rbcLa* itself but to biological (e.g., polyploidy and hybridization) and technical (misidentifications or taxonomic misclassifications) problems. Third, the few available registers in the BOLD and GenBank databases (more than half of our species, 52.72%, did not have previous *rbcLa* sequence records). Indeed, we contributed new 13 species to the GenBank Taxonomy Database and 63 new sequences for *rbcLa* in BOLD and GenBank. We found preliminary evidence suggesting that the ability of the marker to discriminate species is not randomly distributed among taxa. Herb and shrub species in the Asteraceae, Ericaceae and Euphorbiaceae families showed the highest genetic distance using *rbcLa*, which can be helpful to distinguish congeneric species. Contrastingly, we detected nil genetic divergence among congeneric species in long-lived tree genera, *Quercus, Oreopanax*, and *Daphnopsis.* Nonetheless, the accuracy for discriminating species can be substantially improved using tree-based analysis. While BLAST and genetic distance approaches could not differentiate *Quercus* species, NJ and ML could successfully separate white oaks (*Quercus martinezii)* from red oaks (*Q. callophylla* and *Q. laurina)*. Also, most species in the Solanaceae family that showed unsuccessful BLAST results and low genetic distance could be discriminated against with ML phylogenetic tree. The ML phylogenetic tree showed the most robust phylogeny (87.27%) of all our 55 studied species of the tropical montane cloud forest of San Miguel Cuevas in Oaxaca state, Mexico. The establishment of this local barcode database will be valuable for a broad range of potential ecological, conservational, and phylogenetic applications.

Acknowledgments

Raúl Rivera García and Heladio Luna, assisted in field work. We are thankful to San Miguel Cuevas Community at Oaxaca State, México, for permitting access to their forest.

References

Abdullah M T. 2017. Conserving the biodiversity of Kuwait through DNA barcoding the flora. D. Phil. Thesis, University of Edinburgh & Royal Botanic Garden Edinburgh.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of molecular biology, 215(3), 403-410.

Bafeel, S. O., Arif, I. A., Bakir, M. A., Al Homaidan, A. A., Al Farhan, A. H., & Khan, H. A. (2012). DNA barcoding of arid wild plants using rbcL gene sequences. Genet. Mol. Res, 11(3), 1934-41.

Birch, J. L., Walsh, N. G., Cantrill, D. J., Holmes, G. D., & Murphy, D. J. (2017). Testing efficacy of distance and tree-based methods for DNA barcoding of grasses (Poaceae tribe Poeae) in Australia. PLoS One, 12(10), e0186259.

Braukmann, T. W., Kuzmina, M. L., Sills, J., Zakharov, E. V., & Hebert, P. D. (2017). Testing the efficacy of DNA barcodes for identifying the vascular plants of Canada. PLoS one, 12(1), e0169515.

Burgess, K. S., Fazekas, A. J., Kesanakurti, P. R., Graham, S. W., Husband, B. C., Newmaster, S. G., ... & Barrett, S. C. (2011). Discriminating plant species in a local temperate flora using the rbcL+ matK DNA barcode. Methods in Ecology and Evolution, 2(4), 333-340.

CBOL Plant Working Group (2009) A DNA barcode for land plants. Proceedings of the National Academy of Sciences of the United States of America, 106, 12794–12797.

Centre for Biodiversity Genomics, University of Guelph (2021). The Global Taxonomy Initiative 2020: A Step-by-Step Guide for DNA Barcoding. Technical Series No. 94. Secretariat of the Convention on Biological Diversity, Montreal, 66 pages.

China Plant, B. O. L. G., Li, D. Z., Gao, L. M., Li, H. T., Wang, H., Ge, X. J., ... & Duan, G. W. (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 Sciences of the United States of America, 108(49), 19641-6.

De Groot, G. A., During, H. J., Maas, J. W., Schneider, H., Vogel, J. C., & Erkens, R. H. (2011). Use of rbcL and trnL-F as a two-locus DNA barcode for identification of NW-European ferns: an ecological perspective. PLoS one, 6(1), e16371.

Denk, T., Grimm, G. W., Manos, P. S., Deng, M., & Hipp, A. L. (2017). An updated infrageneric classification of the oaks: review of previous taxonomic schemes and synthesis of evolutionary patterns. Oaks physiological ecology. Exploring the functional diversity of genus Quercus L., 13-38.

Doumenge, C.; Gilmour, D.; Ruíz-Pérez, M., and Blockhus, J. 1995. Tropical montane cloud forests: conservation status and management issues. In: Hamilton, L. S.; Juvick, J. O., and Scatena, F. N., Ed. Tropical Montane Cloud Forests. New York: Springer-Verlag; pp. 24-37.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research, 32(5), 1792-1797.

Fazekas, A. J., Burgess, K. S., Kesanakurti, P. R., Graham, S. W., Newmaster, S. G., Husband, B. C., ... & Barrett, S. C. (2008). Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PloS one, 3(7), e2802.

Fazekas, A. J., Kuzmina, M. L., Newmaster, S. G., & Hollingsworth, P. M. (2012). DNA barcoding methods for land plants. In DNA barcodes (pp. 223-252). Humana Press, Totowa, NJ.

Fernandez-Eguiarte A., Zavala-Hidalgo, J., Romero-Centeno, R., 2020. Atlas Climático Digital de México (versión 2.0). Centro de Ciencias de La Atmósfera. Universidad Nacional Autónoma de México. Servicio Meteorológico Nacional, Comisión Nacional Del Agua. http://atlasclimatico.unam.mx/atlas/kml/ (accessed 13 March 2020).

Gaston, K. J. and Williams, P. H. 1996. Spatial patterns in taxonomic diversity. In: Gaston, K. J. Biodiversity: A biology of numbers and difference. Oxford, UK: Blackwell Science, Limited 202-229.

Gemeinholzer, B., & Bachmann, K. (2005). Examining morphological and molecular diagnostic character states of Cichorium intybus L.(Asteraceae) and C. spinosum L. Plant Systematics and Evolution, 253(1), 105-123.

Giovino, A., Marino, P., Domina, G., Scialabba, A., Schicchi, R., Diliberto, G., ... & Scibetta, S. (2016). Evaluation of the DNA barcoding approach to develop a reference data-set for the threatened flora of Sicily. Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology, 150(4), 631-640.

Group, C., Hollingsworth, P.W. et al. A DNA barcode for land plants. Proc. Natl. Acad. Sci. USA, 106 31, 12794-12797 (2009).

Hajibabaei, M., Singer, G. A., Hebert, P. D., & Hickey, D. A. (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. TRENDS in Genetics, 23(4), 167-172.

Hebert, P. D., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences, 270(1512), 313-321.

Hollingsworth, P. M. (2011). Refining the DNA barcode for land plants. Proceedings of the National Academy of Sciences, 108(49), 19451-19452.

Hvistendahl, M. (2021). A Tale of Two Worlds. Scientific American Special Edition, The age of humans 30 (4), 37-41.

Instituto Nacional de Estadística y Geografía (INEGI), 2005. Prontuario de Informaci´on Geogr´afica Municipal de los Estados Unidos Mexicanos: Santiago Juxtlahuaca, Oaxaca. http://www3.inegi.org.mx/contenidos/app/mexicocifras/datos_ geograficos/20/20469.pdf (accessed 20 March 2020).

Kesanakurti, P. R., Fazekas, A. J., Burgess, K. S., Percy, D. M., Newmaster, S. G., Graham, S. W., ... & Husband, B. C. (2011). Spatial patterns of plant diversity below-ground as revealed by DNA barcoding. Molecular Ecology, 20(6), 1289-1302.

Kress, W. J., & Erickson, D. L. A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. PLoS one 2 6, e508 (2007).

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular biology and evolution, 35(6), 1547.

Kuzmina M.L. 2011. Canadian Center for DNA Barcoding (CCDB) Protocols. Primers sets for plants and fungi. Available: http://www.dnabarcoding.ca

Lahaye, R., Van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., ... & Savolainen, V. (2008). DNA barcoding the floras of biodiversity hotspots. Proceedings of the National Academy of Sciences, 105(8), 2923-2928.

Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic acids research, 47(W1), W256-W259.

Maloukh, L., Kumarappan, A., Jarrar, M., Salehi, J., El-Wakil, H., & Rajya Lakshmi, T. V. (2017). Discriminatory power of rbcL barcode locus for authentication of some of United Arab Emirates (UAE) native plants. 3 Biotech, 7(2), 1-7.

Nitta, J. H. (2008). Exploring the utility of three plastid loci for biocoding the filmy ferns (Hymenophyllaceae) of Moorea. Taxon, 57(3), 725-736.

Pang, X., Liu, C., Shi, L., Liu, R., Liang, D., Li, H., ... & Chen, S. (2012). Utility of the trnH–psbA intergenic spacer region and its combinations as plant DNA barcodes: a meta-analysis. PloS one, 7(11), e48833.

Percy, D. M.; Argus, G. W.; Cronk, Q. C.; Fazekas, A. J.; Kesanakurti, P. R.; Burgess, K. S.; Husband, B. C.; Newmaster, S. G.; Barrett, S. C. H., and Graham, S. W. 2014. Understanding the spectacular failure of DNA barcoding in willows (Salix): Does this result from a trans-specific selective sweep? Molecular Ecology 23:4737-4756.

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(1), 72-83.

Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). Molecular ecology notes, 7(3), 355-364.

Smith, S. A., & Donoghue, M. J. (2008). Rates of molecular evolution are linked to life history in flowering plants. science, 322(5898), 86-89.

Tan, S. L., Luo, Y. H., Hollingsworth, P. M., Burgess, K. S., Xu, K., Li, D. Z., & Gao, L. M. (2018). DNA barcoding herbaceous and woody plant species at a subalpine forest dynamics plot in Southwest China. Ecology and Evolution, 8(14), 7195-7205.

Tate, J. A., & Simpson, B. B. (2003). Paraphyly of Tarasa (Malvaceae) and diverse origins of the polyploid species. Systematic Botany, 28(4), 723-737.

The Catalogue of Life Partnership (2017). APG IV: Angiosperm Phylogeny Group classification for the orders and families of flowering plants. Checklist dataset https://doi.org/10.15468/fzuaam accessed via GBIF.org on 2022-02-21.

Thomas, W. W. 1999. Conservation and monographic research on the flora of Tropical America. Biodiversity & Conservation 8:1007-1015.

Toledo-Aceves, T.; Meave, J. A.; González-Espinosa, M., and Ramírez-Marcial, N. 2011. Tropical montane cloud forests: Current threats and opportunities for their conservation and sustainable management in Mexico. Journal of Environmental Management 92:974-981.

Trujillo-Argueta, S., Del Castillo, R. F., Tejero-Diez, D., Matias-Cervantes, C. A., & Velasco-Murguía, A. (2021). DNA barcoding ferns in an unexplored tropical montane cloud forest area of southeast Oaxaca, Mexico. Scientific reports, 11(1), 1-12.

Villaseñor, J. (2010). El bosque húmedo de montaña en México y sus plantas vasculares: catálogo florístico-taxonómico. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad- Universidad Nacional Autónoma de México. 40 pp. México, D.F.

Table 1(on next page)

Studied species of the tropical montane cloud forest of San Miguel Cuevas Juxtlahuaca, Oaxaca, Mexico.

Each column indicates sample ID, family, morphological species IUCN status, bp length, BOLD Process ID, and GenBank accession number.

Table 1. Studied species of the tropical montane cloud forest of San Miguel Cuevas Juxtlahuaca, Oaxaca, Mexico. Each column indicates sample ID, family, morphological species IUCN status, bp length, BOLD Process ID, and GenBank accession number

Sample ID	Family	Morphological species	IUCN status	bp lenght	Process ID	GenBank Accession No.
SMC7	Rubiaceae	*Deppea grandiflora* Schltdl.	LC	553	TFOAX001-19	ON002500
SMC10	Styracaceae	*Styrax glabrescens* Benth.	DD	535	TFOAX002-19	ON002540
SMC19	Thymelaeaceae	*Daphnopsis selerorum* Gilg	LC	535	TFOAX003-19	ON002498
SMC28	Solanaceae	*Solanum nigricans* M. Martens & Galeotti	LC	553	TFOAX004-19	ON002539
SMC29	Polygalaceae	*Monnina xalapensis* Kunth	LC	553	TFOAX005-19	ON002512
SMC37	Ericaceae	*Comarostaphylis longifolia* (Benth.) Klotzsch	DD	540	TFOAX006-19	ON002495
SMC41	Rubiaceae	*Deppea grandiflora* Schltdl.	LC	553	TFOAX007-19	ON002501
SMC61	Euphorbiaceae	*Tragia aff. nepetifolia* Cav.	DD	541	TFOAX008-19	ON002541
SMC70	Rubiaceae	*Hoffmannia longipetiolata* Pol.	DD	553	TFOAX009-19	ON002505
SMC75	Olaceae	*Osmanthus americanus* (L.) Benth. & Hook. f. ex A. Gray	DD	553	TFOAX010-19	ON002521
SMC76	Commelinaceae	*Commelina coelestis* Willd.	DD	553	TFOAX011-19	ON002497
SMC94	Thymelaeaceae	*Daphnopsis tuerckheimiana* Donn. Sm.	NT	540	TFOAX012-19	ON002499
SMC97	Rubiaceae	*Deppea guerrerensis* Dwyer & Lorence	DD	536	TFOAX013-19	ON002502
SMC99	Solanaceae	*Physalis philadelphica* Lam.	LC	535	TFOAX014-19	ON002525
SMC104	Apocinaceae	*Vallesia aurantiaca* (M. Martens & Galeotti) J.F. Morales	DD	553	TFOAX015-19	ON002545
AVM2	Betulaceae	*Alnus acuminata* Kunth	LC	560	DVHTF001-19	ON002486
AVM12	Berberidaceae	*Berberis lanceolata* Benth.	DD	557	DVHTF002-19	ON002489
AVM13	Araliaceae	*Oreopanax sanderianus* Hemsl.	VU	557	DVHTF003-19	ON002519
AVM14	Scrophulariaceae	*Buddleja cordata* Kunth	LC	557	DVHTF004-19	ON002490
AVM15	Convolvulaceae	*Ipomoea elongata* Choisy	DD	557	DVHTF005-19	ON002506
AVM16	Solanaceae	*Solanum hispidum* Pers	DD	553	DVHTF006-19	ON002537
AVM17	Solanaceae	*Solandra maxima* (Sessé & Moc.) P.S. Green	DD	558	DVHTF007-19	ON002536
AVM27	Commelinaceae	*Commelina coelestis* Willd.	DD	557	DVHTF008-19	ON002496
AVM30	Lycopodeaceae	*Lycopodium clavatum* L.	LC	560	DVHTF009-19	MZ771330

AVM32	Primulaceae	*Myrsine juergensenii* (Mez) Ricketson & Pipoly	LC	557	DVHTF010-19	ON002516
AVM33	Ericaceae	*Vaccinium leucanthum* Schltdl.	DD	558	DVHTF011-19	ON002544
AVM34	Fagaceae	*Quercus martinezii* C.H. Mull.	LC	560	DVHTF012-19	ON002531
AVM35	Fagaceae	*Quercus laurina* Bonpl.	LC	558	DVHTF013-19	ON002530
AVM36	Melastomataceae	*Miconia glaberrima* (Schltdl.) Naudin	LC	545	DVHTF014-19	ON002511
AVM40	Pteridaceae	*Adiantum andicola* Liebm.	DD	557	DVHTF015-19	ON002485
AVM43	Rosaceae	*Rubus sapidus* Schltdl.	DD	557	DVHTF016-19	ON002534
AVM45	Araliaceae	*Oreopanax xalapensis* (Kunth) Decne. & Planch.	LC	556	DVHTF017-19	ON002520
AVM46	Piperaceae	*Piper umbellatum* L.	DD	557	DVHTF018-19	ON002528
AVM47	Lauraceae	*Ocotea helicterifolia* (Meisn.) Hemsl.	DD	557	DVHTF019-19	ON002517
AVM48	Fabaceae	*Calliandra houstoniana* (Mill.) Standl.	LC	558	DVHTF020-19	ON002491
AVM50	Lauraceae	*Ocotea helicterifolia* (Meisn.) Hemsl.	DD	553	DVHTF021-20	ON002518
AVM52	Primulaceae	*Parathesis donnell-smithii* Mez	LC	553	DVHTF022-20	ON002522
AVM53	Fagaceae	*Quercus calophylla* Schltdl. & Cham.	LC	553	DVHTF023-20	ON002529
AVM54	Euphorbiaceae	*Cnidoscolus aconitifolius* (Mill.) I.M. Johnst.	LC	553	DVHTF024-20	ON002494
AVM59	Amaranthaceae	*Iresine diffusa* Humb. & Bonpl. ex Willd.	DD	553	DVHTF025-20	ON002507
AVM62	Solanaceae	*Solanum nigricans* M. Martens & Galeotti	LC	553	DVHTF026-20	ON002538
AVM64	Rubiaceae	*Arachnothryx buddleioides* (Benth.) Planch.	LC	553	DVHTF027-20	ON002488
AVM65	Passifloraceae	*Passiflora quadraticordata* Lozada-Pérez	DD	553	DVHTF028-20	ON002524
AVM67	Urticaceae	*Urera killipiana* Standl. & Steyerm.	LC	553	DVHTF029-20	ON002543
AVM69	Rubiaceae	*Hoffmannia longipetiolata* Pol.	DD	553	DVHTF030-20	ON002504
AVM71	Gesneriaceae	*Moussonia deppeana* (Schltdl. & Cham.) Hanst.	DD	553	DVHTF031-20	ON002515
AVM73	Pinaceae	*Pinus montezumae* Lamb.	LC	553	DVHTF032-20	ON002526
AVM78	Dicksoniaceae	*Lophosoria quadripinnata* (J.F. Gmel.) C. Chr.	DD	553	DVHTF033-20	ON002509
AVM79	Passifloraceae	*Passiflora quadraticordata* Lozada-Pérez	DD	553	DVHTF034-20	ON002523
AVM80	Cyperaceae	*Rhynchospora aristata* Boeck.	DD	553	DVHTF035-20	ON002532
AVM82	Marattiaceae	*Marattia weinmanniifolia* Liebm.	DD	553	DVHTF036-20	ON002510
AVM83	Cupressaceae	*Juniperus flaccida* Schltdl.	LC	553	DVHTF037-20	ON002508
AVM84	Pinaceae	*Pinus pseudostrobus* Lindl.	LC	553	DVHTF038-20	ON002527

AVM87	Asteraceae	*Montanoa tomentosa* Cerv.	DD	553	DVHTF039-20	ON002513
AVM89	Meliaceae	*Guarea glabra* Vahl	LC	553	DVHTF040-20	ON002503
AVM90	Meliaceae	*Trichilia havanensis* Jacq.	LC	553	DVHTF041-20	ON002542
AVM93	Rubiaceae	*Arachnothryx buddleioides* (Benth.) Planch.	LC	553	DVHTF042-20	ON002487
AVM95	Gesneriaceae	*Moussonia deppeana* (Schltdl. & Cham.) Hanst.	DD	553	DVHTF043-20	ON002514
AVM98	Solanaceae	*Cestrum commune* C.V. Morton ex Mont.-Castro	DD	553	DVHTF044-20	ON002492
AVM100	Poaceae	*Zeugites hintonii* Hartley	DD	553	DVHTF045-20	ON002546
AVM101	Asteraceae	*Roldana angulifolia* (DC.) H.Rob. & Brettell	DD	553	DVHTF046-20	ON002533
AVM103	Verbenaceae	*Citharexylum hexangulare* Greenm.	DD	474	DVHTF047-20	ON002493
AVM106	Lamiaceae	*Salvia clarkcowanii* B.L. Turner	DD	553	DVHTF048-20	ON002535

3 Least Concern (LC), Data Deficit (DD), Near Threatened (NT) and Vulnerable (VU).

Table 2(on next page)

Best BLASTn match found on queries against *rbcL* nucleotides sequences in the database of GenBank for those species with previously published sequences in the GenBank.

Sample ID Morphological species Identity (%) Best BLAST match accession number are shown.

1

2 Table 2. Best BLASTn match found on queries against *rbcL* nucleotides sequences in the
3 database of GenBank for those species with previously published sequences in the GenBank.

Sample ID	Morphological species	Identity (%)	Best BLAST match	Accession No.
AVM2	*Alnus acuminata*	100	Δ *Alnus nepalensis*	NC_039991.1
SMC7	*Deppea grandiflora*	99.46	*Cosmibuena grandiflora*	AM117220.1
AVM12	*Berberis lanceolata*	99.82	*Berberis thunbergii*	KX162895.1
AVM16	*Solanum hispidum*	100	*Solanum torvm*	MN218087.1
SMC29	*Monnina xalapensis*	**100**	***Monnina xalapensis***	AM234184.1
AVM30	*Lycopodium clavatum*	100	*Lycopodium clavatum*	KF977478.1
AVM35	*Quercus laurina*	100	*Δ Quercus phillyraeoides*	NC_048488.1
AVM40	*Adiantum andicola*	99.82	○ *Adiantum feei*	MH019567.1
AVM46	*Piper umbellatum*	100	○ *Piper umbellatum*	KF496838.1
AVM52	*Parathesis donnell-smithii*	100	○ *Stylogyne longifolia*	MF786262.1
AVM53	*Quercus callophylla*	100	Δ *Quercus phillyraeoides*	NC_048488.1
AVM54	*Cnidoscolus aconitifolius*	**100**	***Cnidoscolus aconitifolius***	MZ045411.1
AVM59	*Iresine diffusa*	**100**	***Iresine diffusa***	JQ590112.1
AVM64	*Arachnothryx buddleioides*	100	○ *Arachnothryx monteverdensis*	JQ594656.1
AVM70	*Hoffmannia longipetiolata*	99.64	○ *Omiltemia filisepala*	AM117251.1
AVM73	*Pinus montezumae*	100	Δ *Pinus arizonica*	KC156714.1
SMC75	*Osmanthus americanus*	99.81	Δ *Osmanthus americanus*	NC_048503.1
AVM78	*Lophosoria quadripinnata*	**99.64**	***Lophosoria quadripinnata***	MW138175.1
AVM80	*Rhynchospora aristata*	99.82	○ *Rhynchospora sp*	JQ594519.1
AVM82	*Marattia weinmanniifolia*	100	○ *Marattia douglasii*	MT657852.1
AVM83	*Juniperus flaccida*	100	○ *Juniperus flaccida*	HM024304.1
AVM84	*Pinus pseudostrobus*	100	*Pinus flexilis*	MG215114.1
AVM87	*Montanoa tomentosa*	100	○ *Montanoa tomentosa*	MT189234.1
AVM89	*Guarea glabra*	100	*Ruagea pubescens*	MN454793.1
AVM90	*Trichilia havanensis*	99.82	Meliaceae* *sp.*	EU042974.1
SMC99	*Physalis philadelphica*	100	Δ *Physalis minima*	NC_048515.1

In **bold** morphological species corresponding with only the studied species in GenBank database.

○ two species more same percent identity

Δ > 30 species with the same high identity percent

* *T. havanensis* showed a best match with the GenBank published sequence of another unidentified species of the Meliaceae

4

5

Table 3(on next page)

Intergeneric genetic distances of the tropical montane cloud forest species in the Mixteca Baja, Oaxaca, Mexico, found in multi-species genera.

The Bold Process ID is shown below the scientific names.

Table 3. Intergeneric genetic distances of the tropical montane cloud forest species in the Mixteca Baja, Oaxaca, Mexico, found in multi-species genera. The Bold Process ID is shown below the scientific names.

Genus	Species comparision		Genetic divergence
Daphnopsis	*Daphnopsis tuerckheimiana* TFOAX012-19	*Daphnopsis selerorum* TFOAX003-19	0.0000
Deppea	*Deppea guerrerensis* TFOAX013-19	*Deppea grandiflora* TFOAX001-19	1.5076
Deppea	*Deppea guerrerensis* TFOAX013-19	*Deppea grandiflora* TFOAX007-19	1.5076
Oreopanax	*Oreopanax xalapensis* DVHTF017-19	*Oreopanax sanderianus* DVHTF003-19	0.0000
Pinus	*Pinus pseudostrobus* DVHTF038-20	*Pinus montezumae* DVHTF032-20	2.0189
Quercus	*Quercus laurina* DVHTF013-19	*Quercus martinezii* DVHTF012-19	0.0000
Quercus	*Quercus callophylla* DVHTF023-20	*Quercus martinezii* DVHTF012-19	0.0000
Quercus	*Quercus callophylla* DVHTF023-20	*Quercus laurina* DVHTF013-19	0.0000
Solanum	*Solanum hispidum* DVHTF006-19	*Solanum nigricans* TFOAX004-19	0.7269
Solanum	*Solanum nigricans* DVHTF026-20	*Solanum hispidum* DVHTF006-19	0.7269

Table 4(on next page)

Mean intergeneric and interspecific genetic distances in the multigenera and multispecies families of this study.

Table 4. Mean intergeneric and interspecific genetic distances in the studied multigenera and multispecies families of this study.

Family	No. Genera	No. Species	Mean Divergence (%)
Asteraceae	2	2	3.3344
Ericaceae	2	2	5.5723
Euphorbiaceae	2	2	4.5869
Meliaceae	2	2	0.7286
Primulaceae	2	2	1.2784
Rubiaceae	3	4	1.6207
Solanaceae	4	5	1.6558

Table 5(on next page)

Monophyly tree-based results observed using *rbcLa*.

Proportion of resolved monophyletic species and support value obtained (bootstrap /posterior probabilities) with different phylogenetic techniques using plant core barcoding gene *rbcLa* in 55 studied species of the tropical montane cloud forest, Mixteca Baja, Mexico.

Table 5. Monophyly tree-based results observed using *rbcLa*. Proportion of resolved monophyletic species and support value obtained (bootstrap /posterior probabilities) with different phylogenetic techniques using plant core barcoding gene *rbcLa* in 55 studied species of the tropical montane cloud forest, Mixteca Baja, Mexico.

	Neighbor Joining tree	Maximum Likelihood tree	Bayesian Inference tree
Percent of monophyletic species resolved	100.00	100.00	85.45
Percent species resolved with support value > 70%	70.91	87.27	92.73

Figure 1

Graphic results of BLASTn analysis and genus and species resolution

Fraction of the new *rbcL* sequences published in the GenBank (left), and resolution at genus and species levels for those species with previously available sequences released in the GenBank (accessed in March 2022) using BLASTn (right).

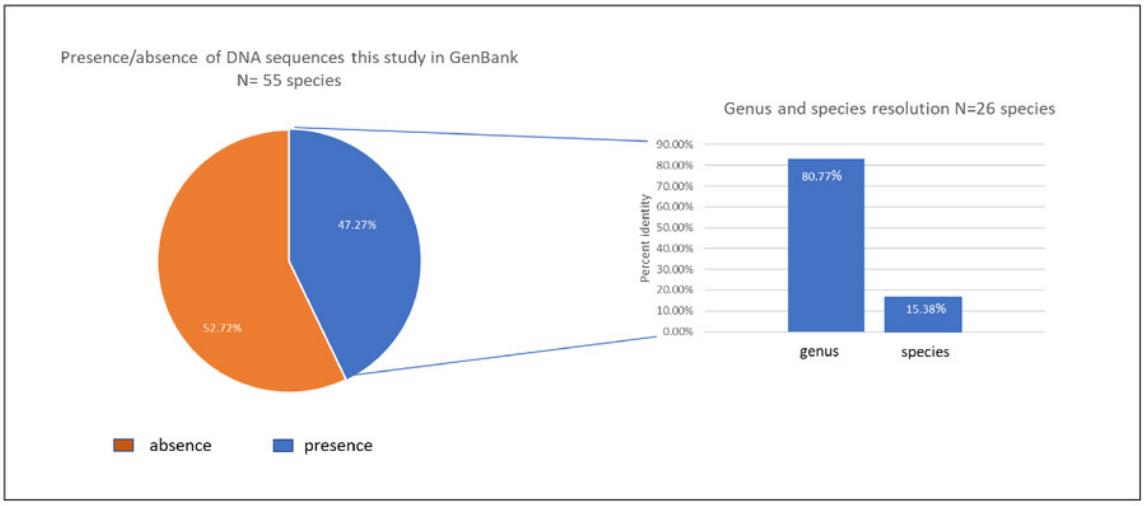

Figure 1. Graphic results of BLASTn analysis.

Figure 2

rbcLa intra- and interspecific genetic distances between 55 plant species from 38 families of tropical montane cloud forest in the Mixteca Baja, Oaxaca, Mexico.

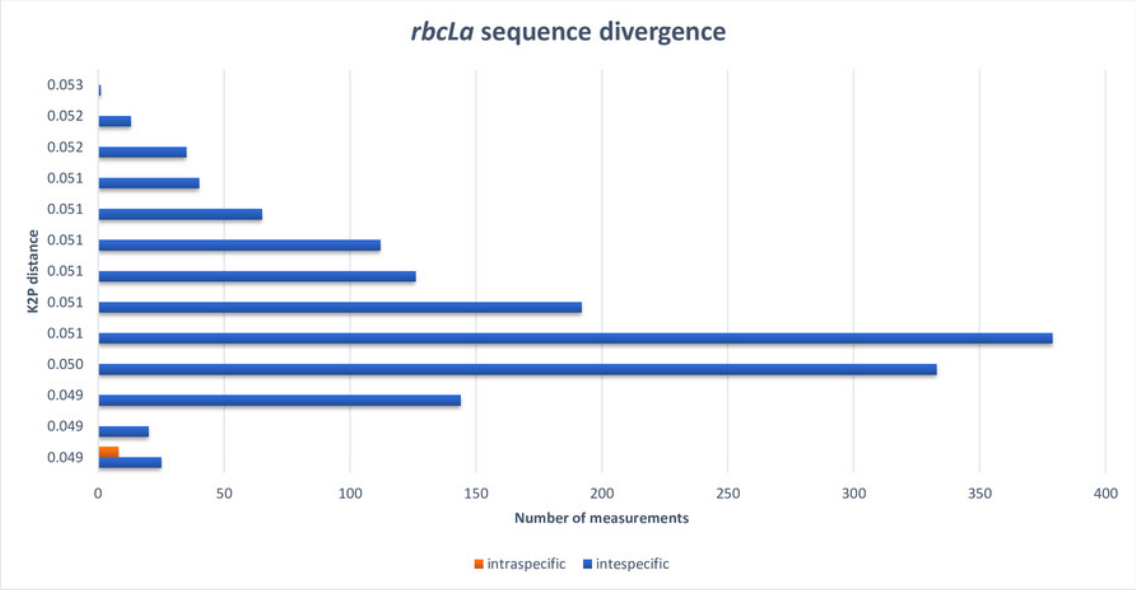

Figure 2. Intraspecific and interspecific genetic distances obtained in 55 species of plants of the Mixteca Baja, Oaxaca, México

Figure 3

Maximum Likelihood phylogenetic tree obtained with barcode *rbcLa*.

Monophyletic species resolution of 55*rbcLa* plant sequences obtained from maximum likelihood phylogenetic tree analysis of tropical montane cloud forest species, Mixteca Baja, Oaxaca, Mexico. Bootstrap values based on 1000 replications are listed as percentages at branching points.

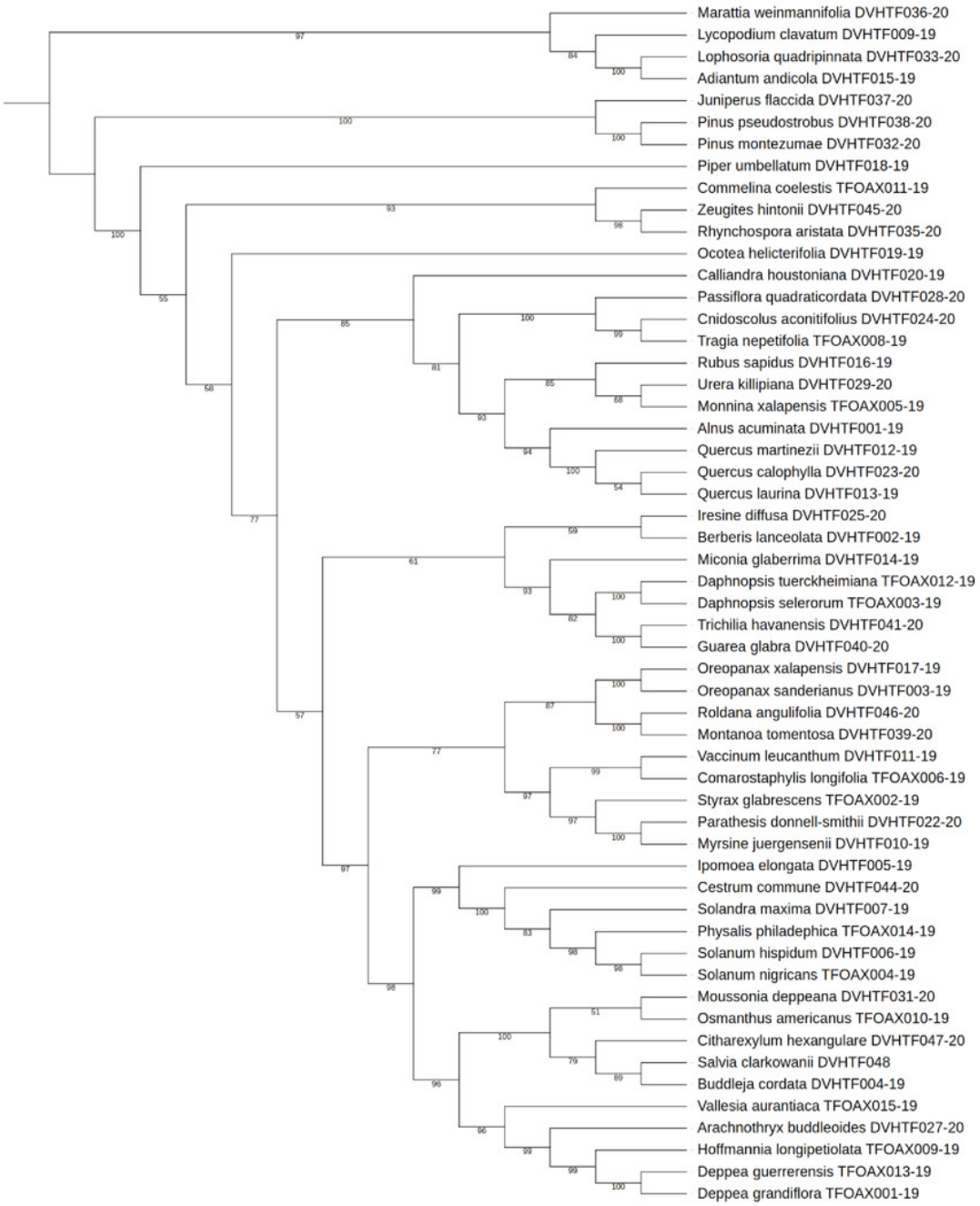

Figure 3. Maximum Likelihood phylogenetic tree obtained with barcode *rbcLa*.