HAYATI
Journal of Biosciences

Vol. 30 No. 3, May 2023 543-550
DOI:10.4308/hjb.30.3.543-550
ISSN: 1978-3019
EISSN: 2086-4094

An Early Reference to DNA Barcode for the Anacardiaceae Family

Alberto Ryadi[1], Iskandar Siregar[1], Carina de Melo Moura[2], Oliver Gailing[2], Fitri Yola Amandita[3*]

[1]Department of Silviculture, Faculty of Forestry and Environment, IPB University, Bogor 16680, Indonesia
[2]Department of Forest Genetics and Forest Tree Breeding, University of Goettingen, Germany
[3]Research Center for Environment and Clean Technology, BRIN, South Tangerang, Indonesia

ARTICLE INFO

Article history:
Received August 15, 2022
Received in revised form January 16, 2023
Accepted January 25, 2023

KEYWORDS:
matK,
molecular identification,
phylogenetic tree,
rbcL,
species identification

ABSTRACT

Anacardiaceae is well-known for its edible fruits and economically important species in Indonesia. Approximately 3% of Indonesia's endangered and vulnerable species belong to this family. Fast and accurate species identification is crucial to support the conservation efforts for this family, such as employing DNA sequences. Species identification using DNA sequences, known as DNA barcoding, has been widely used in many applied fields. So far, the application of DNA barcoding for Anacardiaceae plant species is limited to several genera only, such as Mangifera and Spondias. This research aimed to enrich the DNA barcode references of Anacardiaceae and to evaluate the most suitable and effective genetic marker as DNA barcodes to identify species of 35 samples representing sixteen species of the Anacardiaceae family using chloroplast markers *matK* and *rbcL* as barcode regions. When comparing the morphological identification with the molecular assignments, the barcode accuracy was 62.50% (*matK*), 60.61% (*rbcL*), and 73.33% (*matK*+*rbcL*) at the genus level. All the markers failed to show a barcoding gap, even though the t-test showed that the intraspecific and interspecific genetic distances were significantly different for *matK* and *rbcL*+*matK*. Among others, *Gluta walichii* (Hook.f.) Ding Hou and *Melanochyla caesia* Jack were the only species successfully resolved by all markers. Nevertheless, new DNA barcodes of six Anacardiaceae species were made available by this study, enriching the genetic references of tropical flora diversity.

1. Introduction

Anacardiaceae is a common tree, shrub, or climber, often with resinous bark, including circa 70 genera and 600 species distributed throughout the tropics and subtropics (Hou 1974). The family is well-known for its edible and commercially used seeds and fruits, such as cashew, pistachio, and mango (Pell 2004). Many produce valuable tannins (Araújo *et al.* 2012; da Costa *et al.* 2018; Sobeh *et al.* 2018). Based on IUCN data (2022), two out of 443 Anacardiaceae species recorded globally, and even two species endemic to Borneo are already extinct in the wild (*Mangifera casturi* Kosterm. and *M. rubropetala* Kosterm.), four species are endangered and eleven species are vulnerable in Indonesia. The rich diversity of the Anacardiaceae family provides vital ecosystem services, such as food, yet genetic information for its species conservation and utilization is insufficient (Labdelli *et al.* 2020).

The current problem related to Anacardiaceae's conservation effort is to obtain a quick, precise, and accurate morphological identification of the specimens (Herrera *et al.* 2018; Santos *et al.* 2021). The morphology-based plant species identification was carried out by analyzing all parts of intact specimens, including flowers and fruits, which are seasonally available (Hassoon *et al.* 2018; Wäldchen *et al.* 2018). Major challenges of morphological identification, especially in tropical regions such as Indonesia, involve the limited access to herbarium collections, a restricted number of literature studies, and the comparatively low number of qualified taxonomists, making the accurate identification process time-consuming and inefficient (Gaston and O'Neill 2004; Goodwin *et al.* 2015; Mata-Montero and Carranza-Rojas 2016). Therefore, alternative methods that can provide faster and more precise species identification are advantageous to enhance the exploration of the Anacardiaceae family.

* Corresponding Author
E-mail Address: fitri.yola.amandita@brin.go.id

DNA barcoding is a molecular method widely utilized to quickly identify species using a short DNA sequence extracted from the organism's tissue (Hebert *et al.* 2003; Kress and Erickson 2012). Many studies have reported the success of using DNA barcoding to identify animal species, such as fishes (Ghouri *et al.* 2020), insects (Javal *et al.* 2021), birds (Tizard *et al.* 2019), reptiles (Nagy *et al.* 2012), and mammals (Galimberti *et al.* 2015). Regarding plant identification, the application of DNA barcoding is constrained by the absence of a universal diagnostic marker (CBOL Plant Working Group 2009). However, the markers maturase-K (*matK*) and ribulase-1,3-biphosphate carboxylase oxygenase (*rbcL*) are the recommended barcodes most used for plants (Kress *et al.* 2005; Lahaye *et al.* 2008; Hollingsworth *et al.* 2011). The effectiveness of using these two markers for species identification has been reported by Amandita *et al.* (2019), in which more than 700 flowering plant species were included in the study. Moreover, matK is considered a marker with the highest species discrimination for plant identification (Cowan and Fay 2012). Meanwhile, *rbcL* has been widely used to retrace plant group evolutionary relationships (Omonhinmin and Onuselogu 2022); thus, these two markers and/or a combination of them are highly advised.

So far, the reports on using DNA barcoding for Anacardiaceae species are very limited and focus only on a few genera, such as *Mangifera* (Hidayat *et al.* 2012) and *Spondias* (Silva *et al.* 2015). In this study, nine genera of Anacardiaceae plants, namely *Mangifera, Buchanania, Melanochyla, Dracontomelon, Semecarpus, Campnosperma, Parishia, Drimycarpus,* and *Baouea*, collected from Indonesia's lowland forest were analyzed to test the effectiveness and accuracy of the matK and rbcL as DNA barcoding markers for the identification of these plant species.

2. Materials and Methods

2.1. DNA Barcode Datasets

The data used in this study were matK and rcbL sequences of Anacardiaceae plant species, consisting of sixteen species (Table 1) obtained from "plant DNA barcoding" research activities in the CRC990-EFForTS project (https://www.uni-goettingen.de/efforts). The Anacardiaceae plant samples were collected from 50 m × 50 m inventory plots in and around the area of Bukit Duabelas National Park (1°51'S 102°39'E)

Table 1. List of Anacardiaceae plant species included in the current study

Species name	Conservation status*	Available reference sequences**	Number of samples	Sample ID
Mangifera indica L.	Data deficient	*matK, rbcL*, ITS	3	KR0618, KR5174, KR5172
M. torquenda Kosterm.		ITS	2	KR5176, KR5128
M. foetida Lour.	Unspecified	*matK, rbcL, trnH-psbA*, ITS	2	KR5122, KR5459
M. caesia Jack	Near threatened	-	1	KR4544
M. laurina Blume	Unspecified	*matK, rbcL ITS, trnH-psbA*	2	KR1894, KR5348
Buchanania sessilifolia Blume	Unspecified	*matK, rbcL*	5	KR4735, KR2825, KR1967, KR1966, KR4934
Melanochyla caesia (Blume) Ding Hou	Vulnerable	*matK*	2	KR4964, KR1613
M. beccariana Oliv.	Unspecified	-	1	KR4575
M. tomentosa Hook.f.	Unspecified	-	1	KR3359
Dracontomelon dao (Blanco) Merr. & Rolfe	Least concern	*matK, rbcL, trnH-psbA*	4	KR2827, KR2397, KR4182, KR4281
Semecarpus cf. caesia Blume	Unspecified	*matK*	2	KR3418, KR3323
Campnosperma auriculatum (Blume) Hook.f.	Least concern	*matK, rbcL*	2	KR2617, KR1904
Gluta wallichii (Hook.f.) Ding Hou	Least concern	*matK, rbcL, trnH-psbA*	2	KR1386, KR1363
Parishia insignis Hook.f.	Least concern	-	1	KR0486 (matK), KR1377 (rbcL)
Drimycarpus luridus (Hook.f.) Ding Hou	Unspecified	-	1	KR0698
Bouea macrophylla Griff.	Unspecified	-	-	KR4432, KR4433

*The conservation status according to the IUCN Red List database (2022), **The availability of reference sequences according to the NCBI database (http://www.ncbi.nlm.nih.gov/)

and Harapan Rainforest (2°14'S 103°19'E) in Jambi province, Sumatra, Indonesia.

From each sample, leaf tissues (approximately 2 cm^2) were collected freshly and dried in silica gel for DNA analysis. Herbarium vouchers were prepared and stored in Herbarium Bogoriensis and BIOTROP Herbarium, Bogor, Indonesia. Several high-quality photographs of the herbarium vouchers were taken for further identification and to be uploaded along with the DNA barcodes to the DNA barcoding database (https://www.boldsystems.org). The DNA sequences were extracted, amplified, and sequenced at the Department of Forest Genetics and Forest Tree Breeding, Goettingen University, Germany, following Amandita *et al.* (2019).

The morphological identification of the samples included in this study was conducted by taxonomists affiliated with the CRC990-EFForTS project and had been reported by Rembold *et al.* (2017). For the data analysis, all the samples in this study were referred to the species name based on morphological identification. The availability of reference sequences for each species in this study was checked in the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/), as shown in Table 1.

2.2. Editing and Alignment

Forward and reverse sequences of each sample were combined into a consensus sequence (contig) using CodonCode Aligner software (CodonCode Corp., USA). The concatenated sequences were saved as nexus or fasta files, which were further converted into mega files using the MEGA11 software (Tamura *et al.* 2021). Sequence alignments were performed using the ClustalW algorithm in MEGA11 software (Tamura *et al.* 2021). Furthermore, *matK* and *rbcL* alignments were combined to create multiple alignments for both *matK*+*rbcL* markers using SequenceMatrix software (Vaidya *et al.* 2011).

2.3. Identification Suitability Analysis

The identification suitability analysis was performed to compare the morphological and molecular identification results by comparing the aligned sequences to the NCBI (http://www.ncbi.nlm.nih.gov/) database using Basic Local Alignment Search Tools (BLAST) algorithm (Altschul *et al.* 1990). The best match for each sample was the reference sequence with the highest similarity score and the lowest e-value. The identification suitability was measured by calculating the number of consistent morphological (the species name according to the data source) and molecular identifications (according to the BLAST result) at the family, genus, and species levels. When the result of BLAST with the highest score matched only the family but not the genus, it was counted to match only up to the family level; and if it matched the genus but not the species name, the identification was considered to be correct only up to the genus level. When the morphological identification did not match the BLAST result, it was determined as misidentified and was excluded from the dataset (Amandita *et al.* 2019).

2.4. Genetic Distance Analysis

Each marker's sequence alignment dataset was used to calculate intra- and interspecific genetic distances using MEGA11 software with Tamura 3 parameter model (Tamura *et al.* 2021). The intra- and interspecific genetic distances were depicted as histograms to evaluate the presence of a barcoding gap (Amandita *et al.* 2019) using Microsoft Excel ver. 16. Student' t-tests were performed to compare the mean values of intra- and interspecific distances of each marker, and ANOVA was done to compare the mean values of intra- and interspecific distances between the markers. All the statistical analyses were carried out using JMP Statistical software (SAS Institute Inc., USA).

2.5. Phylogenetic Tree Reconstruction

The phylogenetic analysis was conducted in this study to evaluate if *matK, rbcL*, and the combined barcodes could resolve the investigated taxa into appropriate taxonomical groups. The phylogenetic tree reconstruction was carried out using Maximum Likelihood as an accurate method for DNA sequence assignments (Hall 2005). The multiple alignments were uploaded to the MEGA11 as a fasta file into the Phylogeny menu. The calculation was done with 1,000 bootstrap replications using Tamura 3 parameter model, and the rate among sites was adjusted as Gamma distributed with invariant sites (G+I) (Tamura *et al.* 2021). *matK* and *rbcL* sequences of *Dacryodes laxa* (A.W. Benn) H.J. Lam, *D. rostrata* (Blume) H.J. Lam, and *D. rugosa* (Blume) H.J. Lam of the family Burseraceae, which is one of the closest group to the Anacardiaceae (Pell 2004), were included in the reconstruction of phylogenetic trees as the outgroup.

3. Results

3.1. Comparison of Molecular and Morphological Identification

Table 2 shows the identification suitability between morphological identification and molecular identification. Out of 35 samples, only five samples were correctly identified up to the species level using matK, namely KR4735 and KR4934 (*Buchanania sessilifolia*), KR1386 and KR1363 (*Gluta wallichii*), and KR2617 (*Campnosperma auriculatum*). None of the samples was correctly identified up to the species level by *rbcL*, and only one sample (KR4934) was correctly assigned to the species level by molecular identification using the combined marker set, *matK+rbcL*.

3.2. Barcoding Gap

The distribution of intra- and interspecific genetic distances of the two markers (*matK, rbcL*) and their combination (Figure 1) shows that intra- and interspecific distances are overlapping, meaning no barcoding gaps were formed by the dataset.

All the minimum values of interspecific distances calculated here are equal to the minimum values of intraspecific distances, which are zero (Table 3). These null values of the interspecific distances were present when pairing two species with identical sequences. Furthermore, the *rbcL* sequences tend to have a wider overlap between the intra- and interspecific distances compared to *matK* and *matK+rbcL*. The t-test shows that the intra- and interspecific distances significantly differed when *matK* and *matK+rbcL* markers were used but not with *rbcL*. Meanwhile, a one-way ANOVA revealed significant differences in interspecific distances among the markers.

3.3. Phylogenetic Analysis

The phylogenetic relationships between sixteen species of Anacardiaceae were estimated based on DNA barcodes of *matK, rbcL*, and the combined dataset (*matK+rbcL*) using the Maximum Likelihood

Table 2. Identification of suitability percentage based on taxonomical rank

Identification match	*matK* (%)	*rbcL* (%)	*matK+rbcL (%)*
Up to species level	15.63	0.00	3.33
Up to genus level	62.50	60.61	73.33
Up to family level	21.87	39.39	23.34

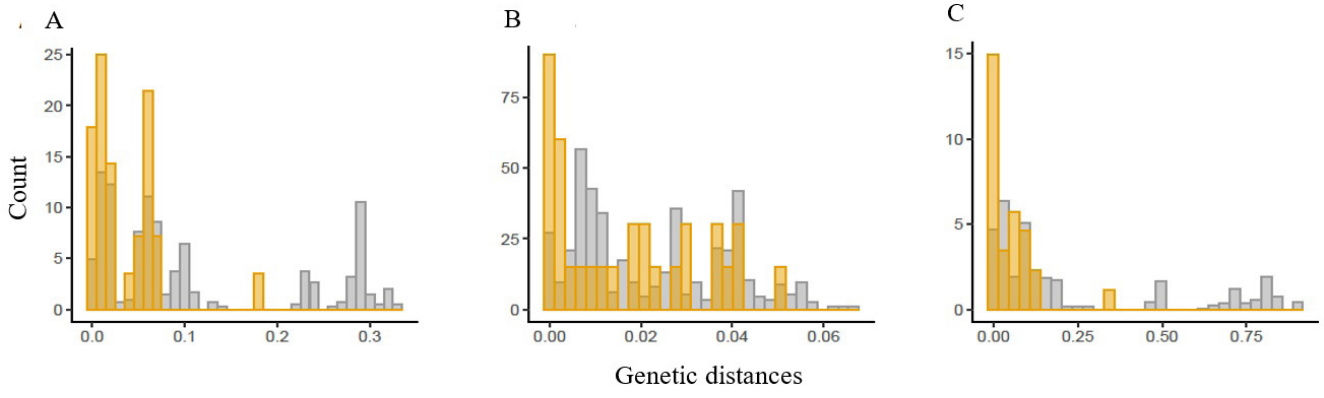

Figure 1. Distribution of intraspecific (yellow) and interspecific (grey) genetic distances for markers (A) *matK+rbcL*, (B) *rbcL*, and (C) *matK*.

Table 3. Intra- and interspecific distances of each marker and mean comparison tests

Marker	Intraspecific			Intraspecific			t-test p-value
	Min.	Max.	Mean (SD)	Min.	Max.	Mean (SD)	
matK	0.0000	0.3474	0.0541 (0.0714)	0.0000	0.9021	0.2268 (0.2841)	<0.05*
rbcL	0.0000	0.0503	0.0168 (0.0158)	0.0000	0.0666	0.0225 (0.0163)	>0.05
matK+rbcL	0.0000	0.1756	0.0342 (0.0370)	0.0000	0.3311	0.1070 (0.1058)	<0.05*
ANOVA p-value	>0.05			<0.05*			

*: mean values are significantly different, SD: standard deviation

tree approach (Figure 2). Most bootstrap values, especially in *matK* and *rbcL* trees, were higher than 70%, indicating that the accuracy and stability of the topologies are high. A low resolution was detected for Mangifera spp. branches in the *matK+rbcL* tree, while it seems non-monophyletic in the *matK* tree. Samples of *Gluta wallichii* and *Melanochyla caesia* were found to be monophyletic in all trees. Meanwhile, the other species were non-monophyletic, except *Bouea macrophylla* in the *rbcL* tree. Furthermore, in *matK* and *rbcL* trees, samples of *Semecarpus cf. caesia, Melanochyla tomentosa, M. beccariana*, and *Dracontomelon dao* were clustered together.

4. Discussion

One way to determine the accuracy level of DNA barcodes is by comparing the molecular identification resulting from BLAST searches against reference databases with morphological identification (Meier *et al.* 2006). A matched identification was highest at the genus level for all three markers (Table 2), as also reported by other studies (Amandita *et al.* 2019; Wati *et al.* 2022). Meanwhile, at the family level, at least 20% of the samples were correctly identified as Anacardiaceae, but assigned to different genera; for example, the sequences of Dracontomelon dao (KR2827, KR2397) were assigned to *Choerospondias axillaris* with 97% similarity by using *matK* and *matK+rbcL*, or as *Spondias pinnata* with 98% similarity by *rbcL*. A similar case was detected for sequences of *Semecarpus cf. caesia* (KR3418 and KR3323) assigned to *Melanochyla fulvinervia* with 98% similarity using matK and *matK+rbcL*, while sequences of *Melanochyla* spp. (KR4964, KR1613, KR4575, and KR3359) were assigned to *Semecarpus reticulatus* with >98% similarity by using *rbcL*. Moreover, the *rbcL* sequences of two samples (KR0486 and KR0698) were identified as species belonging to other families, probably due to contamination or mislabeling during the lab analysis; thus, these sequences were excluded from the dataset.

One factor that might be the cause of the unsuitability between morphological and molecular identification in this study is the absence of reference sequences in databases, as pointed out by many studies, e.g., Nilsson *et al.* (2006), Virgilio *et al.* (2012), and Qing *et al.* (2020). Table 1 shows the reference sequences for six species (*Melanochyla caesia, M. beccariana, M. tomentosa, Parishia insignis, Drimycarpus luridus*, and *Bouea macrophylla*) were not available at all in the NCBI database. In contrast, the reference sequences for *Mangifera torquenda* were limited only to the

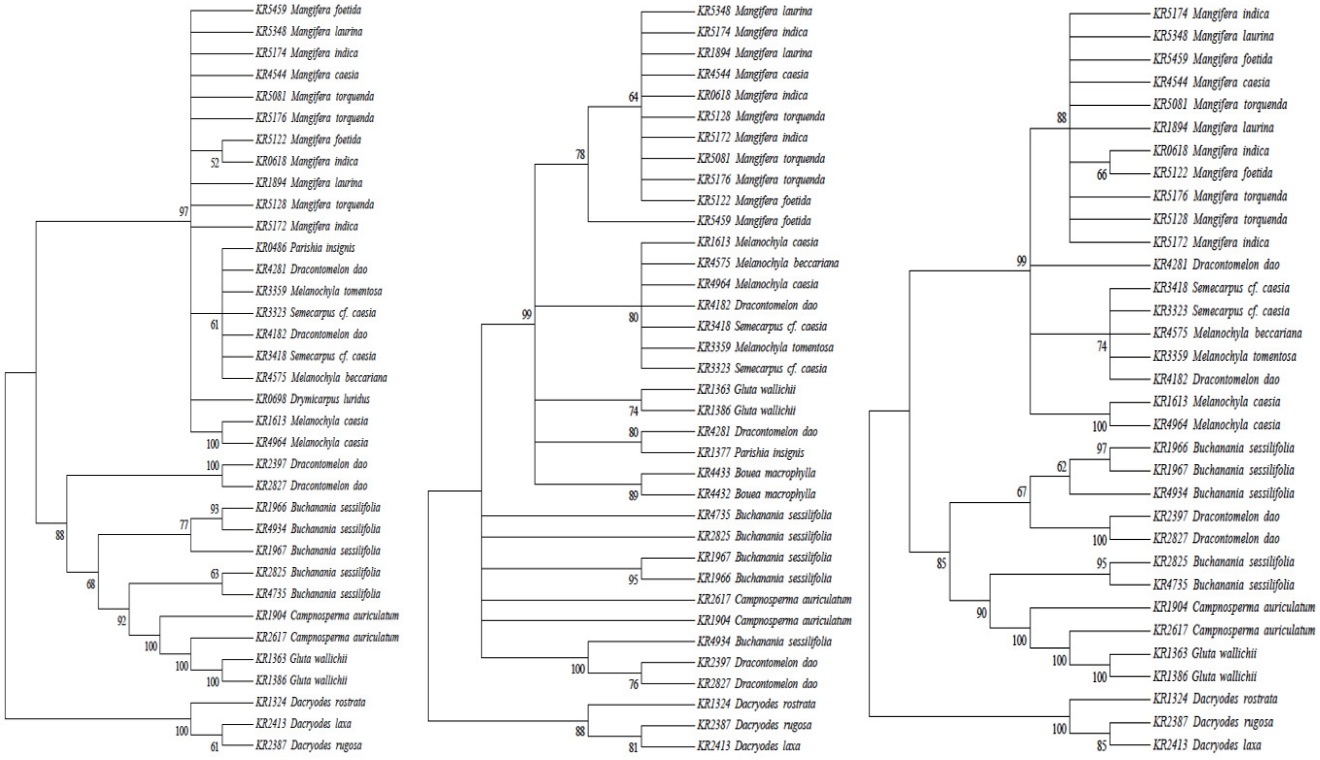

Figure 2. Maximum likelihood tree of Anacardiaceae species based on *matK* (left), *rbcL* (center), and *matK+rbcL* (right)

Inter Transcribed Spacer (ITS) region. Several studies (Bruni *et al.* 2012; Burgess *et al.* 2011; Cowan and Fay 2012) stated that molecular species assignment using NCBI and the Barcode of Life Data system (BOLD, https://www.boldsystems.org/) is greatly influenced by the level of reference sequence availability in such databases. As a result of this study's reference sequences of *Melanochyla caesia, M. beccariana, M. tomentosa, Parishia insignis, Drimycarpus luridus*, and *Bouea macrophylla* are now available in the database mentioned above.

An ideal DNA barcode can be determined by the presence of a barcoding gap, which occurs when the maximum intraspecific distance is lower than the minimum value of the interspecific distance (Meyer and Paulay 2005). The presence of a barcoding gap in plant DNA barcodes has been reported in different studies. Gogoi and Bhau (2018) reported a barcoding gap for *Nepenthes* (Nepenthaceae) samples when using ITS and matK, but not with *rbcL*. Similar results were reported by Trujillo-Argueta *et al.* (2021) for *rbcL* that showed overlapping intra- and interspecific genetic distances. Candek and Kuntner (2014) imply that the barcoding gap size strongly depends on the taxonomic groups in question.

Nevertheless, Ross *et al.* (2008) suggested that the overlap of the distribution of intra- and interspecific distances was a poor predictor of identification success. According to Barret and Hebert (2005) and Liu *et al.* (2011), a marker with a wider overlap of intra- and interspecific distances is considered a less effective DNA barcode. In this regard, the current study's *rbcL* marker is considered less effective than *matK* and *matK+rbcL*.

The taxonomic resolution through phylogenetic analysis allows for assessing a certain marker's effectiveness as a DNA barcode (Kang *et al.* 2017). The clustering of different species within the same clade seems to be related to the mismatch between the morphological and molecular identification previously discussed; the samples of *Semecarpus* spp. were identified as *Melanochyla* spp., and vice versa, as well as four samples of *Dracontomelon dao* were inconsistently assigned to other Anacardiaceae species. Also, Ariyarathne *et al.* (2020) reported that *Semecarpus* is not a monophyletic group. *Semecarpus* and *Melanochyla* belong to the tribe *Semecarpeae*, and, as reported by Weeks *et al.* (2014), clustered next to each other. The failure to recover monophyletic clades for *Semecarpus* and *Melanochyla* indicates that the markers used in this study could not provide a reasonable distinction to discriminate the taxa properly. However, the most recent phylogenetic study on Anacardiaceae (Weeks *et al.* 2014) using the trnL intron and trnL-F intergenic spacer and rps16 in 67 genera, including Semecarpus and Melanochyla, revealed a clear resolution between both genera.

In conclusion, as an early reference to the use of DNA barcodes for Anacardiaceae species identification, this study provides preliminary information on the effectiveness of plant core barcodes, *matK* and *rbcL* and the combined *matK+rbcL*, in assigning samples at least to the genus level. Even though the markers failed to show a barcoding gap for the dataset, the intra- and interspecific genetic distances proved to be significantly different, except for *rbcL*. The phylogenetic analysis showed that neither *matK* nor *rbcL* provided a clear taxonomic resolution for investigated species of Anacardiaceae. Nevertheless, this study provides new reference sequences of six Anacardiaceae species, which may facilitate species assignment in floristic surveys. Further studies should include a reasonable number of samples representing the richness of the family in tropical regions and alternative markers to validate the efficacy of DNA barcoding. On a global scale, DNA barcoding has been proven to be a reliable method for biodiversity assessment and phylogenetic studies to support conservation efforts of valuable species.

Acknowledgements

The authors thank the DFG-funded project "CRC990-EFForTS Program" (project ID 192626868 – SFB 990) for supporting this study. The authors are also grateful for the technical assistance of Alexandra Dolynska and Larissa Kunz in the Department of Forest Genetics and Forest Tree Breeding, Goettingen University, Germany.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403-410. https://doi.org/10.1016/S0022-2836(05)80360-2

Amandita, F.Y., Rembold, K., Vornam, B., Rahayu, S., Siregar, I.Z., Kreft, H., Finkeldey, R., 2019. DNA barcoding of flowering plants in Sumatera, Indonesia. *Ecol. Evol.* 9, 1-11. https://doi.org/10.1002/ece3.4875

Araújo, T.A.S., e Castro, V.T.N.A., de Amorim, E.L.C., de Albuquerque, U.P., 2012. Habitat influence on antioxidant activity and tannin concentrations of *Spondias tuberosa*. *Pharm. Biol.* 50, 754-759. https://doi.org/10.3109/13880209.2011.630673

Ariyarathne, M., Yakandawala, D., Barfuss, M., Heckenhauer, J., Samuel, R., 2020. Molecular phylogeny and chromosomal evolution of endemic species of Sri Lankan Anacardiaceae. *J. Nat. Sci. Found. Sri Lanka*. 48, 289-303. https://doi.org/10.4038/jnsfsr.v48i3.9368

Barrett, R.D., Hebert, P.D.N., 2005. Identifying spider through DNA barcodes. *Can. J. Zool.* 83, 481-491. https://doi.org/10.1139/z05-024

Bruni, I., De Mattia, F., Martellos, S., Galimberti, A., Savadori, P., Casiraghi, M., Nimis, P.L., Labra, M., 2012. DNA barcoding as an effective tool in improving a digital plant identification system: a case study for the area of Mt. Valerio, Trieste (NE Italy). *PLoS ONE.* 7, e43256. https://doi.org/10.1371/journal.pone.0043256

Burgess, K.S., Fazekas, A.J., Kesanakurti, P.R., Graham, S.W., Husband, B.C., Newmaster, S.G., Percy, D.M., Hajibabaei, M., Barrett, S.C.H., 2011. Discriminating plant species in a local temperate flora using the rbcL plus matK DNA barcode. *Methods Ecol. Evol.* 2, 333-340. https://doi.org/10.1111/j.2041-210X.2011.00092.x

Candek, K., Kuntner, M., 2014. DNA barcoding gap: reliable species identification over morphological and geographical scales. *Mol. Ecol. Res.* 15, 268-277. https://doi.org/10.1111/1755-0998.12304

Cowan, R.S., Fay, M.F., 2012. Challenges in the DNA barcoding of plant material: plant DNA fingerprinting and barcoding. *Methods Mol. Biol.* 826, 23-33. https://doi.org/10.1007/978-1-61779-609-8_3

CBOL Plant Working Group. 2009. A DNA barcode for land plants. *Proc. Natl. Acad. Sci. USA.* 106, 12794-12797.

da Costa Cordeiro, B.M.P., de Lima Santos, N.D.,Ferreira, M.R.A., de Araújo, L.C.C.,Carvalho Junior, A.R.,da Conceição Santos, A.D., de Oliveira, A.P., da Silva, A.G., da Silva Falcão, A.P., Correia, M.T.D.S., da Silva Almeida, J.R.G., da Silva, L.C.N., Soares, L.A.L., Napoleão, T.H., da Silva, M.V., Paiva, P.M.G., 2018. Hexane extract from Spon*dias tuberosa* (Anacardiaceae) leaves has antioxidant activity and is an anti-Candida agent by causing mitochondrial and lysosomal damages. *BMC Complement Altern. Med.* 18, 284. https://doi.org/10.1186/s12906-018-2350-2

Galimberti, A., Sandionigi, A., Bruno, A., Bellati, A., Casiraghi, M., 2015. DNA barcoding in mammals: what's new and where next. *Hystrix*. 26, 1.

Gaston, K.J., O'Neill, M.A., 2004. Automated species identification: why not? *Phil. Trans. R. Soc. Lond. B.* 359, 655-667. https://doi.org/10.1098/rstb.2003.1442

Ghouri, M.Z., Ismail, M., Javed, M.A., Khan, S.H., Munawar, N., Umar, A.B., Nisa, M., Aftab, S.O., Amin, S., Khan, Z., Ahmad, A., 2020. Identification of edible fish species of Pakistan through DNA barcoding. *Front. Mar. Sci.* 7, 554183. https://doi.org/10.3389/fmars.2020.554183

Gogoi, B., Bhau, B.S., 2018. DNA barcoding of the genus *Nepenthes* (Pitcher plant): a preliminary assessment towards its identification. *BMC Plant Biol*. 18, 153. https://doi.org/10.1186/s12870-018-1375-5

Goodwin, Z.A., Harris, D.J., Filer, D., Wood, J.R.I., Scotland, R.W., 2015. Widespread mistaken identity in tropical plant collections. *Curr. Biol*. 25, 1057-1069. https://doi.org/10.1016/j.cub.2015.10.002

Hall, B.G., 2005. Comparison of the accuracies of several phylogenetic methods using protein and DNA sequences. *Mol. Biol. Evol*. 22, 792-802.

Hassoon, I.M., Kassir, S.A., Altaie, S.M., 2018. A review of plant species identification techniques. *Int. J. Sci. Res.* 7, 325-328.

Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R., 2003. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B.* 270, 313-321. https://doi.org/10.1098/rspb.2002.2218

Herrera, F., Mitchell, J.D., Pell, S.K., Collinson, M.E., Daly, D.C, Manchester, S.R., 2018. Fruit morphology and anatomy of the Spondioid Anacardiaceae. *Bot. Rev.* 84, 315-393. https://doi.org/10.1007/s12229-018-9201-1

Hidayat, T., Pancoro, A., Kusumawaty, D., Eiadthong, W., 2012. Development *matK* gene as DNA barcode to assess evolutionary relationship of important tropical forest tree genus *Mangifera* (Anacardiaceae) in Indonesia and Thailand. *Jurnal Teknologi*. 59, 17-20. https://doi.org/10.11113/jt.v59.1572

Hollingsworth, P.M., Graham, S.W., Little, D.P., 2011. Choosing and using a plant DNA barcode. *PLoS ONE*. 6,1-13. https://doi.org/10.1371/journal.pone.0019254

Hou, D., 1974. Anacardiaceae, in: van Steenis, C.G.G.J., van Steenis-Kruseman M.J. (Eds.), *Flora Malesiana-Series 1: Spermatophyta*. Noordhoff-Kolff, Jakarta, pp. 395-548.

[IUCN] International Union for Conservation of Nature, 2022. Available at https://www.iucnredlist.org/search?taxonomies=101139searchType=species. [Date accessed: 13 July 2022]

Javal, M., Terblanche, J.S., Conlong, D.E., Delahaye, N., Grobbelaar, E., Benoit, L., Lopez-Vaamonde, C., Haran, J.M., 2021. DNA barcoding for bio-surveillance of emerging pests and species identification in Afrotropical Prioninae (Coleoptera, Cerambycidae). *Biodiv. Data J.* 9, e64499. https://doi.org/10.3897/BDJ.9.e64499

Kang, Y., Deng, Z., Zang, R. Long, W., 2017. DNA barcoding analysis and phylogenetic relationships of tree species in tropical cloud forests. *Sci. Rep.* 7, 12564. https://doi.org/10.1038/s41598-017-13057-0

Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, A.L., Janzen, D.H., 2005. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. USA.* 102, 8369-8374. https://doi.org/10.1073/pnas.0503123102

Kress, W.J., Erickson, D.L., 2012. DNA barcodes: methods and protocols, methods in molecular biology. *Methods Mol. Biol.* 858, 3-8. https://doi.org/10.1007/978-1-61779-591-6_1

Labdelli, A., De La Herrán, R., Arafeh, R., Resentini, F., Trainotti, L., Halis, Y., Adda, A., Merah, O., 2020. Genetic variation in damaged populations of *Pistacia atlantica* Desf. *Plants.* 9,1541-1552. https://doi.org/10.3390/plants9111541

Lahaye, R., Vander, B.M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., Maurin, O., Duthoit, S., Barraclough, T.G., Savolainen, V., 2008. DNA barcoding the floras of biodiversity hotspots. *Proc. Natl. Acad. Sci. USA.* 105, 2923-2928. https://doi.org/10.1073/pnas.0709936105

Liu, J., Moller, M., Gao, L.M., Zhang, D.Q., Li, DJ, 2011. DNA barcoding for the discrimination of *Eurasian yews* (*Taxus* L., Taxaceae) and discovery of cryptic species. *Molec. Biol. Res.* 11, 89-100. https://doi.org/10.1111/j.1755-0998.2010.02907.x

Mata-Montero, E., Carranza-Rojas, J., 2016. Automated plant species identification: challenges and opportunities, in: Mata, F., Pont, A. (Eds.), *ICT for Promoting Human Development and Protecting the Environment*. Springer, Switzerland, pp. 26-36. https://doi.org/10.1007/978-3-319-44447-5_3

Meier, R., Kwong, S., Vaidya, G., Ng, P.K.L., 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Syst. Biol.* 55, 715-728. https://doi.org/10.1080/10635150600969864

Meyer, C.P., Paulay, G., 2005. DNA barcoding: error rates based on comprehensive sampling. *Plos Biology*. 3, 2229-2238. https://doi.org/10.1371/journal.pbio.0030422

Nagy, Z.T., Sonet, G., Glaw, F., Vences, M., 2012. First large-scale DNA barcoding assessment of reptiles in the biodiversity hotspot of Madagascar, based on newly designed COI primers. *PLoS ONE.* 7, e34506. https://doi.org/10.1371/journal.pone.0034506

Nilsson, R.H., Ryberg, M., Kristiansson, E., Abarenkov, K., Larsson, K-H., Kõljalg, U., 2006. Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. *PLoS ONE*. 1, e59. https://doi.org/10.1371/journal.pone.0000059

Omonhinmin, C., Onuselogu, C., 2022. *rbcL* gene in global molecular data repository. *Data Br.* 42, 108090. https://doi.org/10.1016/j.dib.2022.108090

Pell, S.K., 2004. Molecular systematics of the cashew family (Anacardiaceae) [Dissertation]. Louisiana, USA: Louisiana State University.

Qing, X., Wang, M., Karssen, G., Bucki, P., Bert, W., Braun-Miyara, S., 2020. PPNID: a reference database and molecular identification pipeline for plant-parasitic nematodes. *Bioinformatics.* 153, 6, 1052-1056. https://doi.org/10.1093/bioinformatics/btz707

Rembold, K., Mangopo, H., Tjitrosoedirdjo, S.S., Kreft, H., 2017. Plant diversity, forest dependency, and alien plant invasions in tropical agricultural landscapes. *Biol. Cons*. 213, 234-242. https://doi.org/10.1016/j.biocon.2017.07.020

Ross, H.A., Murugan, S., Li, W.L., 2008. Testing the reliability of genetic methods of species identification via simulation. *Syst. Biol*. 57, 216-230. https://doi.org/10.1080/10635150802032990

Santos, V.N., Costa, A.E.S., Santos, C.A.F., 2021. Paternity identification in Spondias tuberosa (Anacardiaceae: Sapindales) polycrosses using microsatellite loci. *Genet. Mol. Res.* 20, gmr18816. https://doi.org/10.4238/gmr18816

Silva, J.N., da Costa, A.B., Silva, J.V., Almeida, C., 2015. DNA barcoding and phylogeny in neotropical species of the genus *Spondias. Biochem. System Ecol.* 61, 240-243. https://doi.org/10.1016/j.bse.2015.06.005

Sobeh, M., Mahmoud, M.F., Hasan, R.A., Abdelfattah, M.A.O, Sabry, O.M., Ghareeb, M.A., El-Shazly, A.M. Wink, M, 2018. Tannin-rich extracts from *Lannea stuhlmannii* and *Lannea humilis* (Anacardiaceae) exhibit hepatoprotective activities *in vivo* via enhancement of the anti-apoptotic protein Bcl-2. *Sci. Rep*. 8, 9343. https://doi.org/10.1038/s41598-018-27452-8

Tamura, K., Stecher, G., Kumar, S., 2021. MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evo*. 38, 3022-3027. https://doi.org/10.1093/molbev/msab120

Tizard, J., Patel, S., Waugh, J., Tavares, E., Bergmann, T., Gill, B., Norman, J., Christidis, L., Scofield, P., Haddrath, O., Baker, A., Lambert, D., Millar, C., 2019. DNA barcoding a unique avifauna: an important tool for evolution, systematics and conservation. *BMC Evol. Biol*. 19, 52. https://doi.org/10.1186/s12862-019-1346-y

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. *Sci. Rep*. 11, 22837. https://doi.org/10.1038/s41598-021-02237-8

Vaidya, G., Lohman, D.J., Meier, R., 2011. SequenceMatrix: concatenation software for the fast assembly ofmulti-gene datasets with character set and codon information. *Cladistics*. 27, 171-180. https://doi.org/10.1111/j.1096-0031.2010.00329.x

Virgilio, M., Jordaens, K., Breman, F.C., Backeljau, T., De Meyer, M., 2012. Identifying insects with incomplete DNA barcode libraries, African fruit flies (Diptera: Tephritidae) as a test case. *PLoS ONE*. 7, e31581. https://doi.org/10.1371/journal.pone.0031581

Wati, R., Amandita, F.Y., Brambach, F. Siregar, I.Z., Gailing, O., de Melo Moura, C.C., 2022 Filling gaps of reference DNA barcodes in Syzygium from rainforest fragments in Sumatra. *Tree Genet. Genomes*. 18, 1-15. https://doi.org/10.1007/s11295-022-01536-z

Wäldchen, J., Rzanny, M., Seeland, M., Mäder, P., 2018. Automated plant species identification-trends and future directions. *PLoS Comput. Biol.* 14, e1005993. https://doi.org/10.1371/journal.pcbi.1005993

Weeks, A., Zapata, F., Pell, S.K., Daly, D.C., Mitchell, J.D., Fine, PVA, 2014. To move or to evolve: contrasting patterns of intercontinental connectivity and climatic niche evolution in "Terebinthaceae" (Anacardiaceae and Burseraceae). *Front. Genet*. 5, 409. https://doi.org/10.3389/fgene.2014.00409