

DNA barcoding of flowering plants in Sumatra, Indonesia

Fitri Y. Amandita^{1,2}  | Katja Rembold^{3,4} | Barbara Vornam¹ | Sri Rahayu⁵ | Iskandar Z. Siregar⁶ | Holger Kreft³ | Reiner Finkeldey^{1,7}

¹Department of Forest Genetics and Forest Tree Breeding, Georg-August University, Göttingen, Germany

²Research and Development Center for Environmental Quality and Laboratory

³Biodiversity, Macroecology, and Biogeography Research Group, Georg-August University, Göttingen, Germany

⁴Botanical Garden of the University of Bern, Bern, Switzerland

⁵Bogor Botanical Garden, Lembaga Ilmu Pengetahuan Indonesia, Bogor, Indonesia

⁶Department of Silviculture, Bogor Agricultural University (IPB), Bogor, Indonesia

⁷University of Kassel, Kassel, Germany

Correspondence

Fitri Yola Amandita, Ministry of Environment and Forestry, Research and Development Center for Environmental Quality and Laboratory, Tangerang Selatan, Indonesia. Email: fitri.amandita@menlhk.go.id

Funding information

Deutsche Forschungsgemeinschaft, Grant/Award Number: EFForTS project

Abstract

The rapid conversion of Southeast Asian lowland rainforests into monocultures calls for the development of rapid methods for species identification to support ecological research and sustainable land-use management. Here, we investigated the utilization of DNA barcodes for identifying flowering plants from Sumatra, Indonesia. A total of 1,207 *matK* barcodes (441 species) and 2,376 *rbcl* barcodes (750 species) were successfully generated. The barcode effectiveness is assessed using four approaches: (a) comparison between morphological and molecular identification results, (b) best-close match analysis with TaxonDNA, (c) barcoding gap analysis, and (d) formation of monophyletic groups. Results show that *rbcl* has a much higher level of sequence recoverability than *matK* (95% and 66%). The comparison between morphological and molecular identifications revealed that *matK* and *rbcl* worked best assigning a plant specimen to the genus level. Estimates of identification success using best-close match analysis showed that >70% of the investigated species were correctly identified when using single barcode. The use of two-loci barcodes was able to increase the identification success up to 80%. The barcoding gap analysis revealed that neither *matK* nor *rbcl* succeeded to create a clear gap between the intraspecific and interspecific divergences. However, these two barcodes were able to discriminate at least 70% of the species from each other. Fifteen genera and twenty-one species were found to be nonmonophyletic with both markers. The two-loci barcodes were sufficient to reconstruct evolutionary relationships among the plant taxa in the study area that are congruent with the broadly accepted APG III phylogeny.

KEYWORDS

DNA marker, EFForTS project, *matK*, molecular identification, *rbcl*, Sumatra

1 | INTRODUCTION

DNA barcoding is a species identification method, using a short, standardized DNA region, so-called DNA barcode (Hebert, Cywinska, Ball, & de Waard, 2003a). In principle, DNA barcodes contain variation that can be posed as a character to differentiate species. Although the utility of DNA barcoding for species identification

has raised debates over its feasibility (Collins & Cruickshank, 2013; Krisnamurthy & Francis, 2012), the method has been increasingly applied during the last decade, especially to facilitate biodiversity studies of very diverse but taxonomically poorly known regions (Blaxter, 2004; Hajibabaei et al., 2005), such as Sumatran tropical rainforests.

Sumatran tropical rainforests are very rich in flora and fauna (Davis, Heywood, & Hamilton, 1995; Laumonier, 1997; Whitten,

Damanik, Anwar, & Hisyam, 2000); nonetheless, they are only sparsely studied compared to other islands in the Malayan Archipelago (Laumonier, 1997). In terms of plant diversity, the Sumatran forests are comparable to the forests of Borneo and are richer than those found in Java and Sulawesi (Meijer, 1981). Sumatra is reported as one of the global centers of vascular plant diversity with a species density of 3,000 to 5,000 species per 10,000 km² (Barthlott, Mutke, Rafiqpoor, Kier, & Kreft, 2005). Roos, Keßler, Gradstein, and Baas (2004) estimated a total number of 10,600 plant species in Sumatra with more than 300 endemic species. Laumonier (1997) argued that many scientists mistakenly consider that the flora of Sumatra is sufficiently well known since it is similar to that of the Malaysian peninsula, but many parts, especially the center of the island, are floristically unexplored territories.

Despite the importance of conserving the ecosystem, the total forest area in Sumatra has decreased from over 23 million hectares to probably less than 16 million hectares between 1985 and 1997 (World Bank, 2001). The southern provinces of Sumatra have lost most of their lowland forests, including those in protected areas (Lambert & Collar, 2002). Approximately 7.5 million hectares of primary forest loss were recorded in Sumatra during 1990–2010 and an additional 2.3 million hectares of primary forest were degraded (Margono et al., 2012). Between 2000 and 2010, the deforestation rate was estimated to be above 5% per year in the eastern lowlands of Sumatra (Miettinen, Shi, & Liew, 2011). The total deforested areas in Sumatra within 2011 alone were recorded to be approximately 2,200 hectares or as much as 3,520 soccer fields (BP-REDD+, 2015). The causes of these massive deforestation and forest degradation are a large-scale conversion into timber or estate crop plantations, illegal logging, and forest fires. By 2010, 3.9 million hectares of Sumatran lowland forests had been converted into oil palm (*Elaeis guineensis*) plantations (Koh, Miettinen, Liew, & Ghazoula, 2011).

The extensive loss of natural habitat puts a great number of species at risk and may lead to the loss of tropical fauna including forest-dwelling birds (Koh et al., 2011), mammals (Maddox, Priatna, Gemita, & Salampessy, 2007), and orangutan (Gaveau et al., 2009). Undoubtedly, the destruction also affects the plant diversity (Brook, Sodhi, & Ng, 2003; Corlett, 1992; Rembold, Mangopo, Tjitrosoedirdjo, & Kreft, 2017; Turner et al., 1994). The rate of species loss in tropical forests seems to be higher than the species exploration due to lack of resources and sound species conservation management such as limited number of taxonomists working in this region, inadequate herbarium collections, and inaccessible taxonomic literature (Kiew, 2002; Meyer & Paulay, 2005; Tautz, Arctander, Minelli, Thomas, & Vogler, 2003). Species explorations become more challenging when the species cannot be identified morphologically. Identification keys based upon morphological characteristics can be difficult to use if features are not present (e.g., in sterile or juvenile specimens) or not well developed.

The use of DNA barcoding might help to overcome the limitations of morphological characters and might help to speed up species identification. This has been made possible because DNA barcoding can identify organisms at any stage of development (e.g.,

Barber & Boyce, 2006; Hausmann et al., 2011; Heimeier, Lavery, & Sewell, 2010; Ko et al., 2013), or at particular gender (e.g., Elsasser, Floyd, Herbert, & Schulte-Hostedde, 2009), or specimens isolated from small and incomplete tissue, whether it is fresh, broken, or old (e.g., Hajibabaei et al., 2006; Valentini, Pompanon, & Taberlet, 2008). DNA barcoding may also help to discover new species and to identify cryptic species (e.g., Hebert, Penton, Burns, Janzen, & Hallwachs, 2004; Pauls, Blahnik, Zhou, Wardwell, & Holzenthal, 2010; Ward, Costa, Holmes, & Steinke, 2008).

DNA barcoding is now well established for animals (Crawford et al., 2013; Hebert, Cywinska, Ball, & deWaard, 2003a; Hebert, Ratnasingham, & de Waard, 2003b; Hebert et al., 2004; Lim, 2012; Nagy, Sonet, Glaw, & Vences, 2012; Ward, Zemlak, Innes, Lasr, & Hebert, 2005) by using the mitochondrial DNA CO1 (cytochrome c oxidase subunit 1) as a standard region. However, this region is ineffective for plant identification due to generally low nucleotide substitution rates in plant mitochondria (Chase et al., 2005; Fazekas, Kesanakurti, & Burgess, 2009).

A number of candidate gene regions were suggested as potential barcodes for plants including coding genes and noncoding genes in the nuclear and plastid genomes (e.g., Chase, Cowan, & Hollingsworth, 2007; Kress & Erickson, 2007; Kress, Wurdack, Zimmer, Weigt, & Janzen, 2005; Taberlet et al., 2007). Some studies suggested DNA barcoding based on a single chloroplast region (e.g., Lahaye et al., 2008) or a combination of different regions (e.g., Chase et al., 2007; Hollingsworth et al., 2009a; Kress & Erickson, 2007). A study by Kress and Erickson (2007) showed that the various combinations of two loci were all more powerful at differentiating between species than either locus individually. In 2009, the Plant Working Group under The Consortium for Barcode of Life (CBOL) suggested that there were no other two-loci or multi-loci barcode provided appreciably greater species resolution than the *matK+rbcl* combination. However, in some complex groups, such as in the genus *Berberis* (Roy et al., 2010), the combination of *matK* with *rbcl* is not sufficient to distinguish all species. The investigation of these markers will contribute to the development of useful barcode information for plant identification and to document plant species globally.

This study aims to generate DNA barcodes of flowering plant species in four land-use systems in Jambi Province (Sumatra) using two DNA chloroplast markers (*matK* and *rbcl*) and to evaluate the effectiveness of these two markers as DNA barcodes for flowering plants. Crucial characteristics for evaluating the performance of DNA barcodes include universal applicability, ease of data retrieval, and sufficient variability of the used marker (Fazekas et al., 2008; Kress & Erickson, 2007).

2 | METHODS

2.1 | Study sites

This study was carried out in the EFForTS project sites (<https://www.uni-goettingen.de/efforts>) in Jambi Province (Sumatra, Indonesia)

TABLE 1 Universal primers of *matK* and *rbcL* used in DNA amplification and sequencing

No.	Region	Name of primer	Primer sequence (5' → 3')	References
1	<i>matK</i>	<i>3F_KIM_f</i>	CGTACAGTACTTTTGTGTTTACGAG	Ki-Joong Kim (unpublished)
		<i>1R_KIM_r</i>	ACCCAGTCCATCTGGAAATCTGGTTC	Ki-Joong Kim (unpublished)
		<i>390f</i>	CGATCTATTCATTCAATATTTTC	Cuenoud et al. (2002)
		<i>990r</i>	GGACAATGATCCAATCAAGGC	Dayananda, Ashton, Williams, & Primack (1999)
2	<i>rbcL</i>	<i>rbcLa_f</i>	ATGTCACCACAAACAGAGACTAAAGC	Krees and Erickson (2007)
		<i>rbcLa_r</i>	GAAACGGTCTCTCCAACGCAT	Fazekas et al. (2008)

comprises of 32 core plots sized 50 m × 50 m. Details about the EFForTS project sites and plot design are described in Drescher et al. (2016).

2.2 | Specimen collection and identification

Herbarium specimens were collected from three individuals of as many as possible vascular plant species within the 32 core plots. The plant survey included all trees with a diameter at breast height (DBH) ≥10 cm within the entire plot and all vascular plants within five 5 m × 5 m subplots nested within each core plot. Leaf tissue (approximately 2 cm²) was collected from each fresh herbarium specimen and dried in silica gel for DNA barcoding analysis. Herbarium vouchers were prepared, morphologically identified, and deposited at the herbarium of the Southeast Asian Regional Centre for Tropical Biology (SEAMEO-BIOTROP), the Herbarium Bogoriense—Research Center for Biology, LIPI, and herbarium of the University of Jambi. The results of the morphological identification were then compared to the molecular identification results. Molecular identification was conducted for all samples that were successfully barcoded, but only samples that have been morphologically identified were included in the further analysis.

2.3 | DNA analysis

Based on the result of morphological species identification, two specimens per species were selected for genetic analysis. DNA extractions were performed on healthy dried leaf tissues from all selected samples using the DNeasy 96 Plant Kit (Qiagen, Hilden, Germany) following the manufacturer's protocols. The concentration and quality of the extracted DNA were checked by 0.8%–1% agarose gel electrophoresis with Lambda DNA as standard (Roche), visualized by UV illumination and saved using a polaroid camera.

Each extracted DNA was amplified by performing polymerase chain reaction (PCR) using universal primers listed in Table 1. For *rbcL*, the amplification was straightforward, while for *matK*, two different amplification reactions were performed. First, the DNA of all investigated samples were amplified using the universal primer pair 1RKIM_f and 3FKIM_r (Table 1). The second amplification reaction, using the primer pair 390f and 990r (Table 1), included only those samples which showed no amplification product or produced multiple PCR products in the first amplification reaction.

The sequencing reactions were performed using the ABI Prism™ Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit v1.1 (Applied Biosystems), based on the principles described by Sanger, Nicklen, and Coulson (1977). Data were collected from capillary electrophoresis on an ABI Prism 3100® Genetic Analyzer with the Sequence Analysis Software v3.1 (Applied Biosystems). The sequencing was performed with the same primers used for amplification in both directions. The amplification and sequencing reaction mixtures are shown in Supporting Information Appendix 1, while the temperature profiles of the PCR for amplification and sequencing are shown in Supporting Information Appendix 2.

2.4 | Sequence analysis

To ensure the generated DNA barcodes were as accurate as possible, sequence editing was performed using CodonCode Aligner software (CodonCode Corporation, Dedham, USA). Furthermore, each of these edited barcodes was assigned to a particular taxon by comparing it with the nucleotide sequences in GenBank database and Barcode of Life Database (BOLD).

Moreover, the results of sequence identification were cross-checked with the morphological identification results. The match between morphological and molecular identification results was counted into three levels: species, genus, and family. The following decisions were made for correct identification assignments, namely: (a) when the species name from the molecular identification matched the species name from the morphological identification, then it was counted as a correct species identification, (b) when the identification result only matched the genus or family, then it was counted as correct genus or family identification, and (c) when the result between morphological and molecular identification did not match, it was counted as incorrect identification if *matK* and *rbcL* both showed similar results at least at family level, or it was counted as mislabeling/contamination if the results of *matK* and *rbcL* were different. Herbarium specimens were double-checked in cases of incorrect identification.

Sequence alignment was carried out independently for each marker in two stages. First, multiple sequences were aligned according to their families using the ClustalW program (Thompson, Higgins, & Gibson, 1994) embedded in MEGA6 (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013). Reference sequences were downloaded from GenBank/BOLD and included in the alignment for those species represented with only one sample. The alignment results were subsequently

checked for the occurrence of ambiguities caused by the presence of indels and/or substitutions and edited if necessary. In the second stage, all aligned sequences from each family were manually aligned with sequences from other families. Gaps were added if necessary, and the final alignment was trimmed at both ends. The aligned sequences of *rbcl* and *matK* were combined to obtain two-loci DNA barcodes using SequenceMatrix software (Vaidya, Lohman, & Meier, 2011).

Identification success was also calculated with best-close match analysis as implemented in TaxonDNA (Meier, Kwong, Vaidya, & Ng, 2006). This analysis only included the species with at least two representatives. A threshold value T was determined for each dataset as a divergence percentage in which 95% of all intraspecific distances were found. In this method, all recovered barcodes were formatted as both database and query. A query can only be identified if the corresponding sequence has a match in the dataset that falls into the 0% to $T\%$ interval. If the species name was identical, the query was considered to be successfully identified. A query was considered ambiguously identified when it matched more than one sequence of different species besides the correct species. On the other hand, a query was considered incorrectly identified when it matched to sequences belonging to other species. All queries without such a match would remain unidentified.

Pairwise distance matrices were created to calculate the genetic distance using MEGA6 (Tamura et al., 2013) based on the Tamura-Nei model (1993) assuming the differences in substitution rate between nucleotides and the inequality of nucleotide frequencies with gamma-distributed rates between sites and the pattern between lineages were assumed to be heterogeneous. The calculation results of intra- and interspecific divergences in these matrices were separated using EXCALIBAR (Aliabadian et al., 2014) to facilitate the measures of distance range and distance mean of each type of divergence. Frequency (%) distribution of intra- and interspecific divergences of each marker was calculated and depicted in graphics using Excel to find possible "gap" between these two divergences. This so-called barcoding gap illustrates the effectiveness of DNA barcodes in discriminating query species from one to another. An ideal barcode can be determined by the presence of a barcoding gap, which occurs when the minimum value of the interspecific divergence is higher than the maximum level of intraspecific divergence (Meyer & Paulay, 2005).

Based on the aligned sequences, phylogenetic trees were reconstructed using MEGA6 (Tamura et al., 2013) with three different algorithms: maximum parsimony (MP), maximum likelihood (ML), and neighbor joining (NJ). Percentages of species, genus, and family monophyletic clades were calculated from each reconstructed tree. Furthermore, ordinal-level phylogenies were reconstructed based on maximum likelihood trees of each used marker and were compared to APG III (APG III 2009) phylogenies to see if there were inconsistencies between these two topologies.

3 | RESULTS

From all 5,328 samples collected from the field, only 2,590 samples were included in the study due to time restriction. The selection of

studied samples was based on the consideration to involve as much species as possible, and each of these species should be represented at least by two samples. Species with only one sample were still included, but the barcodes generated from single-sampled species were excluded from the pairwise analysis.

We extracted DNA from dried leaf specimens without noticeable difficulties. The amplification and sequencing, however, turned out to be more problematic especially when using *matK* primers. Recoverability of DNA sequences for *rbcl* was overall high (amplification and sequencing success were 96.9% and 94.7%, respectively). The amplification and sequencing results using the primer of *matK* were only moderately successful (79.1% and 65.8%, respectively). A total of 1,207 *matK* barcodes representing 441 species of 97 families of 40 orders, and 2,376 *rbcl* barcodes representing 750 species of 126 families of 44 orders, were generated in this study.

For both markers, the highest match between morphological and molecular identification was at genus level (46.6% with *matK* and 51.3% with *rbcl*). The matched identification at species level was higher with *matK* than with *rbcl* (30.2% and 22.4%, respectively). Meanwhile, incorrect identification was relatively low for both regions (3.5%). To maintain the accuracy of the analysis, we excluded all misidentified or presumably mislabeled barcodes from the dataset. Since the study aims at comparing the performance of *matK* and *rbcl* and to generate two-loci barcodes, only samples from which both *matK* and *rbcl* barcodes were successfully recovered were included in the further analysis. Consequently, only 322 samples from 161 species (two samples per species) were included in best-close match and barcode-gap analysis and 334 samples from 334 species (one sample per species) were included in phylogenetic analysis.

According to the best-close match analysis, *matK* has higher overall species identification success compared to *rbcl* (78.3% and 71.4%, respectively), and the highest correct species identification was obtained by the combination of both markers (81.1%). There were 22 species which remained unidentified by each marker and the two-loci marker.

Furthermore, this study showed that the mean value of intraspecific divergences (0.0008–0.0014) was very low and the mean value of the interspecific divergences (0.1–0.3) was significantly higher (unpaired t -test, $p < 0.01$). The frequency (%) distribution of intraspecific and interspecific divergence using three markers (Figure 1) showed that no barcode gaps existed as the intraspecific divergences overlapped with interspecific divergences.

As expected, *matK* had a higher discrimination level than *rbcl* (80% and 73%, respectively) but the difference was not significant (one-way ANOVA, $p > 0.05$). The combination of *matK* and *rbcl* improved the discrimination up to 89%. Forty-four out of 161 species could not be discriminated by *rbcl* and eleven of them were not discriminated by any of the markers including the two-loci barcode. These species were mostly from species-rich genera, such as *Ficus* (Moraceae), *Santiria* (Bursaceae), and *Litsea* (Lauraceae).

Nine phylogenetic trees (Supporting information Appendix 3–11) were constructed based on multiple sequence alignments of *matK*,

rbcl, and *matK+rbcl* using three different methods: maximum parsimony (MP), neighbor joining (NJ), and maximum likelihood (MP). Each tree was observed and similar topologies were found amongst these trees (Table 2).

Seventeen families were not included in the calculation of family-level monophyletic percentage as these families were presented with only one taxon. The two-loci marker provided 100% taxonomic resolution at family level with all three different methods. Twenty-two species were nonmonophyletic in all phylogenetic trees (Supporting information Appendix 12). The nonmonophyletic species mostly originated from species-rich families, such as Burseraceae, Myristicaceae, Moraceae, Phyllanthaceae, Lauraceae, Sapindaceae, and Annonaceae.

The ordinal-level phylogeny of flowering plants shows the relationship between orders of flowering plants and the grouping of these orders (Figure 2). The *matK* marker misplaced Myrtales and failed to separate Laurales from Magnoliales. Meanwhile, the *rbcl* marker misplaced Aquifoliales and grouped Malpighiales and Brassicales into one monophyletic clade. This marker also failed to make Santalales a monophyletic clade. However, this marker successfully separated Laurales from Magnoliales. Finally, the combination of *matK* and *rbcl* improved the topologies of the tree and put nearly all orders into the right position compared to APG III phylogeny.

4 | DISCUSSION

4.1 | Recoverability and quality of *matK* and *rbcl* barcodes

The *rbcl* universality as DNA barcode observed in this study confirms that DNA sequences could be easily obtained with *rbcl* primers from a wide range of tropical plant species (e.g., Gonzales et al., 2009; Lahaye et al., 2008; Parmentier et al., 2013). In contrast to *rbcl*, *matK* seems to be less suitable for tropical floras compared to temperate one (e.g., Bruni et al., 2012; de Vere et al., 2012; Gonzales et al., 2009). This might be due to higher evolutionary rates in tropical compared to temperate plants (Gillman, Keeling, Gardner, & Wright, 2010). The PCR of *matK* performed in this study was using two pairs of primers which were found to be effective to generate

DNA barcodes from specific taxa, such as *Tetrastigma* (Fu, Jiang, & Fu, 2011), *Hedyotis* (Guo, Simmons, But, Shaw, & Wang, 2011), or Asteraceae (Gao et al., 2010). These primers, however, became less effective when they were used for a wide range of species (Gonzales et al., 2009; Kress et al., 2010). A certain primer pair did not always yield a PCR product in all members of a group of seemingly closely related taxa, indicating that the primers themselves are not conserved.

The use of *matK* as a barcode has been criticized mainly because universal primers are not available (e.g., Bafeel et al., 2011; Dong et al., 2015). A study by Fazekas et al. (2008) showed a relatively high rate of sequencing success for this marker after using up to 10 primer pairs. The usefulness of *matK* primers is proven when they are used in specific species or taxa, such as *Camellia sinensis* (Stoeckle et al., 2011), Lamiaceae (De Mattia et al., 2011), or palms (Jeanson, Labat, & Little, 2011). In a review of the best barcode for plants, Hollingsworth, Graham, and Little (2011) indicated that *matK* still needs optimization in regard to primer combinations and needs to be adapted to specific taxonomic groups.

4.2 | Plant species identification success using *matK* and *rbcl*

As one way to evaluate the success rate of species identification, we compared the results from morphological identification with the results from molecular identification. Some authors suggested a superiority of molecular identification in comparison with morphological identification (Newmaster, Ragupathy, & Janovec, 2009; Stace, 2005). However, this study showed that DNA barcoding alone is not sufficient to assign all DNA sequences to a correct species name. Only 22%–30% of the samples were correctly assigned to the correct species, while the majority of correct identifications was limited to genus level (46%–51%).

Approximately three percent of mismatch between morphological identification results and DNA identification results were found in this study that could be due to several reasons. A specimen could be misidentified when it was found to have the highest similarity to a reference sequence that was falsely identified. The mismatch between morphological and molecular identification could also happen when the taxonomist misidentified the voucher. Morphological identification is difficult in the absence of certain features, such as

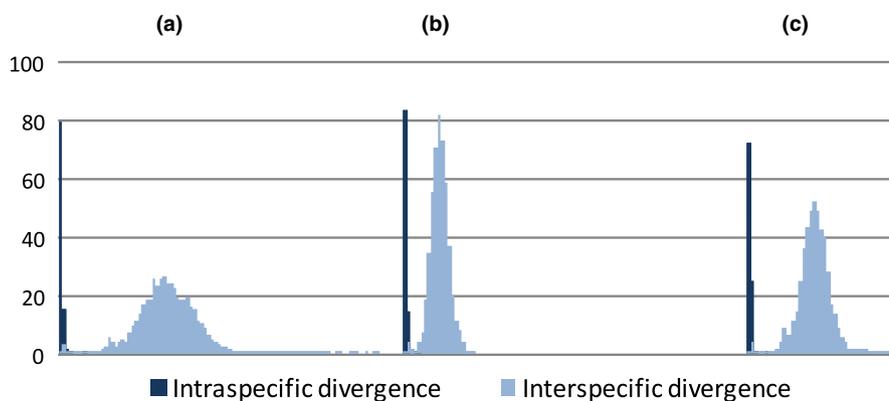
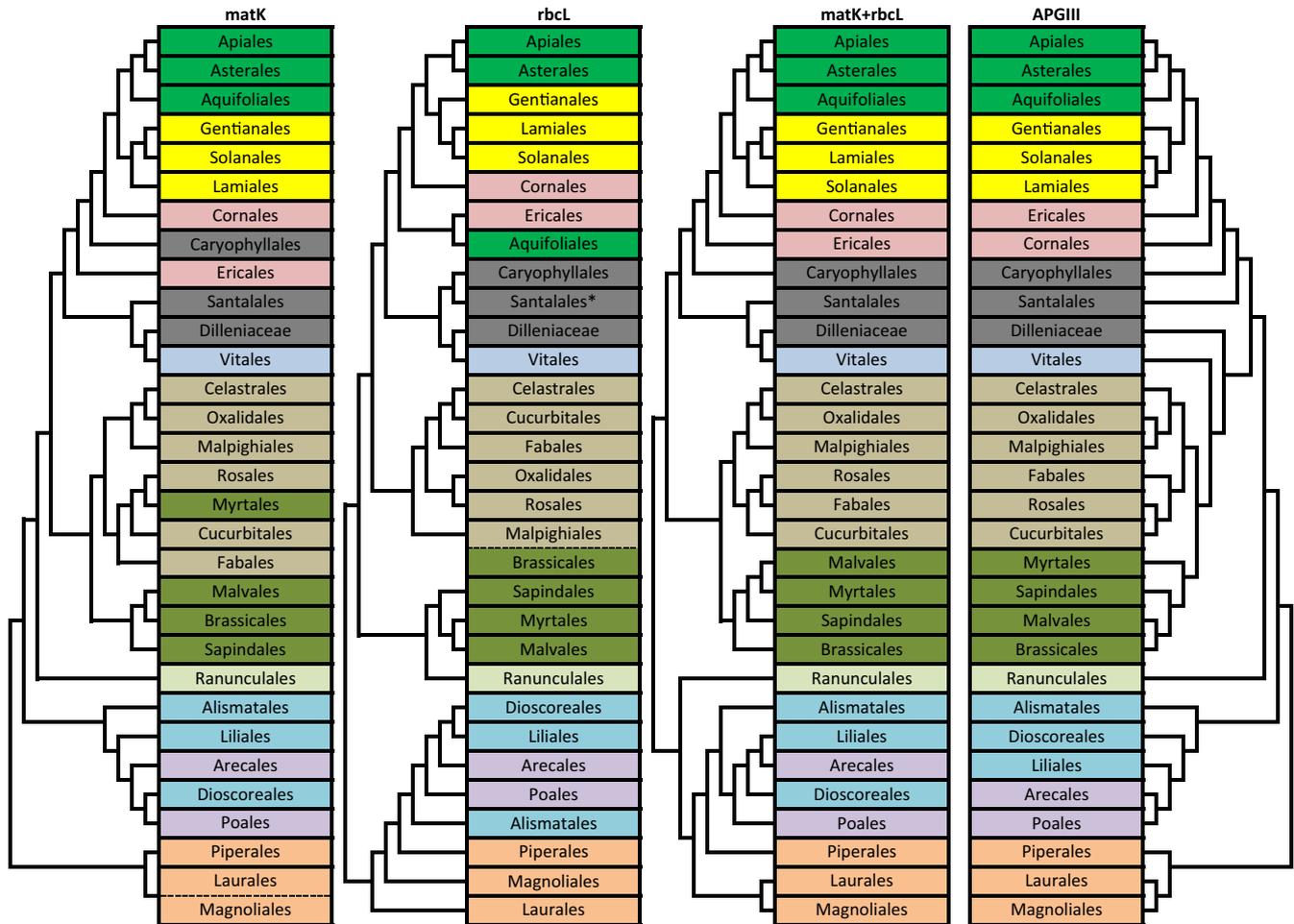


FIGURE 1 Frequency (%) distribution of intraspecific and interspecific divergences of pairwise sequences of *matK* (a), *rbcl* (b), and *matK+rbcl*(c)

TABLE 2 Percentage of monophyletic clades recovered in nine reconstructed phylogenetic trees

Barcode	Monophyletic with support value >70%								
	Maximum Parsimony (MP)			Neighbor Joining (NJ)			Maximum Likelihood (ML)		
	Family	Genus	Species	Family	Genus	Species	Family	Genus	Species
<i>matK</i>	95.9	68.4	73.9	93.9	66.7	69.6	98.0	64.9	68.9
<i>rbcl</i>	95.9	63.2	60.3	93.9	63.2	64.0	89.9	63.2	55.9
<i>matK+rbcl</i>	100.0	71.9	73.3	100.0	64.9	73.9	100.0	70.2	75.2

**FIGURE 2** Comparison between ordinal-level phylogeny of flowering plants based on DNA barcodes and APG III (2009). The dash lines indicate that the two orders are not clearly separated. *Santalales in *rbcl* phylogeny tree is a nonmonophyletic clade

flowers or fruits, especially when dealing with species-rich groups. A high percentage of nonfertile material is particularly common in ecological projects such as ours. In the case of incorrect morphological identification, the herbarium vouchers of corresponding samples should be verified morphologically once again.

The success of species identification using DNA barcoding depends very much on the taxa in question, as much as the utilized marker. For example, in this study, the family Piperaceae resulted in high species-matched identification when using *matK* (60%) but no success at all when using *rbcl*. Meanwhile, for the family Asteraceae, the species-matched identification was higher with *rbcl* (50%) than with *matK* (30%).

Another factor affecting the success of species identification using DNA barcoding is the availability of nucleotide data of the corresponding taxa in the DNA sequences database such as GenBank and BOLD. Through this study, 303 newly barcoded tropical plant species have been uploaded to BOLD. Forty-one percent of the 772 species investigated in this study still had no nucleotide data in BOLD and Genbank. Thus, a significant proportion of samples belonging to species which were not yet recorded in the reference databases lead to increased rates of unassigned samples. Incorrect specimen assignment is more often due to the incompleteness of molecular datasets rather than the data analysis (Bruni et al., 2010; Burgess et al., 2011; Cowan & Fay, 2012). An accurate and complete molecular

database, especially for plant species, is still far from being achieved in the present state. Such a database will hopefully be developed in the future as many studies and projects of plant DNA barcoding are going on (e.g., <http://botany.si.edu/projects/DNAbarcode/intro.htm>; <http://xmalesia.info/index.html>).

4.3 | Discriminatory power of *matK* and *rbcL*

None of the markers used in this study successfully obtained a DNA barcoding gap. All of the minimum values of interspecific divergence obtained from three different markers were lower than the maximum values of intraspecific divergence. In studies of DNA barcoding of specific plant taxa, for example, *Ludwigia* (Ghahramanzadeh et al., 2013), *Abies*, *Cupressus* (Armenise, Simeone, Piredda, & Schirone, 2012), and *Tetrastigma* (Fu, Jiang, & Fu, 2011), the distribution of intra- versus interspecific distances was relatively well separated. Meanwhile, large-scale plant diversity inventories (Lahaye et al., 2008; Parmentier et al., 2013) reported the absence of barcoding gaps by using a combination of potential markers. The richness of the dataset might have contributed to the wider distribution of the intra- and interspecific divergences which then increase the possibility of them to overlap. This implies that the sampling intensity and variety would influence the distribution of the intra- and interspecific variation within the dataset.

Despite the absence of barcoding gaps, the barcodes generated in this study have relatively high discriminatory power. According to Hollingsworth et al. (2011), most of the plant barcodes would have discriminatory power of more than 70%. Studies by Kress et al. (2009) and Burgess et al. (2011) showed that barcoding of distantly related taxa typically results in high levels of discriminatory power.

The *matK+rbcL* marker has the highest number of discriminated species compared to *matK* or *rbcL* alone. This is because the use of two-loci barcodes maximized the genetic variation, thus minimizing the number of identical barcodes between different species. All species that could not be discriminated have barcodes identical to other species from the same family. Identical barcodes across different genera of the same family were uncommon with *matK* but more common with *rbcL*. However, *matK* and *rbcL* mostly failed to discriminate different species from the same genus. These two plastid markers are therefore not variable enough to be effective barcodes for closely related species in certain taxa.

To improve the analysis of closely related taxa, noncoding plastid genes, such as *trnH-psbA*, could be used as an additional marker (Hollingsworth et al., 2011). A study by Kress and Erickson (2007) showed that *trnH-psbA* has dramatically higher sequence variability than the coding genes because it has a higher number of single-nucleotide polymorphisms (SNPs). Hence, *trnH-psbA* can be a suitable marker to discriminate among closely related species. Moreover, nuclear genomic regions, such as the internal transcribed spacer (ITS) region, were suggested as potential DNA barcodes by Kress et al. (2005). ITS sequences generally show high levels of interspecific sequence variability (Cowan & Fay, 2012) and has been used successfully to classify angiosperms (Li et al., 2011).

4.4 | The phylogeny of flowering plants of Jambi based on *matK* and *rbcL*

Both *matK* and *rbcL* showed high family-level resolution, and the combination of *matK* and *rbcL* succeeded to resolve all of the families into monophyletic clades with high bootstrap value. Furthermore, the taxonomic resolution at the genus level was much lower compared to the family level which was expected. Surprisingly, the genus-level monophyletic percentages were found slightly lower compared to the species level in all trees, except for MP and ML trees using *rbcL*. A similar study by Gonzalez et al. (2009) reported larger numbers of monophyletic genera compared to monophyletic species. This difference can be explained by the fact that the proportion of distantly related species included in the dataset in this study was higher than the proportion of closely related species. Thus, the probability of resolving monophyletic-species clades was higher than to resolve the monophyletic-genus clade. Finally, the species-level resolution in this study is comparable to similar studies (Gonzalez et al., 2009; de Vere et al., 2012). However, the two-loci barcode did not improve the species-level resolution significantly. Combining these two chloroplast markers was not sufficient to provide 100% of species monophyly.

Of 76 families included in the phylogenetic tree reconstruction, Burseraceae and Phyllanthaceae were the families with the highest number of unresolved genera. Most of the species in these genera were found to have identical sequences, so they could not be separated from each other. Identical sequences between species of different genera could be common if the marker was not variable enough, such as *matK* and *rbcL*. In this study, it was revealed that *matK* and *rbcL* were not sufficiently variable for species-rich groups.

The phylogenetic trees based on the *rbcL* marker resulted in larger numbers of unresolved species than *matK*. At least eighteen species were nonmonophyletic according to *rbcL* but monophyletic according to *matK*. The unresolved species found in this study could be explained by two reasons. First, these species might have identical genetic information with other species belonging to the same genera/family. Second, these species might have higher intraspecific than interspecific divergence; thus, they were grouped with the allo-species but not with the conspecifics.

A number of constraints are limiting DNA barcoding of plant species including slow evolution rates (Palmer et al., 2000) and high incidence of hybridization (Knobloch, 1972). The genetic variation caused by hybridization cannot be simply detected by plastid markers (Fazekas et al., 2008, 2009). Nevertheless, none of the plant DNA markers are perfect in every case (Hollingsworth et al., 2011). Indeed, one of the future challenges for plant DNA barcoding is to find the most suitable marker to tackle these problems. As the DNA sequencing technology and bioinformatic tools are progressively advancing, the development of new primers will be much easier and at the end will increase the success of DNA barcoding. The application of next-generation sequencing (NGS) technology will enhance the capability of DNA barcoding as a powerful tool in the studies of

ecology, evolution, and conservation biology (Kress, Garcia-Robledo, Uriarte, & Erickson, 2014).

5 | CONCLUSION

We conclude that the two plastid markers *matK* and *rbcl* as plant barcodes work reasonably well in identifying flowering plant species in Sumatran lowland rainforest and surrounding agricultural systems, at least up to genus level. However, there are taxa that are difficult to be distinguished using *matK* and *rbcl*. These taxa mostly belong to species-rich clades with low interspecific divergences. DNA barcoding of closely related species results in low success, especially when using coding plastid markers, such as *matK* and *rbcl*.

The success of species identification strongly depends on the availability of an accurate and complete molecular database. Such database should include sufficient barcodes for each species distributed over its entire distribution range to cover the full range of its intraspecific variability. Thus, future studies ideally include all congeneric species from a geographic region and maximize the geographic diversity of samples for each species. Moreover, utilization of supplement markers, such as *psbA-trnH* or ITS, is highly recommended in combination with *matK* and *rbcl*.

All of DNA barcodes generated in this study, comprises more than 500 species of flowering plants, are uploaded to BOLD. This, coupled with the collection of herbarium vouchers, will improve the usability of DNA barcodes for plant identification.

ACKNOWLEDGMENTS

We would like to thank the German Research Foundation (DFG) for funding this research as part of the EFForTS project in the framework of the Collaborative Research Centre 990 (<http://www.uni-goettingen.de/crc990>). Sincere thanks to Lembaga Ilmu Pengetahuan Indonesia (LIPI), and Ministry of Environment and Forestry of Republic of Indonesia (KLHK). Special thanks to Hardianto Mangopo, Fajar Adityarama, Melanie Schmitt and Alexandra Dolynska for the technical assistances, and to the staff of Harapan Rainforest (PT. Restorasi Ekosistem Indonesia) and Bukit Duabelas National Park. We would also like to thank the Ministry of Research, Technology and Higher Education (RISTEKDIKT) for research permission in Indonesia.

AUTHOR CONTRIBUTIONS

F.Y.A. performed specimen collection, laboratory work, sequence analyses and wrote the manuscript. K.R. performed specimen collection, morphology identification and provided critical review of the manuscript. B.V. supported part of the laboratory work and data analysis, and revised the manuscript. S.R. provided author citation for each botanical name of species barcoded in this study and revised the manuscript. I.Z.S. provided the sample collection permit and mutual transfer agreement (MTA) documents, and revised the manuscript. H.K. and R.F. supervised the research and revised the manuscript.

DATA ACCESSIBILITY

All data for the project were managed in the BOLD database in a project called "DNA Barcoding of Vascular Plants in Jambi, Indonesia" (project code CRCZ). A list of all species barcoded in this study is available as Supporting Information (Table S1).

ORCID

Fitri Y. Amandita  <https://orcid.org/0000-0001-9124-3267>

REFERENCES

- Aliabadian, M., Nijman, V., Mahmoudi, A., Naderi, M., Vonk, R., & Vences, M. (2014). ExcaliBAR: A simple and fast software utility to calculate intra- and inter-specific distances from DNA barcodes. *Contributions to Zoology*, 83, 79–83.
- App III. (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society*, 161, 105–121. <https://doi.org/10.1111/j.1095-8339.2009.00996.x>
- Armenise, L., Simeone, M. C., Piredda, R., & Schirone, B. (2012). Validation of DNA barcoding as an efficient tool for taxon identification and detection of species diversity in Italian conifers. *European Journal of Forest Research*, 131, 1337–1353. <https://doi.org/10.1007/s10342-012-0602-0>
- Bafeel, S. O., Arif, I. A., Bakir, M. A., Khan, H. A., Al Farhan, A. H., ... Thomas, J. (2011). Comparative evaluation of PCR success with universal primers of maturase K (*matK*) and ribulose-1, 5-bisphosphate carboxylase oxygenase large subunit (*rbcl*) for barcoding of some arid plants. *Plant Omics*, 4, 195–198.
- Barber, P., & Boyce, S. L. (2006). Estimating diversity of Indo-Pacific coral reef stomatopods through DNA barcoding of stomatopod larvae. *Proceedings of the Royal Society B: Biological Sciences*, 273, 2053–2061. <https://doi.org/10.1098/rspb.2006.3540>
- Barthlott, W., Mutke, J., Rafiqpoor, D., Kier, G., & Kreft, H. (2005). Global centers of vascular plant diversity. *Nova Acta Leopoldina NF*, 92, 61–83.
- Blaxter, M. L. (2004). The promise of a DNA taxonomy. *Philosophical Transactions of the Royal Society B*, 359, 669–679. <https://doi.org/10.1098/rstb.2003.1447>
- Brook, B. W., Sodhi, N. S., & Ng, P. K. L. (2003). Catastrophic extinctions follow deforestation in Singapore. *Nature*, 424, 420–423. <https://doi.org/10.1038/nature01795>
- Bruni, I., De Mattia, F., Galimberti, A., Galasso, G., Banfi, E., Casiraghi, M., & Labra, M. (2010). Identification of poisonous plants by DNA barcoding approach. *International Journal of Legal Medicine*, 124, 595–603. <https://doi.org/10.1007/s00414-010-0447-3>
- Bruni, I., De Mattia, F., Martellos, S., Galimberti, A., Savadori, P., Casiraghi, M., ... 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.
- Burgess, K. S., Fazekas, A. J., Kesanakurti, P. R., Graham, S. W., Husband, B. C., ... Barrett, S. C. H. (2011). Discriminating plant species in a local temperate flora using the *rbcl* plus *matK* DNA barcode. *Methods in Ecology and Evolution*, 2, 333–340.
- BP-REDD+ (2015). National Forest Reference Emission Level for Deforestation and Forest Degradation in the Context of the Activities Referred to in Decision 1/CP.16, Paragraph 70 (REDD+) Under the UNFCCC: A Reference for Decision Makers, Published by BP-REDD+ Indonesia
- 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. <https://doi.org/10.1073/pnas.0905845106>
- Chase, M. W., Salamin, N., Wilkinson, M., Dunwell, J. M., Kesanakurti, P. R., Haidar, N., & Savolainen, V. (2005). Land plants and DNA barcodes: Short-term and long-term goals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360, 1889–1895. <https://doi.org/10.1098/rstb.2005.1720>
- Chase, M. W., Cowan, R. S., Hollingsworth, P. M., van den Berg, C., Madriñán, S., Petersen, G., ... Wilkinson, M. (2007). A proposal for a standardized protocol to barcode all land plants. *Taxon*, 56, 295–299.
- China Plant BOL Group, 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*, 108, 19641–19646. <https://doi.org/10.1073/pnas.1104551108>
- Collins, R. A., & Cruickshank, R. H. (2013). The seven deadly sins of DNA barcoding. *Molecular Ecology Resources*, 13, 969–975. <https://doi.org/10.1111/1755-0998.12046>
- Corlett, R. T. (1992). The ecological transformation of Singapore, 1819–1990. *Journal of Biogeography*, 19, 411–420. <https://doi.org/10.2307/2845569>
- Cowan, R. S., & Fay, M. F. (2012). Challenges in the DNA barcoding of plant material. *Molecular Biology*, 862, 23–33. https://doi.org/10.1007/978-1-61779-609-8_3
- Crawford, A. J., Cruz, C., Griffith, E., Ross, H., Ibáñez, R., Lips, K. R., ... Crump, P. (2013). DNA barcoding applied to ex situ tropical amphibian conservation program reveals cryptic diversity in captive populations. *Molecular Ecology Resources*, 13, 1005–1018.
- Cuenoud, P., Savolainen, V., Chatrou, L. W., Powell, M., Grayer, R. J., & Chase, M. W. (2002). Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcl*, *atpB*, and *matK* DNA sequences. *American Journal of Botany*, 89, 132–144. <https://doi.org/10.3732/ajb.89.1.132>
- Davis, S. D., Heywood, V. H., & Hamilton, A. C. (1995). *Centres of plant diversity: A guide and strategy for their conservation* (Vol. 2. Asia, Australia, and the Pacific). Cambridge: IUCN Publications Unit.
- Dayanandan, S., Ashton, P. S., Williams, S. M., & Primack, R. B. (1999). Phylogeny of the tropical tree family Dipterocarpaceae based on nucleotide sequences of the chloroplast *rbcl* gene. *American Journal of Botany*, 86, 1182–1190. <https://doi.org/10.2307/2656982>
- De Mattia, F., Bruni, I., Galimberti, A., Cattaneo, F., Casiraghi, M., & Labra, M. (2011). A comparative study of different DNA barcoding markers for the identification of some members of Lamiaceae. *Food Research International*, 44, 693–702. <https://doi.org/10.1016/j.foodres.2010.12.032>
- de Vere, N., Rich, T. C. G., Ford, C. R., Trinder, S. A., Long, C., Moore, C., ... Wilkinson, M. J. (2012). DNA Barcoding the native flowering plants and conifers of Wales. *PLoS ONE*, 7, e37945. <https://doi.org/10.1371/journal.pone.0037945>
- Dong, W., Xu, C., Li, C., Sun, J., Zuo, Y., Shi, S., ... Zhou, S. (2015). *ycf1*, the most promising plastid DNA barcode of land plants. *Scientific Reports*, 5, 8348. <https://doi.org/10.1038/srep08348>
- Drescher, J., Rembold, K., Allen, K., Beckschäfer, P., Buchori, D., Clough, Y., ... Scheu, S. (2016). Ecological and socioeconomic functions across tropical land-use systems after rainforest conversion. *Philosophical Transactions of the Royal Society B*, 371, 20150275. <https://doi.org/10.1098/rstb.2015.0275>
- Elsasser, S. C., Floyd, R., Herbert, P. D. N., & Schulte-Hostedde, A. I. (2009). Species identification of North American guinea worms (Nematoda: Dracunculus) with DNA barcoding. *Molecular Ecology Resources*, 9, 707–712.
- Fazekas, A. J., Burgess, K. S., Kesanakurti, P. R., Graham, S. W., Newmaster, S. G., Husband, B. C., ... Barrett, S. C. H. (2008). Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE*, 3, e2802. <https://doi.org/10.1371/journal.pone.0002802>
- Fazekas, A. J., Kesanakurti, P. R., Burgess, K. S., Percy, D. M., Graham, S. W., Barrett, S. C., ... Husband, B. C. (2009). Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Molecular Ecology Resources*, 9, 130–139. <https://doi.org/10.1111/j.1755-0998.2009.02652.x>
- Fu, Y.-M., Jiang, W.-M., & Fu, C.-X. (2011). Identification of species within *Tetrastigma* (Miq.) Planch. (Vitaceae) based on DNA barcoding techniques. *Journal of Systematics and Evolution*, 49, 237–245. <https://doi.org/10.1111/j.1759-6831.2011.00126.x>
- Gao, T., Yao, H., Song, J., Liu, C., Zhu, Y., ... Chen, S. (2010). Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. *Journal of Ethnopharmacology*, 130, 116–121. <https://doi.org/10.1016/j.jep.2010.04.026>
- Gaveau, D. L. A., Wich, S., Epting, J., Juhn, D., Kanninen, M., & Leader-Williams, N. (2009). The future of forests and orangutans (*Pongo abelii*) in Sumatra: Predicting impacts of oil palm plantations, road construction, and mechanisms for reducing carbon emissions from deforestation. *Environmental Research Letters*, 4, 11. <https://doi.org/10.1088/1748-9326/4/3/034013>
- Ghahramanzadeh, R., Esselink, G., Kodde, L. P., Duistermaat, H., Van Valkenburg, J. L. C. H., Marashi, S. H., ... Van De Wiel, C. C. M. (2013). Efficient distinction of invasive aquatic plant species from non-invasive related species using DNA barcoding. *Molecular Ecology Resources*, 13, 21–31. <https://doi.org/10.1111/1755-0998.12020>
- Gillman, L. N., Keeling, D. J., Gardner, R. C., & Wright, S. D. (2010). Faster evolution of highly conserved DNA in tropical plants. *Journal of Evolutionary Biology*, 23, 1327–1330. <https://doi.org/10.1111/j.1420-9101.2010.01992.x>
- Gonzalez, M. A., Baraloto, C., Engel, J., Mori, S. A., Petronelli, P., Riéra, B., ... Chave, J. (2009). Identification of Amazonian trees with DNA barcodes. *PLoS ONE*, 4, e7483. <https://doi.org/10.1371/journal.pone.0007483>
- Guo, X., Simmons, M. P., But, P.-P.-H., Shaw, P.-C., & Wang, R. J. (2011). Application of DNA barcodes in *Hedyotis* L. (Spermacoceae, Rubiaceae). *Journal of Systematics and Evolution*, 49, 203–212. <https://doi.org/10.1111/j.1759-6831.2011.00130.x>
- Hajibabaei, M., Smith, A., Janzen, D. H., Rodriguez, J. J., Whitfield, J. B., & Hebert, P. D. N. (2006). A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes*, 2006, 959–964. <https://doi.org/10.1111/j.1471-8286.2006.01470.x>
- Hajibabaei, M., de Waard, J. R., Ivanova, N. V., Ratnasingham, S., Dooph, R. T., Kirk, S. L., ... Hebert, P. D. N. (2005). Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society B*, 360, 1959–1967. <https://doi.org/10.1098/rstb.2005.1727>
- Hausmann, A., Haszprunar, G., Segerer, A. H., Speidel, W., Behounek, G., & Hebert, P. D. N. (2011). Now DNA-barcoded: The butterflies and larger moths of Germany (Lepidoptera: Rhopalocera, Macroheterocera). *Spixiana*, 34, 47–58.
- Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003a). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, 270, 313–321.
- Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H., & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences*, 101, 14812–14817. <https://doi.org/10.1073/pnas.0406166101>
- Hebert, P. D. N., Ratnasingham, S., & de Waard, J. R. (2003b). Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Philosophical Transactions of the Royal Society B*, 270, S96–S99. <https://doi.org/10.1098/rsbl.2003.0025>
- Heimeier, D., Lavery, S., & Sewell, M. A. (2010). Using DNA barcoding and phylogenetics to identify Antarctic invertebrate larvae: Lessons

- from a large-scale study. *Marine Genomics*, 3, 165–177. <https://doi.org/10.1016/j.margen.2010.09.004>
- Hollingsworth, M. L., Clark, A., Forrest, L. L., Richardson, J., Pennington, R. T., Long, D. G., ... Hollingsworth, P. M. (2009a). Selecting barcoding loci for plants: Evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Molecular Ecology Research*, 9, 439–457. <https://doi.org/10.1111/j.1755-0998.2008.02439.x>
- Hollingsworth, P. M., Graham, S. W., & Little, D. P. (2011). Choosing and Using a Plant DNA Barcode. *PLoS ONE*, 6, e19254. <https://doi.org/10.1371/journal.pone.0019254>
- Jeanson, M. L., Labat, J.-N., & Little, D. P. (2011). DNA barcoding: A new tool for palm taxonomists? *Annals of Botany*, 108, 1445. <https://doi.org/10.1093/aob/mcr158>
- Kiew, R. (2002). State of biodiversity studies: Sumatra and Peninsular Malesia compared. *Flora Malesiana Bulletin*, 13, 191–196.
- Knobloch, I. W. (1972). Intergeneric hybridization in flowering plants. *Taxon*, 21, 97–103. <https://doi.org/10.2307/1219229>
- Ko, H. L., Wang, Y. T., Chiu, T. S., Lee, M. A., Leu, M. Y., Chang, K.-Z., ... Shao, K.-T. (2013). Evaluating the accuracy of morphological identification of larval fishes by applying DNA barcoding. *PLoS ONE*, 8(1), e53451. <https://doi.org/10.1371/journal.pone.0053451>
- Koh, L. P., Miettinen, J., Liew, S. C., & Ghazoula, J. (2011). Remotely sensed evidence of tropical peatland conversion to oil palm. *Proceedings of the National Academy of Sciences*, 108, 5127–5132. <https://doi.org/10.1073/pnas.1018776108>
- Kress, W. J., & Erickson, D. L. (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. <https://doi.org/10.1371/journal.pone.0000508>
- Kress, W. J., Erickson, D. L., Andrew-Jones, F., Swenson, N. G., Perez, R., Sanjurjo, O., & Bermingham, E. (2009). Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceedings of the National Academy of Sciences*, 106, 18621–18626. <https://doi.org/10.1073/pnas.0909820106>
- Kress, W. J., Erickson, D. L., Swenson, N. G., Thompson, J., Uriarte, M., & Zimmerman, J. K. (2010). Advances in the use of DNA barcodes to build a community phylogeny for tropical trees in a Puerto Rican forest dynamics plot. *PLoS ONE*, 5, e15409. <https://doi.org/10.1371/journal.pone.0015409>
- Kress, W. J., Garcia-Robledo, C., Uriarte, M., & Erickson, D. L. (2014). DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology & Evolution*, 1880, 1–11.
- Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A., & Janzen, D. H. (2005). Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences*, 102, 8369–8374. <https://doi.org/10.1073/pnas.0503123102>
- Krisnamurthy, P. K., & Francis, R. A. (2012). A critical review on the utility of DNA barcoding in biodiversity conservation. *Biodiversity Conservation*, 21, 1901–1919. <https://doi.org/10.1007/s10531-012-0306-2>
- Lahaye, R., van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., ... Savolainen, P. (2008). DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences*, 105, 2923–2928. <https://doi.org/10.1073/pnas.0709936105>
- Lambert, F. R., & Collar, N. J. (2002). The future for Sundaic lowland forest birds: Long-term effects of commercial logging and fragmentation. *Forktail*, 18, 127–146.
- Laumonier, Y. (1997). *The Vegetation and physiography of Sumatra*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Lim, B. K. (2012). Preliminary assessment of Neotropical mammal DNA barcodes: An underestimation of biodiversity. *The Open Zoology Journal*, 1(3), 10–17. <https://doi.org/10.2174/1874336601205010010>
- Maddox, T., Priatna, D., Gemita, E., & Salampessy, A. (2007). The conservation of tigers and other wildlife in oil palm plantations, Jambi Province, Sumatra, Indonesia. ZSL Conservation Report No. 7. London: The Zoological Society of London.
- Margono, B. A., Turubanova, S., Zhuravleva, I., Potapov, P., Tyukavina, A., Baccini, A., ... Hansen, M. C. (2012). Mapping and monitoring deforestation and forest degradation in Sumatra (Indonesia) using Landsat time series data set from 1990 to 2010. *Environmental Research Letters*, 7, 16. <https://doi.org/10.1088/1748-9326/7/3/034010>
- 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. *Systematic Biology*, 55, 715–728. <https://doi.org/10.1080/10635150600969864>
- Meijer, W. (1981). Sumatra as seen by a botanist. *Indonesia Circle*, 25, 17–27. <https://doi.org/10.1080/03062848108723828>
- Meyer, C. P., & Paulay, G. (2005). DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biology*, 3, e422. <https://doi.org/10.1371/journal.pbio.0030422>
- Miettinen, J., Shi, C., & Liew, S. C. (2011). Deforestation rates in insular Southeast Asia between 2000 and 2010. *Global Change Biology*, 17, 2261–2270. <https://doi.org/10.1111/j.1365-2486.2011.02398.x>
- 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>
- Newmaster, S. G., Ragupathy, S., & Janovec, J. (2009). A botanical renaissance: State-of-the-art DNA barcoding facilitates an Automated Identification Technology system for plants. *International Journal of Computer Applications in Technology*, 35, 50–60. <https://doi.org/10.1504/IJCAT.2009.024595>
- Palmer, J. D., Adams, K. L., Cho, Y., Parkinson, C. L., Qiu, Y.-L., & Song, K. (2000). Dynamic evolution of plant mitochondrial genomes: Mobile genes and introns and highly variable mutation rates. *Proceedings of the National Academy of Sciences*, 97, 6960–6966. <https://doi.org/10.1073/pnas.97.13.6960>
- Parmentier, I., Duminil, J., Kuzmina, M., Philippe, M., Thomas, D. W., Kenfack, D., ... Hardy, O. J. (2013). How effective are DNA barcodes in the identification of African rainforest trees? *PLoS ONE*, 8, e54921. <https://doi.org/10.1371/journal.pone.0054921>
- Pauls, S. U., Blahnik, R. J., Zhou, X., Wardwell, C. T., & Holzenthal, R. W. (2010). DNA barcode data confirm new species and reveal cryptic diversity in Chilean Smicridea (Smicridea) (Trichoptera: Hydropsychidae). *Journal of the North American Benthological Society*, 29, 1058–1074. <https://doi.org/10.1899/09-108.1>
- Rembold, K., Mangopo, H., Tjitrosoedirdjo, S. S., & Kreft, H. (2017). Plant diversity, forest dependency, and alien plant invasions in tropical agricultural landscapes. *Biological Conservation*, 213, 34–242. <https://doi.org/10.1016/j.biocon.2017.07.020>
- Roos, M. C., Keßler, P. J. A., Gradstein, S. R., & Baas, P. (2004). Species diversity and endemism of five major Malaysian islands: Diversity-area relationships. *Journal of Biogeography*, 31, 1893–1908. <https://doi.org/10.1111/j.1365-2699.2004.01154.x>
- Roy, S., Tyagi, A., Shukla, V., Kumar, A., Singh, U. M., Chaudhary, L. B., ... Tuli, R. (2010). Universal plant DNA barcode loci may not work in complex groups: A case study with Indian *Berberis* species. *PLoS ONE*, 5, e13674. <https://doi.org/10.1371/journal.pone.0013674>
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74, 5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>
- Stace, C. A. (2005). Plant taxonomy and biosystematics - does DNA provide all the answers? *Taxon*, 54, 999–1007. <https://doi.org/10.2307/25065484>
- Stoeckle, M. Y., Gamble, C. C., Kirpekar, R., Young, G., Ahmed, S., & Little, D. P. (2011). Commercial teas highlight plant DNA barcode

- identification successes and obstacles. *Scientific Reports*, 1(42), 7. <https://doi.org/10.1038/srep00042>
- Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., ... Willerslev, E. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Research*, 35, e14. <https://doi.org/10.1093/nar/gkl938>
- Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology Evolution*, 10, 512–526. <https://doi.org/10.1093/oxfordjournals.molbev.a040023>
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, 30, 2725–2729. <https://doi.org/10.1093/molbev/mst197>
- Tautz, D., Arctander, P., Minelli, A., Thomas, R. H., & Vogler, A. P. (2003). A plea for DNA taxonomy. *Trends in Ecology & Evolution*, 18, 70–74. [https://doi.org/10.1016/S0169-5347\(02\)00041-1](https://doi.org/10.1016/S0169-5347(02)00041-1)
- Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL-W – Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice. *Nucleic Acids Research*, 22, 4673–4680. <https://doi.org/10.1093/nar/22.22.4673>
- Turner, I. M., Tan, H. T. W., Wee, Y. C., Ibrahim, A. B., Chew, P. T., & Corlett, R. T. (1994). A study of plant species extinction in Singapore: Lessons for the conservation of tropical biodiversity. *Conservation Biology*, 8, 705–712. <https://doi.org/10.1046/j.1523-1739.1994.08030705.x>
- Vaidya, G., Lohman, D. J., & Meier, R. (2011). SequenceMatrix: Concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics*, 27, 171–180. <https://doi.org/10.1111/j.1096-0031.2010.00329.x>
- Valentini, A., Pompanon, F., & Taberlet, P. (2008). DNA barcoding for ecologists. *Trends in Ecology and Evolution*, 24, 110–117. <https://doi.org/10.1016/j.tree.2008.09.011>
- Ward, R. D., Costa, F. O., Holmes, B. H., & Steinke, D. (2008). DNA barcoding of shared fish species from the North Atlantic and Australasia: Minimal divergence for most taxa, but *Zeus faber* and *Lepidopus caudatus* each probably constitute two species. *Aquatic Biology*, 3, 71–78. <https://doi.org/10.3354/ab00068>
- Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., & Hebert, P. D. N. (2005). DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B*, 360, 1847–1857. <https://doi.org/10.1098/rstb.2005.1716>
- Whitten, T., Damanik, S. J., Anwar, J., & Hisyam, N. (2000). *The ecology of Sumatra*. Singapore: Periplus Editions (HK) Ltd.
- World Bank (2001). *Indonesia: Environment and natural resource management in a time of transition*. Washington DC: World Bank.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Amandita FY, Rembold K, Vornam B, et al. DNA barcoding of flowering plants in Sumatra, Indonesia. *Ecol Evol*. 2019;9:1858–1868. <https://doi.org/10.1002/ece3.4875>