



HAL
open science

DNA barcoding in a biodiversity hot spot: potential value for the identification of Malagasy *Euphorbia* L. listed in CITES A ppendices I and II

Xavier Aubriot, Porter P. Lowry, Corinne Cruaud, Arnaud Couloux, Thomas Haevermans

► **To cite this version:**

Xavier Aubriot, Porter P. Lowry, Corinne Cruaud, Arnaud Couloux, Thomas Haevermans. DNA barcoding in a biodiversity hot spot: potential value for the identification of Malagasy *Euphorbia* L. listed in CITES A ppendices I and II. *Molecular Ecology Resources*, 2013, 13 (1), pp.57-65. 10.1111/1755-0998.12028 . hal-03865482

HAL Id: hal-03865482

<https://hal.science/hal-03865482v1>

Submitted on 22 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DNA barcoding in a biodiversity hot spot: potential value for the identification of Malagasy *Euphorbia* L. listed in CITES Appendices I and II

XAVIER AUBRIOT,* PORTER P. LOWRY II,*† CORINNE CRUAUD,‡ ARNAUD COULOUX‡ and THOMAS HAEVERMANS*

*Muséum national d'histoire naturelle, Département Systématique et Evolution, UMR 7205 MNHN/CNRS Origine Structure et Evolution de la Biodiversité (OSEB), Herbarium Plantes Vasculaires, CP 39, 57 rue Cuvier, 75231 Paris CEDEX 05, France,

†Missouri Botanical Garden, PO Box 299, St. Louis, MO 63166–0299, USA, ‡Genoscope, Centre National de Séquençage, CP 5706, 2 rue Gaston Crémieux, 91057 Evry CEDEX, France

Abstract

The island of Madagascar is a key hot spot for the genus *Euphorbia*, with at least 170 native species, almost all endemic. Threatened by habitat loss and illegal collection of wild plants, nearly all Malagasy *Euphorbia* are listed in CITES Appendices I and II. The absence of a reliable taxonomic revision makes it particularly difficult to identify these plants, even when fertile, and thereby compromises the application of CITES regulations. DNA barcoding, which can facilitate species-level identification irrespective of developmental stage and the presence of flowers or fruits, may be a promising tool for monitoring and controlling trade involving threatened species. In this study, we test the potential value of barcoding on 41 *Euphorbia* species representative of the genus in Madagascar, using the two widely adopted core barcode markers (*matK* and *rbcL*), along with two additional DNA regions, nuclear internal transcribed spacer (ITS) and the chloroplastic intergenic spacer *psbA-trnH*. For each marker and for selected marker combinations, inter- and intraspecific distance estimates and species discrimination rates are calculated. Results using just the 'official' barcoding markers yield overlapping inter- and intraspecific ranges and species discrimination rates below 60%. When ITS is used, whether alone or in combination with the core markers, species discrimination increases to nearly 100%, whereas the addition of *psbA-trnH* produces less satisfactory results. This study, the first ever to test barcoding on the large, commercially important genus *Euphorbia* shows that this method could be developed into a powerful identification tool and thereby contribute to more effective application of CITES regulations.

Keywords: conservation, DNA barcoding, *Euphorbia*, illegal trade, *matK* and *rbcL*, *psbA-trnH* and ITS, species identification

Received 30 April 2012; revision received 6 July 2012; accepted 17 July 2012

Introduction

During the last decade, DNA barcoding has been promoted as a new tool to facilitate species-level identification of organisms (Hebert *et al.* 2003a,b) by accelerating and automating the process using short, standardized DNA regions as species tags. Material that might be unidentifiable using traditional morphological methods alone can, by means of barcoding, easily and reliably be assigned to a species. The Plant Working Group of the Consortium for the Barcode of Life (CBOL Plant Working Group 2009) has recently selected a set of markers for

plant barcoding, although they have not yet been tested across the full diversity of plant groups. Given the rapidly increasing use of DNA sequencing as both technical and financial constraints diminish, barcoding is becoming an increasingly promising approach, particularly in the context of large-scale biodiversity surveys for which rapid, accurate identifications are required (Lahaye *et al.* 2008; Vernooij *et al.* 2010). Moreover, there is an urgent need to develop a cost-effective system for easy and reliable identification in fields such as drug discovery, verification of herbals and other plant-based products, ecological studies and the national and international regulation of trade involving biological materials—for example, species regulated by the Convention on International Trade in Endangered Species of Wild Fauna and

Correspondence: Xavier Aubriot, Fax: 33 140 79 33 42; E-mail: aubriot@mnhn.fr

Flora (CITES; Chase *et al.* 2005; Cowan *et al.* 2006; Ogden *et al.* 2009; Valentini *et al.* 2009; Johnson 2010; Yesson *et al.* 2011). However, the practical use of DNA barcoding is still in a developmental stage, and it is therefore crucial to conduct real-life tests to evaluate its effectiveness as an identification tool, particularly in the context of applied uses such as the control and management of threatened and/or protected organisms, for which barcoding is often promoted. To our knowledge, only two studies evaluating the barcoding approach for CITES listed plants have been published to date (see Yesson *et al.* 2011 and Parveen *et al.* 2012), which points to the need for additional research to assess its general utility for the conservation and management of illegally traded endangered plants.

Malagasy *Euphorbia* provide a good model for testing the utility of DNA barcode methods for plant identification. *Euphorbia*, the second largest genus of flowering plants (with *c.* 1850–2000 species), is present in almost every part of the globe (Govaerts *et al.* 2000; Steinmann & Porter 2002; Mabberley 2008). Nevertheless, no reliable taxonomic revision or authoritative reference work is available for the genus, and the most recent comprehensive treatment is that of Boissier (1862, 1866), published 150 years ago. Recent taxonomic and phylogenetic studies of *Euphorbia* have focused on establishing a new subgeneric classification (Bruyins *et al.* 2006; Yang *et al.* 2012) or on revising geographically localized species (Binojkumar & Balakrishnan 2010; Pahlevani & Riina 2011; Pahlevani *et al.* 2011) and thus contribute to an ongoing revision of the entire genus.

The island of Madagascar has at least 170 species of *Euphorbia*, 98% of which are endemic, making it a key hot spot for the genus (Haevermans 2003). Despite this exceptional diversity, no general treatment of Malagasy *Euphorbia* is available, and only few recent taxonomic and phylogenetic works have attempted to encompass the full diversity of Malagasy *Euphorbia* (see Haevermans 2003; Rebmann 2007; and also Rauh 1995, 1998 for succulent taxa).

Micro-endemism is a common phenomenon within Malagasy *Euphorbia*, as many species are restricted to small areas and comprise fragmented populations (Haevermans 2003). Owing both to severe degradation and destruction of their natural habitats (Harper *et al.* 2007) and to excessive pressure from illegal collecting, around 80% of Malagasy *Euphorbia* species are threatened, that is, they are either vulnerable, endangered or critically endangered (Haevermans *et al.* 2004), according to the International Union for Conservation of Nature (IUCN) risk of extinction criteria and categories (IUCN 2001). Many Malagasy species of *Euphorbia* are illegally collected in the wild by unscrupulous collectors (X. Aubriot and T. Haevermans, personal observations) and

internationally traded, presenting significant threats as a result of overexploitation, sometimes leading to the total decimation of entire populations. Consequently, nearly all native species are listed in CITES Appendices I or II (Haevermans *et al.* 2004). All taxa currently listed in Appendix I belong to the endemic subgenus *Lacanthia* (Raf.) M.G.Gilbert (*c.* 110 taxa *sensu* Haevermans *et al.* 2009, 2004), which includes the most critically threatened life forms (i.e. spiny and succulent species as well as geophytes). Moreover, Malagasy *Euphorbia* are notoriously difficult to identify, and most taxa have been described by amateurs in obscure publications, making reliable application of species concepts and identification even more problematical. Consequently, although the genus shows a tremendous range of morphologies, closely related species are often difficult to distinguish from one another, even when in flower or fruit. Because Malagasy *Euphorbia* are highly threatened by habitat loss and illegal collection for commercial purposes and are perplexingly difficult to identify, they constitute an exemplary group for which DNA barcoding might serve as a particularly valuable identification tool for management and control.

The work presented here is based on a representative sample comprising around 25% of the Malagasy *Euphorbia* species, which is used to estimate sequence quality and the discriminatory power of barcode markers. Specifically, we aim to determine whether the currently adopted core barcode markers, *matK* and *rbcL*, are adequate to identify Malagasy *Euphorbia*. We also seek to test the potential value of two additional markers, *psbA-trnH* (Kress *et al.* 2005; CBOL Plant Working Group 2009) and nuclear internal transcribed spacer (ITS; Kress *et al.* 2005; Sass *et al.* 2007), taken alone or as complements to the core markers. Then, we explore the potential use of the DNA barcoding method to help clarify species limit within Malagasy *Euphorbia*. And finally, we examine issues regarding the technical feasibility of DNA barcoding for identifying and managing threatened species, including those covered by CITES.

Materials and methods

Taxon and marker sampling

Recent targeted fieldwork in Madagascar, supplemented by the study of living plants in several European collections (in particular those in Wageningen and Heidelberg), allowed us to compile our core sampling, which covers at least 60% of Malagasy species (i.e. around 100 species). However, given the diversity of the genus in Madagascar (at least 170 spp.) and the current lack of clarity regarding the delimitation of some members of the group, taxon selection is a key issue for any molecular

study aiming to test species concepts. To minimize errors in identification, we therefore limited our sampling for the present study to material that could be identified unambiguously. We also excluded a few specimens belonging to possible new species or for which species-level identification was uncertain. *In fine*, the core sampling comprised material of 41 species (representing around 25% of the total in Madagascar), including members of each of the infrageneric groups defined by Haevermans (2003), with the sole exception of the monotypic *E. antso* group. At least two accessions of each species were used, resulting in a total of 148 samples and 592 sequences (Table 1; Appendix S2, Supporting information). Following the recommendations of the CBOL Plant Working Group (2009), we sequenced the two core barcodes (*matK* and *rbcL*), both of which represent portions of coding plastid genes. In an attempt to explore other potentially useful loci, we also sequenced the plastid intergenic spacer *psbA-trnH* and the nuclear ribosomal ITS region (comprising ITS1, 5.8s and ITS2) (For a complete list of the molecular markers used, see the Appendix S1, Supporting Information).

DNA extraction, amplification, sequencing and data assembly

Total DNA was extracted from leaf fragments, flowers or young green shoots (for leafless individuals), which were dried immediately after collection in silica gel (Chase & Hills 1991). Dried material was ground using a Mixer Mill MM 300™ with tungsten carbide beads (Qiagen Inc.) and with molecular grade Fontainebleau sand. Total DNA was extracted with the Dneasy™ Plant kit (Qiagen Inc.) following the manufacturer's protocol. All PCRs were performed in a total reaction volume of 25 µL containing 15.75 µL of autoclaved ion-exchanged water, 5 µL of Taq & Load Mastermix (MP Biomedicals Inc.), 0.75 µL of DMSO (1 M), 1 µL of each primer (10 µM) and 1.5 µL of unquantified DNA template.

The entire ITS region was amplified using the following PCR protocol: 2-min initial denaturation at 94 °C, 30 cycles of 1-min denaturation at 94 °C, 1-min annealing at 52 °C and 3-min extension at 72 °C followed by a final extension of 7 min at 72 °C. PCRs for *matK* were performed with a 2-min initial denaturation at 94 °C, 32 cycles of 1-min denaturation at 94 °C, 1-min annealing at 53 °C, and 1.5-min extension at 72 °C followed by a final extension of 4-min at 72 °C. PCRs for *rbcL* were carried out with a 1-min initial denaturation at 94 °C, 35 cycles of 1-min denaturation at 94 °C, 1-min annealing at 48 °C, and 1.5-min extension at 72 °C followed by a final extension of 4-min at 72 °C. Finally, PCRs for *psbA-trnH* were performed with a 5-min initial denaturation at 96 °C, 34 cycles of 1-min denaturation at 96 °C, 1-min annealing at

Table 1 List of the Malagasy *Euphorbia* species included in the present study, sorted by infrageneric clade affiliation

| Sub-generic clades (<i>vide</i> Haevermans 2003) | Species | No. of accessions sequenced for all 4 markers |
|---|-----------------------------------|---|
| Clade of <i>E. milii</i> | <i>Euphorbia</i> | 2 |
| | <i>ambarivatoensis</i> | |
| | <i>Euphorbia ankarensis</i> | 2 |
| | <i>Euphorbia banae</i> | 2 |
| | <i>Euphorbia berorohae</i> | 3 |
| | <i>Euphorbia bulbispina</i> | 2 |
| | <i>Euphorbia</i> | 5 |
| | <i>capmanambatoensis</i> | |
| | <i>Euphorbia</i> | 4 |
| | <i>capsaintemariensis</i> | |
| | <i>Euphorbia</i> cf. <i>milii</i> | 5 |
| | <i>Euphorbia cylindrifolia</i> | 3 |
| | ssp. <i>cylindrifolia</i> | |
| | <i>Euphorbia didiereoides</i> | 2 |
| | <i>Euphorbia gotlbebei</i> | 2 |
| | <i>Euphorbia hoffstaetteri</i> | 3 |
| | <i>Euphorbia horombensis</i> | 8 |
| | <i>Euphorbia iharanae</i> | 2 |
| | <i>Euphorbia labatii</i> | 3 |
| | <i>Euphorbia leuconeura</i> | 3 |
| | <i>Euphorbia lophogona</i> | 3 |
| | <i>Euphorbia millotii</i> | 7 |
| | <i>Euphorbia neohumbertii</i> | 2 |
| | <i>Euphorbia paulianii</i> | 2 |
| | <i>Euphorbia rossii</i> | 3 |
| | <i>Euphorbia thuarsiana</i> | 3 |
| | <i>Euphorbia viguieri</i> | 3 |
| <i>Euphorbia waringiae</i> | 3 | |
| Clade of <i>E. pyriformis</i> | <i>Euphorbia ankaranae</i> | 2 |
| | <i>Euphorbia</i> | 6 |
| | <i>mahabokensis</i> | |
| Clade of <i>E. alluaudii</i> | <i>Euphorbia martinae</i> | 2 |
| | <i>Euphorbia alluaudii</i> | 14 |
| Clade of <i>E. cedrorum</i> | <i>Euphorbia cedrorum</i> | 2 |
| | <i>Euphorbia</i> | 4 |
| Clade of <i>E. pachysantha</i> | <i>Euphorbia</i> | 4 |
| | <i>mananarensis</i> | |
| Clade of <i>E. tirucalli</i> | <i>Euphorbia mandravioky</i> | 6 |
| | <i>Euphorbia analalavensis</i> | 2 |
| Clade of <i>E. tetraptera</i> and <i>E. tirucalli</i> | <i>Euphorbia decorsei</i> | 3 |
| | <i>Euphorbia enterophora</i> | 4 |
| | <i>Euphorbia intisy</i> | 3 |
| | <i>Euphorbia kamponii</i> | 2 |
| | <i>Euphorbia</i> | 4 |
| | <i>randrianjohanyi</i> | |
| | <i>Euphorbia rauhii</i> | 4 |
| | <i>Euphorbia tirucalli</i> | 5 |
| Clade of <i>E. denisii</i> | <i>Euphorbia denisii</i> | 4 |
| Clade of <i>E. orthoclada</i> | <i>Euphorbia orthoclada</i> | 4 |
| | 41 Species | 148 |
| | | Accessions |

All four markers (*matK*, *rbcL*, *psbA-trnH* and internal transcribed spacer) were sequenced for each accession.

52 °C, and 0.5-min extension at 72 °C followed by a final extension of 5-min at 72 °C.

Purification and sequencing of PCR products was performed by the National Centre for Sequencing (Génoscope) in Evry, France, using the same primers, except for ITS and *rbcl*, for which a pair of interior primers was used (see Appendix S1, Supporting information). The complete sequences were assembled and manually edited using Sequencher® 4.9 (Gene Codes Corp.). The sequences and voucher information are archived in GenBank (Appendix S2, Supporting information; GenBank Accession number JQ951967–JQ952558).

Data analysis

To estimate the quality of *Euphorbia* sequences, we used a recently published Barcode quality index, B_{30} (Little 2010; Jeanson *et al.* 2011), which takes into account sequence quality and the extent of the overlap between complementary DNA strands of the same sequence. B_{30} values that tend towards 1 indicate that sequence quality is maximal and that overlap between the DNA strands is nearly complete. B_{30} index values were calculated for each sequence (with the exception of those lacking quality values and those with mono-strand sequences), and variation in sequence quality among and within marker was tabulated. We used the pairwise *t*-test in R (R 2.14.0; <http://cran.r-project.org/>) to assess statistical differences in sequence quality among barcode markers.

For each marker, sequences were automatically aligned using MUSCLE 3.6 software (Edgar 2004), visually adjusted as needed and trimmed to the outermost external barcoding primers. However, when the sequences obtained at the 5' and 3' extremities were of poor quality, as was often the case, they were deleted to avoid biases in the analyses. Thus, for the *matK* barcoding region, the first 53 bases and the last 40 bases were excluded from the aligned matrix. Similarly, for *psbA-trnH*, the first and last 47 bases were removed, and the last 14 bases were trimmed off the 3' extremity of ITS sequences. The pairwise *p* distances were calculated and analysed for each region using the MEGA 5.0 Kimura two-parameter (K2P) model (Tamura *et al.* 2011). The frequency distributions of interspecific and intraspecific sequence divergences for each barcode region were then estimated.

To test the rate at which accurate identifications of material were made based on the barcode sequences in our data set, we used the program TAXONDNA (Meier *et al.* 2006), both for each marker taken individually and for selected combinations of two or more markers. By assessing the pairwise *p* distances of an aligned data set, this software determines the closest match of a given sequence by comparing it to all others. To determine the proportion of successful identifications for each marker

taken individually, we used the 'Best match' and 'Best close match' criteria in TAXONDNA. We then assessed the identification scores for the concatenation of the two core barcodes (*matK* and *rbcl*), to which we subsequently added the ITS and *psbA-trnH* sequence data. The 'Best match' method determines the closest match for a given sequence; if the closest match belongs to the same species the identification is considered a success, if it belongs to another species it is regarded as a misidentification and if several equally good 'best matches' are recovered referring to two or more different species, the identification is scored as ambiguous. Using the frequency distribution of the intraspecific pairwise distances, the TAXONDNA program calculates a similarity threshold value below which are found 95% of all intraspecific pairwise distances (Meier *et al.* 2006; Newmaster *et al.* 2008). The 'Best close match' method identifies all barcode matches whose value exceeds this threshold; these are regarded as unidentified. If a barcode match is lower than the threshold value, the results are interpreted in a manner similar to that of the 'Best match' criteria.

Results

Sequence quality

The best sequence qualities were obtained for *psbA-trnH* (mean B_{30} = 0.86) and the two core barcodes *matK* and *rbcl* (mean B_{30} = 0.85 for both); 75% of the sequences obtained from *rbcl*, *matK* and *psbA-trnH* had a B_{30} value above ca. 0.80. Sequence quality for these three markers was not statistically different (Pairwise *t*-test, P = 0.05). Internal transcribed spacer had the lowest sequence quality (mean B_{30} = 0.74), a value that is statistically different from those of the other three markers (Pairwise *t*-test, P = 0.05); 25% of ITS sequences had a low B_{30} , falling below 0.66 (Fig. 1).

Sequence divergence and identification success

The two core barcode markers (matK and rbcl). The aligned *rbcl* matrix (excluding external primers) was 1450 pb in length, with 70 variable sites and no indels. Interspecific distances (ranging from 0 to 0.027) were generally larger than intraspecific values (0–0.002). However, no discontinuity was found between intra- and interspecific ranges. Around 30% of the interspecific distances are below 0.002 (Fig. 2). On its own, *rbcl* is of very limited value for correctly matching sequences to species: under both the 'best match' and 'best close match' models, <50% of the *rbcl* sequences were correctly identified and most of the identifications (*c.* 55%) were ambiguous (Table 2).

The *matK* region was 825 pb long when aligned, with 121 variable sites and two gaps (resulting from insertions

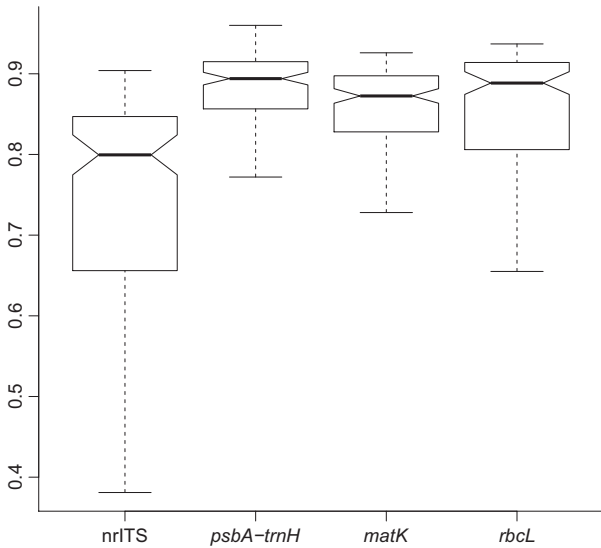


Fig. 1 Box plots of sequence quality (B_{30}) (y -axis), for each of the four barcoding markers used (x -axis).

present in some samples). Interspecific distances (0–0.094) were indistinguishable from intraspecific values (0–0.005), with ca. 27% of the interspecific distances

below 0.005 (Fig. 2). As with *rbcL*, *matK* alone does not perform well for making identifications based on sequences, with only c. 50% proving successful.

When data for the two core barcodes are concatenated, they still performed rather poorly, with only c. 60% successful identifications. Intraspecific distances (0–0.003) overlapped with interspecific distances (0–0.05), with around 23% of the latter below 0.003.

Additional markers: psbA-trnH and ITS. When aligned, the *psbA-trnH* region was 1128 pb in length, with 219 variable sites, numerous indels, and a single inversion (from bases 69 to 92). The effect of inversions on the barcode analyses has been documented in several studies (Whitlock *et al.* 2010; Jeanson *et al.* 2011). Depending on the data set, correcting for inversion resulted either in a positive impact on species identification (Whitlock *et al.* 2010) or a negative one (Jeanson *et al.* 2011). To estimate the effects of the inversion found in *psbA-trnH* sequences on the identification of Malagasy *Euphorbia* species, we first conducted our analyses using an ‘uncorrected’ *psbA-trnH* matrix and then a ‘corrected’ matrix, in which the sequences of the inversion were manually reinverted.

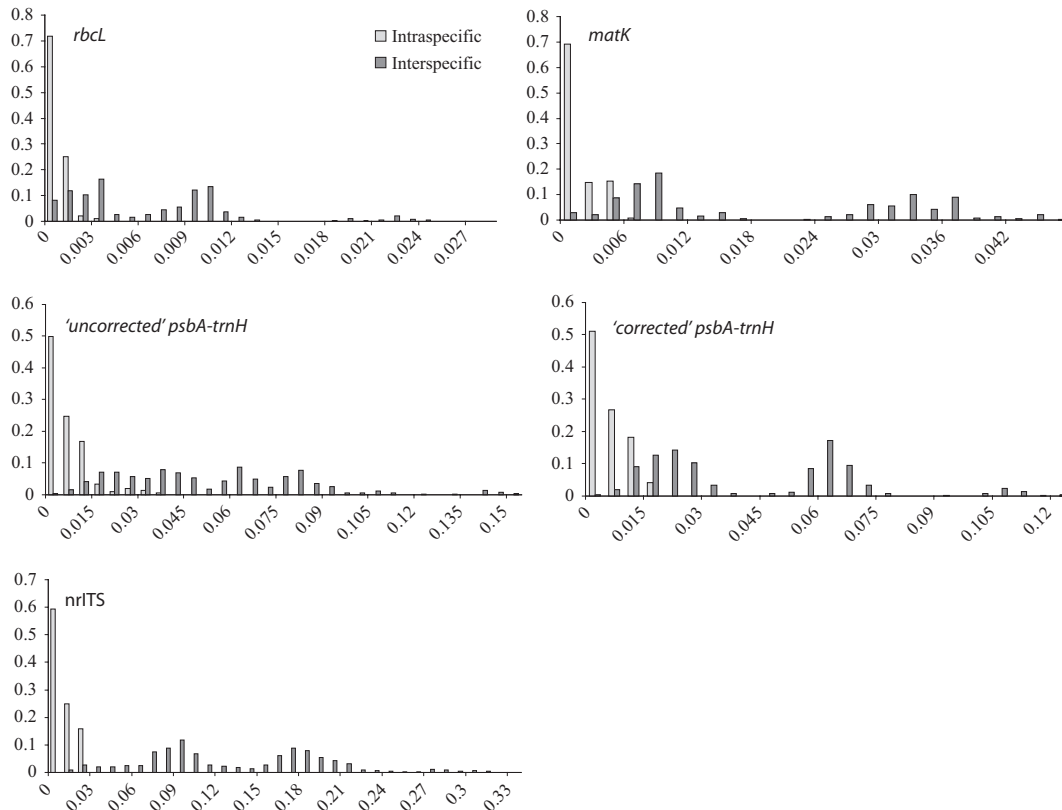


Fig. 2 Histograms of the frequencies (y -axes) of pairwise intra- and interspecific divergence distances (x -axes) among all 148 samples of Malagasy *Euphorbia* for each barcoding marker.

Table 2 Identification success of Malagasy *Euphorbia* based on the 'best match' and 'best close match' functions of the program TAXON-DNA (Meier *et al.* 2006) for each barcoding marker and for selected combinations of two or more markers

| Regions | Best match | | | Best close match | | | No match (%) | Threshold (%) |
|--|-----------------------------|------------------|-------------------|-----------------------------|-------------------|-------------------|----------------|---------------|
| | Successfully identified (%) | Ambiguous (%) | Misidentified (%) | Successfully identified (%) | Ambiguous (%) | Misidentified (%) | | |
| <i>matK</i> | 72 (48.64) | 70 (47.29) | 6 (4.05) | 71 (47.97) | 70 (47.29) | 6 (4.05) | 1 (0.67) | 0.27 |
| <i>rbcL</i> | 63 (42.56) | 82 (55.4) | 3 (2.02) | 58 (39.18) | 76 (51.35) | 3 (2.02) | 11 (7.43) | 0.06 |
| <i>matK</i> + <i>rbcL</i> | 88 (59.45) | 55 (37.16) | 5 (3.37) | 87 (58.78) | 54 (36.48) | 5 (3.37) | 2 (1.35) | 0.13 |
| <i>psbA-trnH</i> 'non-corrected' | 118 (79.72) | 17 (11.48) | 13 (8.78) | 118 (79.72) | 17 (11.48) | 13 (8.78) | 0 (0.0) | 1.90 |
| <i>psbA-trnH</i> 'corrected' | 119 (80.4) | 16 (10.81) | 13 (8.78) | 119 (80.4) | 16 (10.81) | 13 (8.78) | 0 (0.0) | 0.93 |
| <i>matK</i> + <i>rbcL</i> + <i>psbA-trnH</i> 'non-corrected' | 123 (83.1) | 11 (7.43) | 14 (9.45) | 123 (83.1) | 11 (7.43) | 14 (9.45) | 0 (0.0) | 0.50 |
| <i>matK</i> + <i>rbcL</i> + <i>psbA-trnH</i> 'corrected' | 125 (84.45) | 12 (8.1) | 11 (7.43) | 125 (84.45) | 12 (8.1) | 11 (7.43) | 0 (0.0) | 0.33 |
| nrITS | 147 (99.32) | 0 (0.0) | 1 (0.67) | 147 (99.32) | 0 (0) | 1 (0.67) | 0 (0.0) | 1.30 |
| <i>matK</i> + <i>rbcL</i> + nrITS | 147 (99.32) | 0 (0.0) | 1 (0.67) | 147 (99.32) | 0 (0.0) | 1 (0.67) | 0 (0.0) | 0.40 |
| <i>matK</i> + <i>rbcL</i> + <i>psbA-trnH</i> 'non-corrected' + nrITS | 143 (96.62) | 2 (1.35) | 3 (2.02) | 140 (94.59) | 2 (1.35) | 2 (1.35) | 4 (2.7) | 0.54 |
| <i>matK</i> + <i>rbcL</i> + <i>psbA-trnH</i> 'corrected' + nrITS | 145 (97.97) | 2 (1.35) | 1 (0.67) | 145 (97.97) | 2 (1.35) | 1 (0.67) | 0 (0.0) | 0.48 |

Highest values for each barcoding marker and for each function of the program ('best match' and 'best close match') are given in bold. The preferred barcoding option for identification of the Malagasy *Euphorbia* is highlighted in grey. ITS, internal transcribed spacer.

As expected, distances between sequences were lower when using the 'corrected' *psbA-trnH* matrix, which yielded interspecific values ranging from 0 to 0.119 as compared to distances from 0 to 0.150 for the 'uncorrected' matrix. Intraspecific values were 0–0.013 using the 'corrected' *psbA-trnH* matrix vs. 0–0.032 for the 'uncorrected' matrix. The overlap between interspecific and intraspecific distances is greater when using the 'uncorrected' matrix (35% of the interspecific distances below 0.032) than for the 'corrected' matrix (22% of the interspecific distances below 0.013; Fig. 2).

The *psbA-trnH* region, whether 'corrected' or not for the inversion, provides a comparatively high reliability of identification, with a success rate of *c.* 80% (the rate of accurate identification being slightly better using the 'corrected' matrix). However, when the *psbA-trnH* region is concatenated with the two core barcodes, identification success increases only slightly and does not exceed 85% (Table 2).

The aligned ITS region was 713 pb in length, with 337 variable sites and several indels. Interspecific distances (0–0.324) showed little overlap with intraspecific values (0–0.015), and only 2% of the interspecific values were below 0.015 (Fig. 2). The ITS region was the most efficient of the four markers examined at successfully discriminating species, with a rate of nearly 100%, which was unchanged with the addition of *matK* and *rbcL* (Table 2).

When all four barcode regions (*rbcL*, *matK*, *psbA-trnH*, and ITS) were concatenated together, the rate of accurate identification dropped slightly to *c.* 97%. As for

psbA-trnH alone, identification power of the fully concatenated data set was slightly better using the 'corrected' *psbA-trnH* matrix.

Discussion

The illegal exploitation and sale of wildlife represents a flourishing multi-billion dollar industry that contributes significantly to the mounting threats faced by thousands of plant and animal species (Johnson 2010). The ability of countries that have signed the CITES convention to regulate international trade of biological materials is, however, severely hampered by a lack of reliable procedures to make accurate identifications of samples that require verification. Barcoding offers the prospect of providing just such a tool. Our results indicate that this method indeed provides a powerful way to identify species of *Euphorbia* from Madagascar, and they suggest that barcoding may be useful for discriminating the *c.* 2000 species that comprise the genus as a whole, including the succulent taxa whose trade is regulated under CITES.

The successful application of barcoding requires high-quality DNA sequences (CBOL Plant Working Group 2009; Little 2010). For the sample of Malagasy *Euphorbia* studied here, mean sequence quality (mean B_{30}) ranged between 0.74 and 0.86 (with a minimum of 0.38 and a maximum of 0.96), values that are consistent with those obtained in other studies, which fell between 0.40 and 0.93 (Little 2010; Jeanson *et al.* 2011). These values also match the criteria indicated by the CBOL Plant Working

Group (2009) as sufficiently reliable for barcoding. However, species discrimination is ambiguous when only the two core barcode markers, *matK* and *rbcL*, are used, whether individually or together. This is consistent with earlier reports; the limited power of *matK* and *rbcL* has been observed in several other plant groups, including the family *Myristicaceae* (Newmaster *et al.* 2008) and the palm tribe *Caryoteae* (Jeanson *et al.* 2011). When two supplemental markers (ITS and *psbA-trnH*) were used, however, the accuracy of identification increased noticeably. Despite the fact that ITS had the lowest sequence quality of the four markers studied, it nevertheless provides a high level of successful identification. Identification success using ITS approached 100%, whether used alone or in combination with the two core barcode markers, far exceeding the 72% rate obtained by the core markers promoted by the CBOL Plant Working Group (2009). Our results also show that ITS is much more informative than *psbA-trnH*, and should be used preferentially for the barcoding of Malagasy *Euphorbia*. Moreover, the numerous insertions/deletions found in *psbA-trnH* make sequence alignment difficult and compromise its potential value as a barcode region (Lahaye *et al.* 2008; Jeanson *et al.* 2011). As the inversions in *psbA-trnH* do not appear to be species-specific within Malagasy *Euphorbia*, if this marker is used, the matrix should always either be 'corrected' or the portion of the sequence corresponding to the insertions should be deleted. Whitlock *et al.* (2010), in a study of six species of *Gentianaceae*, showed that maintaining inversions in a data set can lead to an overestimate of intraspecific distances, whereas Jeanson *et al.* (2011) found that correcting for inversion tends to compromise the diagnostic value of a mostly interspecific set of bases and thereby lower the identification power of *psbA-trnH* region. Regardless of whether correction for inversions is required in a particular group, the fact that *psbA-trnH* cannot universally be used without first considering this constraint limits its overall utility as a barcode marker.

An important potential drawback of ITS as a barcoding region is that it may be subject to paralogy, that is, the presence of multiple copies that may not be identical to one another, a situation that has been reported in the genome of *Euphorbia* (Yang & Berry 2011). Sampling of divergent paralogues for a given specimen and/or species would potentially lead to misidentification, reducing the value of ITS as a barcode marker. Notwithstanding this potential problem, we observed only a single ITS band on the agarose gels we ran, and we did not obtain overlapping signals on any of the corresponding electropherograms, suggesting that at least in the case of Malagasy *Euphorbia*, either ITS is represented by a single copy or if multiple copies exist, their sequences have undergone concerted evolution.

Based on a broad study of numerous groups, the CBOL Plant Working Group (2009) formally designated *matK* and *rbcL* as its 'standard 2-locus barcode'. Our results show, however, that the concatenation of these two 'official' barcodes with ITS does not increase identification power over ITS when used alone. Thus, at least for Malagasy *Euphorbia*, the use of *matK* and *rbcL* as barcode markers to support application of the CITES convention would be superfluous. Internal transcribed spacer offers a reliable barcoding option for rapid, economic and accurate species discrimination among the Malagasy *Euphorbia* included in our study.

Considering the high rate of successful identification we obtained using ITS, DNA barcoding may also be of value for clarifying the chaotic taxonomy and species circumscription among Malagasy *Euphorbia*. Recent results based on molecular data have shown that several species and subspecies, as currently delimited, are para- or polyphyletic (Aubriot, unpublished data). In particular, DNA barcoding may help elucidate relationships within certain groups such as the *E. pyriformis* Lam. clade, in which most species concepts are unclear and a critical taxonomic assessment is badly needed (Haevermans 2003). The contribution of DNA barcoding to resolving taxonomic issues within the *E. pyriformis* clade will be addressed in a forthcoming paper.

Given the tremendous taxonomic and structural diversity found in *Euphorbia* and considering the chronic difficulty of accurately identifying material based on morphology alone, DNA barcoding may be a useful identification tool for the genus worldwide. The utility of several markers, including ITS, *psbA-trnH*, *matK* and *rbcL*, for phylogenetic analysis of *Euphorbia* has already been demonstrated in several studies that successfully employed extensive sampling without noticeable problems of amplification, sequencing or alignment (Steinmann & Porter 2002; Bruyns *et al.* 2006; Zimmermann *et al.* 2010; Yang & Berry 2011; Horn *et al.* 2012). The expanded use of these markers for further efforts to explore phylogenetic relationships within the genus seems assured, but our results suggest that ITS alone may be sufficient for barcoding efforts aimed at facilitating the accurate identification of species of *Euphorbia*.

Currently, all succulent species of *Euphorbia* are covered under the CITES convention. Yet, it is not always possible to determine whether a plant, cutting, seed or other propagule belongs to a succulent species, and as a consequence, material of other species may erroneously be subjected to CITES control. Moreover, some succulent *Euphorbia* currently covered under CITES are in fact common and widespread, and international trade does not represent a threat, yet it is often impossible to distinguish them from other succulent species that are highly threatened. Reliable

barcoding procedures for all Malagasy *Euphorbia* could help resolve these issues, enabling control for those species that require it while allowing trade for those that are not threatened, and thereby providing a much-needed source of sustainable income for Madagascar.

Acknowledgements

We thank the CNRE (Centre National de Recherche sur l'Environnement), MNP (Madagascar National Parks) and the local office of the Missouri Botanical Garden, all in Antananarivo, for in-country assistance and for help in obtaining permits. We also wish to thank Wilbert Hetterscheid for help with collecting some crucial *Euphorbia* samples. The authors are grateful to Frederic Austerlitz, Damon Little and Peter Hollingsworth for the technical assistance and thoughtful comments. We thank the BoEM molecular laboratory (MNHN, Paris) for welcoming us and providing support for the molecular work. Financial support for field and laboratory work was provided by the MNHN ATM 'Taxonomie moléculaire: DNA barcode et gestion durable des collections' and grants from Sud Expert Plantes and ANR (EVORANGE 6^{ème} extinction: ANR-09-PEXT-011). This work is part of the @SPEED-ID project entitled 'Accurate SPEciEs Delimitation and IDentification of eukaryotic biodiversity using DNA markers' led by F-BoL, the French Barcode of Life initiative.

References

- Binojkumar MS, Balakrishnan NP (2010) *The Genus Euphorbia L. (Euphorbiaceae) in India. A Taxonomic Revision*. Bishen Singh Mahendra Pal Singh, New Delhi, 430 pp.
- Boissier PE (1862) Euphorbiaceae subordo I. Euphorbieae. In: A. DC., *Prodromus*, **15**, 3–188.
- Boissier PE (1866) Euphorbiaceae subordo I. Euphorbieae. (Addenda et Corrigenda). In: A. DC., *Prodromus*, **15**, 1261–1269.
- Bruyns PV, Mapaya RJ, Hedderson TJ (2006) A new subgeneric classification for *Euphorbia* (Euphorbiaceae) in southern Africa based on ITS and *psbA-trnH* sequence data. *Taxon*, **55**, 397–420.
- 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.
- Chase MW, Hills HH (1991) Silica gel: an ideal material for field preservation of leaf samples for DNA studies. *Taxon*, **40**, 215–220.
- Chase MW, Salamin N, Wilkinson M *et al.* (2005) Land plants and DNA barcodes: short-term and long-term goals. *Philosophical Transactions of the Royal Society. Series B, Biological Sciences*, **360**, 1889–1895.
- Cowan RS, Chase MW, Kress WJ, Savolainen V (2006) 300,000 Species to identify: problems, progress, and prospects in DNA barcoding of land plants. *Taxon*, **55**, 611–616.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.
- Govaerts R, Frodin DG, Radcliffe-Smith A (2000) *World Checklist and Bibliography of Euphorbiaceae (and Pandanaceae)*. Royal Botanic Gardens, Kew, 1661 pp.
- Haevermans T (2003) *Le genre Euphorbia à Madagascar: phylogénie moléculaire et systématique*. Ph.D. Dissertation, MNHN, Paris, France.
- Haevermans T, Hoffmann P, Lowry PP II, Labat JN, Randrianjohany E (2004) Phylogenetic analysis of the Madagascan *Euphorbia* subgenus *Lacanthis* based on ITS sequence data. *Annals of the Missouri Botanical Garden*, **91**, 247–259.
- Haevermans T, Rouhan G, Hetterscheid W *et al.* (2009) Chaos revisited: nomenclature and typification of the Malagasy endemic *Euphorbia* subgenus *Lacanthis* (Raf.) MG Gilbert. *Adansonia*, **31**, 279–299.
- Harper GJ, Steininger MK, Tucker CJ, Juhn D, Hawkins F (2007) Fifty years of deforestation and forest fragmentation in Madagascar. *Environmental Conservation*, **34**, 325–333.
- Hebert PDN, Ratnasingham S, de Waard JR (2003a) Barcoding animal life: cytochrome c oxidase subunit I divergences among closely related species. *Proceedings of the Royal Society. Series B, Biological Sciences*, **270**, S96–S99.
- Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003b) Biological identifications through DNA barcodes. *Proceedings of the Royal Society. Series B, Biological Sciences*, **270**, 313–321.
- Horn JW, van Ee BW, Morawetz JJ *et al.* (2012) Phylogenetics and the evolution of major structural characters in the giant genus *Euphorbia* L. (Euphorbiaceae). *Molecular Phylogenetics and Evolution*, **63**, 305–326.
- IUCN (2001) *IUCN Red List Categories and Criteria, Version 3.1*. IUCN, Gland, Switzerland & Cambridge, ii + 30 p.
- Jeanson ML, Labat JN, Little DP (2011) DNA barcoding: a new tool for palm taxonomists? *Annals of Botany*, **108**, 1445–1451.
- Johnson RN (2010) The use of DNA identification in prosecuting wildlife-traffickers in Australia: do the penalties fit the crimes? *Forensic Science, Medicine, and Pathology*, **6**, 211–216.
- Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 8369–8374.
- Lahaye R, Van Der Bank M, Bogarin D *et al.* (2008) DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences*, **105**, 2923–2928.
- Little DP (2010) A unified index of sequence quality and contig overlap for DNA barcoding. *Bioinformatics*, **26**, 2780–2781.
- Mabberley DJ (2008) *Mabberley's Plant-Book*, 3rd edn. Cambridge University Press, Cambridge, 1040 pp.
- Meier R, Shiyang K, Vaidya G, Ng PKL (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology*, **55**, 715–728.
- Newmaster S, Fazekas A, Steeves R, Janovec J (2008) Testing candidate plant barcode regions in the Myristicaceae. *Molecular Ecology Resources*, **8**, 480–490.
- Ogden R, Dawnay N, McEwing R (2009) Wildlife DNA forensics—bridging the gap between conservation genetics and law enforcement. *Endangered Species Research*, **9**, 179–195.
- Pahlevani A, Riina R (2011) A synopsis of *Euphorbia* subgen. *Chamaesyce* (Euphorbiaceae) in Iran. *Annales Botanici Fennici*, **48**, 304–316.
- Pahlevani A, Geltman D, Riina R (2011) Taxonomic revision of *Euphorbia* subsect. *Myrsinitae* in Iran. *Annales Botanici Fennici*, **48**, 483–493.
- Parveen I, Singh HK, Raghuvanshi S, Pradhan UC, Babbar SB (2012) DNA barcoding of endangered Indian *Paphiopedilum* species. *Molecular Ecology Resources*, **12**, 82–90.
- Rauh W (1995) *Succulent and Xerophytic Plants of Madagascar. Volume One*. Strawberry Press, Mill Valley, 343 pp.
- Rauh W (1998) *Succulent and Xerophytic Plants of Madagascar. Volume Two*. Strawberry Press, Mill Valley, 386 pp.
- Rebmann N (2007) Les Euphorbes de Madagascar. Succulentes, Special Issue, 48 pp.
- Sass C, Little DP, Stevenson DW, Specht CD (2007) DNA barcoding in the Cycadales: testing the potential of proposed barcoding markers for species identification of Cycads. *PLoS ONE*, **2**, e1154.
- Steinmann VW, Porter M (2002) Phylogenetic relationships in Euphorbieae (Euphorbiaceae) based on ITS and *ndhF* sequence data. *Annals of the Missouri Botanical Garden*, **89**, 453–490.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.
- Valentini A, Pompanon F, Taberlet P (2009) DNA barcoding for ecologists. *Trends in Ecology & Evolution*, **24**, 110–117.
- Vernooy R, Haribabu E, Muller MR *et al.* (2010) Barcoding life to conserve biological diversity: beyond the taxonomic imperative. *PLoS Biology*, **8**, e1000417.

- Whitlock BA, Hale AM, Groff PA (2010) Intraspecific inversions pose a challenge for the *trnH-psbA* plant DNA barcode. *PLoS ONE*, **5**, e11533.
- Yang Y, Berry PE (2011) Phylogenetics of the Chamaesyce clade (*Euphorbia*, Euphorbiaceae): reticulate evolution and long-distance dispersal in a prominent C4 lineage. *American Journal of Botany*, **98**, 1486–1503.
- Yang Y, Riina R, Morawetz JJ, Haevermans T, Aubriot X, Berry PE (2012) Molecular phylogenetics and classification of *Euphorbia* subgenus *Chamaesyce*. *Taxon*, **61**, 764–789.
- Yesson C, Bárcenas RT, Hernández HM *et al.* (2011) DNA barcodes for Mexican Cactaceae, plants under pressure from wild collecting. *Molecular Ecology Resources*, **11**, 775–783.
- Zimmermann NFA, Ritz CM, Hellwig FH (2010) Further support for the phylogenetic relationships within *Euphorbia* L. (Euphorbiaceae) from nrITS and trnL-trnF IGS sequence data. *Plant Systematics and Evolution*, **286**, 39–58.

X.A. carried out the study. X.A. and T.H. conducted the DNA extraction and amplification. Sequencing was coordinated by C.C. and A.C. X.A. prepared the first draft of the paper; T.H. and P.P.L. II participated in finalizing text.

Data Accessibility

DNA sequences: GenBank Accessions JQ951967–JQ952558.

Please see the Supporting information available with the online version of this manuscript (Appendix S2).

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Primer identities, sequences and source for the three chloroplast loci and nuclear ITS used for PCR and sequencing in this study.

Appendix S2 Details of voucher specimens along with GenBank accession numbers for individual loci.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.