

RESEARCH ARTICLE

The first initiative of DNA barcoding of ornamental plants from Egypt and potential applications in horticulture industry

Hosam O. Elansary^{1*}, Muhammad Ashfaq², Hayssam M. Ali^{3,4}, Kowiyou Yessoufou⁵

1 Floriculture, Ornamental Horticulture and Garden Design Department, Faculty of Agriculture (El-Shatby), Alexandria University, Alexandria, Egypt, **2** Biodiversity Institute of Ontario, University of Guelph, ON, Guelph, Canada, **3** Botany and Microbiology Department, College of Science, King Saud University, Riyadh, Saudi Arabia, **4** Timber Trees Research Department, Sabahia Horticulture Research Station, Horticulture Research Institute, Agriculture Research Center, Alexandria, Egypt, **5** Department of Geography, Environmental Management and Energy Studies, University of Johannesburg, APK campus, Johannesburg, South Africa

* hosammail2003@yahoo.com, helansary@alexu.edu.eg



OPEN ACCESS

Citation: O. Elansary H, Ashfaq M, Ali HM, Yessoufou K (2017) The first initiative of DNA barcoding of ornamental plants from Egypt and potential applications in horticulture industry. PLoS ONE 12(2): e0172170. doi:10.1371/journal.pone.0172170

Editor: Diego Breviario, Istituto di Biologia e Biotechnologia Agraria Consiglio Nazionale delle Ricerche, ITALY

Received: October 16, 2016

Accepted: January 31, 2017

Published: February 15, 2017

Copyright: © 2017 O. Elansary et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This project was supported by the King Saud University, Deanship of Scientific Research, College of Science Research Center. Sequence analysis was funded by the Government of Canada through Genome Canada and the Ontario Genomics Institution in support of the International Barcode of Life Project. The Faculty of Agriculture,

Abstract

DNA barcoding relies on short and standardized gene regions to identify species. The agricultural and horticultural applications of barcoding such as for marketplace regulation and copyright protection remain poorly explored. This study examines the effectiveness of the standard plant barcode markers (*matK* and *rbcl*) for the identification of plant species in private and public nurseries in northern Egypt. These two markers were sequenced from 225 specimens of 161 species and 62 plant families of horticultural importance. The sequence recovery was similar for *rbcl* (96.4%) and *matK* (84%), but the number of specimens assigned correctly to the respective genera and species was lower for *rbcl* (75% and 29%) than *matK* (85% and 40%). The combination of *rbcl* and *matK* brought the number of correct generic and species assignments to 83.4% and 40%, respectively. Individually, the efficiency of both markers varied among different plant families; for example, all palm specimens (Arecaceae) were correctly assigned to species while only one individual of Asteraceae was correctly assigned to species. Further, barcodes reliably assigned ornamental horticultural and medicinal plants correctly to genus while they showed a lower or no success in assigning these plants to species and cultivars. For future, we recommend the combination of a complementary barcode (e.g. *ITS* or *trnH-psbA*) with *rbcl* + *matK* to increase the performance of taxa identification. By aiding species identification of horticultural crops and ornamental palms, the analysis of the barcode regions will have large impact on horticultural industry.

Introduction

Global horticultural industry is one of the fastest growing industries in agricultural sector. According to the US Bureau of Economic Analysis, floriculture related sales in the USA alone in 2012 were USD 27.8 billion while the sales for the global industry surpassed \$60 billion. Unfortunately, the global market of horticultural industry is compromised by a wide range of

Alexandria University funded the collection of material and delivery costs.

Competing interests: The authors have declared that no competing interests exist.

counterfeited ornamental and fruit plants that have been sold without anyone paying intellectual propriety rights or following plant variety protection (PVP) laws [1,2]. Intellectual property infringements in horticultural crops may lead to large economic losses for plant breeders including small and medium size companies and public research institutes whose main revenues and license fees are paid by authorized producers, while illegal traders ignore the payment of such fees, and this results in negative impact not only on the producer but also on the society and global trade [3]. In the face of counterfeited ornamental plants and many other illegal activities in the industry, the development of reliable methods to distinguish among species or specimens of ornamental plants, fruits and vegetables may help in informing and enforcing the market regulations [4]. Traditionally, most plant identifications are based on morphological characters, but such identification is not always reliable and efficient [5]. A wide range of molecular techniques including (but not limited to) random amplified polymorphic DNA (RAPD) [6], amplified fragment length polymorphism (AFLP) [7], restriction fragment length polymorphism (RFLP) [8], microsatellite [9] and single nucleotide polymorphism (SNP) [10] have been proposed to identify plant species/specimen and cultivars. DNA barcoding has emerged as a relatively novel and perhaps more universal tool with which to analyze diversity of both plants and animals and to assign specimens to their species even in the absence of all or key morphological diagnostic features [11, 12]. Although there are still some reserves against the performance of DNA barcoding as compared, for example, to morphology, an early study, through a thorough comparison of DNA barcoding and morphology-based species identification recorded a number of limitations to the morphology particularly when it comes to cryptic species [5].

The earliest use of DNA barcoding to identify insect species [11] has triggered a global campaign that mobilizes scientists and institutions for biodiversity, ecology and phylogenetic studies [13–17]. The technique has become an acceptable taxonomic tool [18] and has been successfully used in large scale biodiversity projects where regional flora and fauna are documented [15,17,19], including regulated and threatened taxa [20]. Although a number of plant loci including, *trnH-psbA* [21], *rpoC1*, *rpoB* [22], *trnL* [23], *rbcL* [24] and *matK* [25] were initially proposed as potential plant barcodes based on assessments of recoverability, sequence quality and levels of species discrimination, the Consortium for the Barcode of Life [12] recommended the 2-loci combination of *rbcL + matK* as the standard plant barcode. However, there are persistent calls to include ITS into the core barcodes [26–28].

This combination *rbcL + matK* has been successful in several specimen identification campaigns across continents such as the barcoding campaign of the African rainforest trees in Cameroon [29], the trees and shrubs of Egypt [13], the forest trees in Panama [30], and the poorly known flora of Australia [31]. Similarly, several projects to barcode specific taxa such as the horticultural crops like *Ocimum* [32], *Ficus* [13,33], *Rhododendron* [34] and Araliaceae [35] have been initiated.

Specifically, some attention has been devoted to barcoding medicinal plants in China [36], and on the African continent it is only in South Africa that exceptional effort to DNA-barcode local and regional floras has been made [15,16,28]. However, in the northern Africa particularly in Egypt, a promising country for the production of ornamental crops [37] and a well-known country for its medicinal and horticultural plant diversity [38], the barcoding effort of local and regional floras is yet to be fueled although the applications of DNA barcoding a wide range of scientific disciplines are mounting; e.g. invasion ecology [4,39], biodiversity assessment [15], conservation efforts [16,40], and phylogeographic studies [40].

The present study is the first initiative of its kind in Egypt and northern Africa aiming to barcode the Egyptian ornamental herbs and palms in private and public nurseries. Specifically, we used the core DNA barcodes (*rbcLa + matK*) data that we generated to explore the

resolution power of each marker in taxa and specimen identification. The DNA barcode data generated in the present study will serve in the future in commercial agricultural, horticultural and medicinal plant industries for the purpose of control of counterfeited product, and could also serve in ecological studies of local flora as demonstrated elsewhere [4,16,40].

Materials and methods

Plant material and tissue sampling

We collected 225 plant specimens from 161 taxa; of this collection, 121 specimens were sampled from the Green Oasis Nursery in Alexandria, 85 from the nursery of the Faculty of Science, Meharam Bek, Alexandria, 12 from Ashor Nursery in Montaza area in Alexandria, 5 from Mostafa Kamel Village Nursery in Mariout area in Alexandria and 2 from Antoniadis garden in Alexandria (May 2016). All owners of the nurseries and gardens approved the sampling and publishing of the data and none of the plants were endangered or protected species. To examine the success of barcoding on ornamental plant cultivars several individuals belonging to the same species and differing in flower or leaf color were collected, and this included *Viola tricolor* (Hornveilchen lila, Frosthart, Hortensis, L., Heartsease, Hornveilchen hellgelb, Simon Shine, Sun Glory, Freefall Purple & White) and *Pericallis x hybrida* (Senetti Blue Bicolor, Senetti Magenta, Senetti Super Blue, Senetti Pink, Jester Pure White) etc.. Samples collections, analyses and vouchering were completed in May 2016. These specimens were geo-referenced with digital pictures and leaf samples were dried in silica gel for subsequent analysis. Specimen information along with images is available on Egypt barcode of life project (www.boldsystems.org) and [S1 Table](#).

DNA extraction, PCR and sequencing

DNA extraction, PCR amplification and sequencing were performed at the Canadian Centre for DNA Barcoding (CCDB) of the Biodiversity Institute of Ontario. DNA extractions, PCR amplifications and sequencing were performed following CCDB protocols ([S1](#) and [S2](#) Sheets). The following primers were used for amplification and sequencing: *rbcl*: *rbcl*-F (TGTCACCACAAACAGAGACTAAAGC) [41], *rbcl*-R (GTAAAATCAAGTCCACCRG) and *matK*: *MatK*-1RKIM-f (CCCAGTCCATCTGGAAATCTTGGTTC), *MatK*-3FKIM-r (GTACAGTACTTTTGTGTTTACGAG). The forward and reverse trace files were trimmed and assembled after sequencing using the CodonCode Aligner V 3.5.4 (CodonCode Co., USA). All the sequences generated are available on Genbank/EBI (*matK* accession No. KX783623—KX783811; *rbcl* accession No. KX783812–KX784028).

Statistical analysis

BLAST tests against the GenBank database were performed for identification of specimens at family, genus and species levels and the resolution efficiency was determined based on Blast1 method (BLAST1: the ID is that of the species associated with the best BLAST hit, and E-value < cut-off. This corresponds to choosing the top hit in the BLAST results) [42]. The correct identification means that the individual is assigned to the right species, genus or family; ambiguous identification means that the individual is assigned to one or several species, genera or families including the right one; incorrect identification means that the individual is assigned to one or several species, genera or families not including the right one [43]. TAXONDNA [44] was used to assess the distribution of interspecific and intraspecific distances in the dataset. Barcode gap analysis of *matK* and *rbcl* was performed using Kimura 2-parameter distance model implemented in Boldsystems [45]. Consensus barcode of each species was obtained using the 'Consensus Barcode Generator' function of TAXONDNA [44]. Consensus barcodes were used in a

neighbor-joining (NJ) trees of *matK*, *rbcL* and the combined *rbcL* + *matK* sequences using evolutionary distances computed based on the Kimura 2-parameter [46] method in MEGA6 [47]. Sequences were trimmed, and aligned using MUSCLE [48] by pairwise deletion and 500 replications of Bootstrap phylogeny test. Distance analyses were performed in MEGA6 between families, within families and among species using consensus barcode sequences. The number of segregation sites and nucleotide diversity value which is the average number of nucleotide differences per site between a pair of randomly chosen sequences [49] was calculated for *matK* and *rbcL* using DnaSP v5 [50]. All alignments are available as [S1–S3](#) Alignments.

Results

Sequencing success

PCR amplifications of 225 plant specimens yielded 217 (96.4%) *rbcL* and 189 (84%) *matK* sequences. Our collection represented 161 plant species, 98.1% of them were successfully sequenced for *rbcL* and 83.9% for *matK*. Sequence length distribution ranged between 506–552bp and 468–894 bp for *rbcL* and *matK*, respectively. The longest *matK* sequences (894 bp) were produced for *Ipomoea*, *Mentha* and *Syngonium* while the shortest (468 bp) in *Mattiola incana* (L.) R.Br. For *rbcL* most species produced similar length (552 bp) except for few short fragments in *Bauhinia retusa* (520 bp), *Papaver rhoeas* L. (531 bp), *Spiraea cantoniensis* (529 bp) and in *Rosa hybrida* L. (215 bp). The GC% ranged from 27.98 to 83.34 with an average of 33.64 in *matK* whereas in *rbcL* it ranged from 40.29 to 43.30 with an average of 36.38. Mean number of specimens examined per species was 1.44 and 1.45 for *rbcL* and *matK*; respectively. Sequencing success varied between families ([S2 Table](#)). The lowest success rates were found in *matK* in several members of Crassulaceae (12.5%), Malvaceae (57.14%) and Brassicaceae (66.7%). Furthermore, some singleton families (represented by one member) were not amplified or sequenced in *matK* such as Balsaminaceae, Oxalidaceae and others. The *rbcL* showed 100% amplification and sequencing success with most families except for few members of the family Linaceae, Piperaceae, Araceae, Lamiaceae, and Asteraceae. Medicinal and horticultural families such as Lamiaceae showed high sequences recoveries (100 *matK*, 91.3% *rbcL*).

Species resolution and barcode analyses

Using *matK* sequences, taxa were correctly assigned by 100, 85.2 and 39.7% at the family, genus and species levels, respectively, whereas ambiguous identification was 6.9 and 36.5% for genus and species levels. Incorrect *matK* identification represented 7.9 and 23.8% for the genus and species, respectively ([Table 1](#)). The *rbcL* successfully identified 100, 74.65 and 29% at the levels of family, genus and species, respectively, whereas ambiguous identification was 13.8 and 38.2% for genus and species levels. Incorrect *rbcL* identification represented 11.5 and 32.7% for genus and species, respectively. Concatenations of *matK* & *rbcL* sequences correctly assign 83.4% taxa to their genus and 39.8% to species while it assigned 11.6% of genera and 46.9% of species ambiguously. By concatenating *rbcL* and *matK*, the incorrect assignments were only 4.9% for the genera and 13.3% for species.

In TaxonDNA, pairwise intraspecific distances in the two barcode loci of all dataset ranged from 0.0–2.7% ([Table 2](#)). The *rbcL*+*matK* showed higher mean intraspecific value than either marker. Pairwise mean interspecific distances were low (0.4%) in *rbcL* and high (1.3%) in *matK*. The concatenation of barcode loci did not increase the interspecific mean distances ([Table 2](#)). The data showed overlapping between intraspecific and interspecific distances of the individual or combined sequences ([Table 2](#)). This overlapping did not differ between *rbcL* (89.3%) and *matK* (89.4%) while being increased in *rbcL*+*matK* (97.2%). The barcode gap analysis provides the distribution of distances within each species and the distance to the nearest

Table 1. Identification success of Egyptian horticultural crops based on Blast1 method.

		Correct (%)	Ambiguously identified (%)	Incorrect identification (%)	Total sample Number	Average sequence Length bp
Species	<i>rbcl</i>	63 (29.03)	83(38.24)	71 (32.72)	217	522
	<i>matK</i>	75 (39.68)	69(36.50)	45 (23.80)	189	870
	<i>rbclLa +matK</i>	72 (39.77)	85(46.96)	24(13.25)	181	1323
Genus	<i>rbcl</i>	162 (74.65)	30 (13.82)	25 (11.52)	217	522
	<i>matK</i>	161 (85.18)	13(6.87)	15 (7.90)	189	870
	<i>rbclLa +matK</i>	151(83.42)	21(11.60)	9(4.90)	181	1323
Family	<i>rbcl</i>	217 (100)	-	-	217	522
	<i>matK</i>	189 (100)	-	-	189	870
	<i>rbclLa +matK</i>	181(100)	-	-	181	1323

doi:10.1371/journal.pone.0172170.t001

neighbor (NN) of each species. The use of barcode gap analysis tool on BOLD for *matK* under K2P distance model (pairwise deletion) showed a higher mean NN distances (4.7) than the mean intraspecific (0.01) indicating the existence of a barcode gap. Based on 189 *matK* sequences 22 species showed a higher (>2%) and 52 showed a lower (<2%) intraspecific divergence. The *rbcl* showed a higher mean (2.3) NN than the mean intraspecific (0.0) distance. The analysis of 217 *rbcl* sequences showed 23 species with higher (>2%) and 91 with lower (<2%) intraspecific distances.

Sequence analysis

Families and genera clustering. The NJ tree for *rbcl+matK* was generated using 182 sequences (S1 Fig) with at least one sequence from each family. Members of each family are clustered on the tree with the largest cluster for the family Lamiaceae in the *matK* (S1 Fig) or *rbcl + matK* trees. Furthermore, each genus was split into sub-clusters. In Solanaceae, 13 individuals from 6 genera were clustered. Barcodes separated all the genera but did not separate a majority of the species. Members of Asparagaceae were analyzed by both markers and formed two subclusters, one joined the genera *Yucca* and *Chlorophytum* and the second joined *Draacaena*, *Sansveiria*, *Asparagus* and *Cordyline* (S1 Fig). In Arecaceae, nine species were examined. Species of Arecaceae were differentiated in *matK* and *rbcl+matK* but formed one cluster. The taxa *Spathiphyllum*, *Monstera*, *Anthurium*, *Aglonema* were discriminated by both loci except for three species belonging to *Phyllodendron*.

Simple diagnostic characters for genera and species. *Mentha* showed simple diagnostic characters as two polymorphic sites in the local species split the genus into three different haplotypes (Fig 1). The first contained 459-T&670-G, exclusively found in *M. longifolia* L. whereas the second and third haplotypes (459-C &670-A /459-C &670-G) were shared in *M. piperita*

Table 2. Pairwise intraspecific and interspecific distances in the barcode loci of all 161 plant species.

Locus (n)	Intraspecific distances (%)			Interspecific distances (%)			Intra-/interspecific distance overlap with 5% error margin on both sides	
	Min	Max	Mean	Min	Max	Mean	Overlapping distance range (%)	Intra-/interspecific sequences in the overlap (%)
<i>matK</i> (189)	0.00	0.15	0.02	0.00	11.35	1.28	0.0–0.24	89.41
<i>rbcl</i>(217)	0.00	0.18	0.01	0.00	3.08	0.43	0.0 to 0.18	89.33
<i>matK+rbcl</i> (182)	0.00	2.71	0.24	0.00	8.31	1.17	0.0 to 3.15	97.17

doi:10.1371/journal.pone.0172170.t002

such as leaves variegation in the former. In *Salvia* (Lamiaceae), two haplotypes were found in each locus; one of them is associated with *S. viridis* L. and the second in two *S. splendens* Sellow ex Schult. Two species in Lamiaceae [*Rosmarinus officinallis* L., *Solenostemon scutellarioides* (different cultivars)] did not show diagnostic characters although the former shows clear morphological differences among subspecies examined. In *Petunia* (Solanaceae), 3 species were examined (*P. x hybrida*, *P. axillaris* and *P. integrifolia*), two haplotypes in each of *matK* and *rbcL* were found. Each of *Petunia axillaris* (Lam.) Britton, Sterns & Poggenb. and *P. integrifolia* subsp. *inflata* had its own haplotype in each barcode marker whereas *P. x hybrida* cultivars contained both haplotypes of each barcode and each barcode marker divided *P. x hybrida* cultivars into two groups based on single nucleotide polymorphism. In *Dracaena* (Asparagaceae) four morphologically divergent species were barcoded and each barcode differentiated each species accurately where four haplotypes were produced in each locus. Furthermore, other species of Asparagaceae such as *Yucca gloriosa* variegata and *Y. aloifolia* purpurea produced two haplotypes in both loci. In Arecaceae, species of the genera *Dypsis*, *Livistona*, *Ravenea* and *cocos* showed clear diagnostic characters. *Monstera*, *Spathiphyllum*, *Anthurium*, *Aglonema*, *Zamiocollas* (Araceae) had their own simple diagnostic characters in both markers. Simple diagnostic characters were found in the closely related genera of *Chrysanthemum* and *Matricaria* of the family Asteraceae.

Genetic distances among families, species and nucleotide diversity. We compared maximum, minimum and average distances for each locus and for the combined *rbcL+matK* sequence. In *matK*, mean distance among families, within families and among species were 0.22, 0.05 and 0.22; respectively (Table 3). In *rbcL* mean distance among families, within families and among species were 0.09, 0.02 and 0.09; respectively. *rbcL+matK* showed the largest distances compared to individual locus. However, minimum distances among families in individual or combined loci were higher than minimum within families or among species.

Nucleotide diversity, number of segregation sites and number of haplotypes for the two barcode loci for all genera represented by several species were calculated (Fig 2). The number of species ranged from 1 (*Hibiscus rosa-sinensis* L.) in *matK* and *rbcL* to 5 in *rbcL* such as in the genus *Kalanchoe* (*K. beharensis* Drake, *K. blossfeldiana* Poelln., *K. manginii* Raym.-Hamet & H.Perrier, *K. thysiflora* Balfour and *K. tomentosa* Golden Girl) as shown in Fig 2 and in S1 Table. The highest number of segregation sites in all genera was in *Pilea* between *P. cadierei* Gagnep. & Guillaumin and *P. serpyllacea* (Kunth) Liebm. The *Pilea* was followed by *Salvia* cultivars and *Justicia* cultivars for *matK*. The nucleotide diversity ratio followed the same trend as for the segregation sites where the highest value was in the *Pilea* followed by *Salvia* (*S. splendens* Sellow ex Schult. & *S. viridis* L.) and *Justicia* (*J. adhatoda* L. & *J. brandegeana* Wash. & L. B.Sm.) in both loci. In *matK*, the number of haplotypes ranged from 1–4 (Fig 2). Nucleotide diversity was highest in *Pilea* (12.9%) followed by *Salvia* (6.5%) and *Justicia* (4.3%) whereas the remaining genera ranged from 2–0%. In *rbcL*, the number of haplotypes was either high (5, 4 and 3) such as in *Kalanchoe*, *Dracaena* and *Narcissus* L., respectively or low (2 or 1) in all remaining genera. The highest values were found in *Pilea* (3.08), *Chrysanthemum* L. (1.45), *Narcissus* (0.97), *Justicia* (0.72) and *Dracaena* (0.36).

Table 3. Genetic distances in barcode loci at three taxonomic levels.

Barcode locus (n = 132)	Among families			within families			species		
	min	Max	Mean	Min	Max	Mean	Min	Max	Mean
<i>matK</i>	0.05	0.38	0.22	0	0.12	0.05	0	0.4	0.22
<i>rbcL</i>	0.02	0.16	0.09	0	0.06	0.02	0	0.17	0.09
<i>rbcL+matK</i>	0.04	0.14	0.24	0	0.08	0.034	0	0.24	0.14

doi:10.1371/journal.pone.0172170.t003

	Locus	Seq Len (bp)	No. of Sp.	No. of Hap	Nuc Div (Pi)%	Segre Sites
<i>Mentha</i>	<i>rbcL</i>	552	4	1	0	0
	<i>matK</i>	720	4	3	0.12	2
<i>Ruellia</i>	<i>rbcL</i>	552	2	2	0.36	2
	<i>matK</i>	796	2	2	2.00	16
<i>Justicia</i>	<i>rbcL</i>	552	2	2	0.72	4
	<i>matK</i>	789	2	2	4.30	34
<i>Petunia</i>	<i>rbcL</i>	552	3	2	0.10	1
	<i>matK</i>	685	3	3	0.17	2
<i>Salvia</i>	<i>rbcL</i>	552	1	1	0	0
	<i>matK</i>	704	2	2	6.45	45
<i>Chrysanthemum</i>	<i>rbcL</i>	552	2	2	1.45	8
	<i>matK</i>	791	2	2	1.90	15
<i>Pericallis</i>	<i>rbcL</i>	552	2	2	0.22	3
	<i>matK</i>	743	2	2	0.36	8
<i>Hibiscus</i>	<i>rbcL</i>	552	1	2	0.09	1
	<i>matK</i>	751	1	1	0	0
<i>Ficus</i>	<i>rbcL</i>	552	2	2	0.18	1
	<i>matK</i>	787	2	2	0.89	7
<i>Pilea</i>	<i>rbcL</i>	552	2	2	3.08	17
	<i>matK</i>	770	2	2	12.86	99
<i>Kalanchoe</i>	<i>rbcL</i>	552	5	5	0.65	8
	<i>matK</i>	----	----	----	----	----
<i>Narcissus</i>	<i>rbcL</i>	552	3	3	0.97	8
	<i>matK</i>	811	2	2	0.37	3
<i>Dracaena</i>	<i>rbcL</i>	552	4	4	0.36	4
	<i>matK</i>	727	4	4	0.28	4
<i>Yucca</i>	<i>rbcL</i>	552	2	2	0.09	1
	<i>matK</i>	501	2	2	0.10	1
<i>Philodendron</i>	<i>rbcL</i>	552	2	1	0	0
	<i>matK</i>	530	2	2	0.13	1
<i>Plectranthus</i>	<i>rbcL</i>	552	2	1	0	0
	<i>matK</i>	768	2	1	0	0
<i>Euphorbia</i>	<i>rbcL</i>	552	2	1	0	0
	<i>matK</i>	726	2	2	0.69	5

Fig 2. Comparison of two loci tested on several genera. No.: Number, Sp.: Species, Seq.: Sequence, Len.: Length, Hap.: Haplotype, Nuc. Div.: Diversity, Segre.: Segregation.

doi:10.1371/journal.pone.0172170.g002

Discussion

DNA barcoding campaign is still at its infancy stage in Africa particularly in northern Africa, although an increasing effort is noted in South Africa [4,15,16]. Our study, a first attempt of

DNA barcoding study of its kind in Egypt and northern Africa, showed a higher sequencing success for *rbcL* than *matK*. Previous studies have shown a similar pattern in other plant groups [24,13]. Our sequencing success of *matK*, however, matched that reported in CBOL [12] but was higher than that reported by Parmentire et al. [29]. The sequences recovery in the family Lamiaceae was higher in *matK* than *rbcL* disagreeing with a study by Theodoridis et al. [51] on the same plant family. We found a higher universality in *rbcL* in genera identification but a lower species resolution than what was observed in *matK* [12]. In addition, we found a barcode gap [52] in *matK* with a higher mean interspecific than the mean intraspecific distances in 189 sequences. In general, the barcode gaps observed in this study is higher than that found in an early study of trees and shrubs in Egypt [13]. Although the existence of barcode gap may not predict the discrimination success [53], it is a key criterion for barcoding assessment. Genetic distance analyses were conducted at different taxonomic levels. Distances with *matK* were 2 times greater than the mean distance of *rbcL* in all cases, indicating a higher resolution power of *matK* for the poorly studied flora of Egypt. Furthermore, concatenated sequences of *rbcL* and *matK* slightly increased distances reflecting improved resolution power using both barcodes which is in agreement with Parmentier et al. [29] and Saarela et al. [54]. Both barcodes indicated that the largest genetic distance was achieved within Rubiaceae between *Pentas lanceolata* (Forssk.) Deflers and *Hoffmania discolor* (Lem.) Hemsl. The family Rubiaceae contains over 13,200 species in 620 genera in addition to numerous unresolved generic complexes [55] and the family harbors a high diversity, especially in southern African countries and South America and has a worldwide distribution. The high number of segregation sites and consequently high nucleotide diversity found in *Pilea* (Urticaceae) species compared to other genera is due to the species richness of this genus as it contains over 700 species and one fifth of the diversity of seed plants [56,57].

As expected, taxonomic assignment decreases from family, genus to specific level [29,13]. A combination of *rbcL+matK* slightly improved the rate of correct species resolution over the individual markers. The combined markers did not improve genus identification, supporting previous report for African flora [29]. The combined markers, however, dramatically reduced the level of incorrect species identification by 60% in *rbcL* and 44% in *matK*. A similar trend was found at the genus level. Correct cultivar assignments were 1.4 and 1.3% for *matK* and *rbcL*, respectively. The lower species discrimination in our study could be attributed to several factors such as floristic affinities (e.g. close relatives are well known for not being easily discriminated by the official barcodes [58]), or the existence of multiple cultivars in our horticultural crops. Further, it is also possible, owing to the Egyptian flora being understudied, that there is a taxonomic confusion (vague morphological parameters leading to misidentifications) in the existing morphology-based species discrimination.

We compared our *matK* sequences from mint species with those in GenBank and constructed a phylogenetic tree (data not shown). The identification rate was low; one potential reason could be possible hybridization, introgression or gene flow between species [59] blurring both genetic and taxonomic delimitations between taxa [60]. It could also be because the GenBank data are questionable as such doubt about public repositories has previously been reported (e.g. [61]). It is also likely that the well-known maternal inheritance associated with plastid regions [62] plays a role in the poor discriminatory power of *rbcL* and *matK*. Furthermore, although we expect nuclear region that could perform better than *rbcL + matK*, several recent studies also reported low performance (i.e. $\leq 50\%$) for *ITS*. For example, the highest performance of *ITS* for Orchids was around 50% [63] and about 30% for Alooidea [28]. Our objective in this study was to build a DNA barcode library for Egyptian flora and demonstrate how DNA barcodes data can be used for biodiversity assessment, and ecological studies of local flora in future studies (e.g. see ref. [16,39] for South Africa's flora).

Although identification rate is known to decrease with an increase in the mean number of species per genus [29], this could not be the case in this study as the mean number of species per genus is lower (1.3) than reported in other studies [64]. Low identification rate in both core barcodes is common as reported in several taxonomic groups: Indian *Berberis* (23%; [65]), Pinaceae (25%; [66]) and vascular plants of Manitoba, Canada (45–55%; [67]) and African Combretaceae (10–61%; [15]). Dong et al. [68] explored the use of *rbcL* as barcode in all plant families and found that the successful species identification rates varied significantly among plant groups, ranging from 24.58% to 85.50%.

Furthermore, NJ-tree analysis shows that both Asparagaceae and Amarillidaceae are sisters in the tree (S1 Fig) which is in agreement with the Angiosperm phylogeny Group (APGIII) tree (<http://www.mobot.org/MOBOT/Research/APweb/welcome.html>). *matK* highly discriminated species of Arecaceae, suggesting that *matK* is a strong DNA barcode candidate for the Egyptian palms. In addition, 13 species demonstrated simple diagnostic characters whereas other species had homologous sequences using both core barcodes. Sequence variation in some cases was associated with morphological variations and in other cases sequences were identical. Our study therefore recommends the use of several combined markers beyond *rbcL* and *matK*. The two species of *Slavia* (*S. splendens* Sellow ex Schult. & *S. viridis* L.) examined showed simple diagnostic characters in both markers, matching the morphological difference between both species based on the flower color (red in *S. splendens* and blue in *S. viridis*). Barcodes discriminated between the two closely genera of *Chrysanthemum* (*C. carinatum* Schousb., *C. morifolium* Ramat.) and *Matricaria* (*M. chamomilla* L.) of the family Asteraceae. Morphologically divergent varieties or hardly known varieties were chosen from these genera to be barcoded in this study. In some cases, we chose varieties showing variation in flowers color such as in *Viola tricolor* (Hornveilchen lila, Frosthart, Hortensis, L., Heartsease, Hornveilchen hellgelb, Simon Shine, Sun Glory, Freefall Purple & White), *Pericallis x hybrida* (Senetti Blue Bicolor, Senetti Magenta, Senetti Super Blue, Senetti Pink, Jester Pure White) and *Antirrhinum majus* L. (pink and white). In other cases, we studied varieties showing variation in leaf-shape and variegation in leaf color such as in *Brassica olearaceae* (Emperor white, L., Dietrich Idaho and Nagoya Red F1), *Hydrangea macrophylla* (Thunb. Ser. and L.) and *Codiaeum variegatum* (L.) Rumph. ex A. Juss. Only four of all species examined either for *matK* or in *rbcL* showed variation among varieties in the ratio of 1.4% and 1.4%; respectively. The inability to distinguish among subspecies/varieties using the core barcodes is well established [29] although few cases where barcodes or plastid regions were successful in discriminating among subspecies as found in *Mentha spicata* L. and *M. x piperita* (Chocolate and L.) and the case of the intergenic spacer *trnH-psbA* (a complementary DNA barcode) in *Silene vulgaris* (Moench) Garcke [69] as well as *matK* and *rbcL* in *Celtis occidentalis* L. [13] were also reported.

Conclusion

The application of DNA barcoding in horticultural and agricultural industry is promising. Both the core barcodes have a high resolution power at genus level and moderate at the species level with *matK* showing higher resolution power at all taxonomic levels. The addition of other barcodes may enhance the discriminatory power of barcoding at genus and species levels. The core DNA barcodes are not always able to discriminate species but have more promise in controlling the market place of horticultural crops and protecting copyrights of new species or cultivars. Nuclear markers are generally advocated for, and the ITS region in particular, although we should acknowledge some controversies around this nuclear marker (see [70]): incomplete lineage sorting, inhomogeneous concerted evolution, divergent paralogous copies within individuals, and pseudogenes; [71]; but see ref. [26]). Overall, we suggest that including more

replicates per species and adopt a more multi-gene approach that includes a nuclear region may result in a more efficient DNA barcode data for horticultural and agricultural industry.

Supporting information

S1 Table. Taxa included in the study.

(DOCX)

S2 Table. Families using each of *matK* & *rbcLa* of BOLD Taxon ID tree analysis showing that there are 189 sequences, 117 species, 114 genus and 50 family using *matK* and 217 sequences, 131 species, 132 genus and 62 family using *rbcLa*. % of PCR success based on families is also illustrated.

(DOCX)

S1 Alignment. *rbcL* alignment.

(FAS)

S2 Alignment. *matK* alignment.

(FAS)

S3 Alignment. *rbcL+matK* alignment.

(FAS)

S1 Fig. NJ tree of *rbcL* and *matK* produced in MEGA6.

(DOCX)

S2 Fig. NJ tree of taxa using *matK*, produced in MEGA 6.

(DOCX)

S3 Fig. NJ tree of *rbcL* produced in MEGA6.

(DOCX)

S1 Sheet. CCDB DNA extraction for plants.

(PDF)

S2 Sheet. CCDB PCR amplification for plants.

(PDF)

Acknowledgments

This project was supported by the King Saud University, Deanship of Scientific Research, College of Science Research Center. This publication arises from a collaborative research program between the Faculty of Agriculture, Alexandria University and the Biodiversity Institute of Ontario (BIO). Sequence analysis was funded by the Government of Canada through Genome Canada and the Ontario Genomics Institution in support of the International Barcode of Life Project.

Author Contributions

Conceptualization: HOE.

Data curation: HOE.

Formal analysis: HOE.

Funding acquisition: HOE.

Investigation: HOE.

Methodology: HOE.

Project administration: HOE.

Resources: HOE.

Software: HOE.

Supervision: HOE.

Validation: MA.

Visualization: HOE.

Writing – original draft: HOE MA.

Writing – review & editing: HOE MA HMA KY.

References

1. Pardee WD: Protecting New Plant Varieties through PVP: Practical Suggestions from a Plant Breeder for Plant Breeders. In: (Krattiger A, Mahoney RT, Nelsen L, editors. Intellectual Property Management in Health and Agricultural Innovation: A Handbook of Best Practices MIHR: Oxford, U.K., and PIPRA: Davis, U.S.A; 2007. Available online at www.ipHandbook.org. (Accessed Sep. 14, 2016)
2. CIOPORA Green Paper. CIOPORA green paper on plant variety protection: Policy statement. René Royon and Ciopora, November 2002, <http://www.ciopora.org/publications> (accessed Sep. 14, 2016)
3. CIOPORA Strategy Paper. on the Negative Effects of Infringements of Plant Variety Rights. 2007. Available from: <http://www.ciopora.org/publications> (accessed Oct. 14, 2016).
4. Hoveka LN, van der Bank M, Boatwright JS, Bezeng BS, Yessoufou K. The noncoding *trnH-psbA* spacer, as an effective DNA barcode for aquatic freshwater plants, reveals prohibited invasive species in aquarium trade in South Africa. *S. Afr. J. Bot.* 2016; 102: 208–216.
5. Ali MA, Gyulai G, Hidvégi N, Kerti B, Al Hemaïd FMA, Pandey AK, Lee J. The changing epitome of species identification—DNA barcoding. *Saudi J. Biolog. Sci.* 2014; 21:204–231
6. Keil M, Griffin AR. Use of random amplified polymorphic DNA (RAPD) markers in the discrimination and verification of genotypes in *Eucalyptus*. *Theor. Appl. Genet.* 1994; 89: 442–450. doi: [10.1007/BF00225379](https://doi.org/10.1007/BF00225379) PMID: [24177893](https://pubmed.ncbi.nlm.nih.gov/24177893/)
7. McKinnon GE, Vaillancourt RE, Steane DA, Potts BM. An AFLP marker approach to lower-level systematics in *Eucalyptus* (Myrtaceae). *Am J Bot.* 2008; 95: 368–380. doi: [10.3732/ajb.95.3.368](https://doi.org/10.3732/ajb.95.3.368) PMID: [21632361](https://pubmed.ncbi.nlm.nih.gov/21632361/)
8. Besnard G, Khadari B, Villemur P, Bervillé A. Cytoplasmic male sterility in the olive (*Olea europaea* L.). *Theor. Appl. Genet.* 2000; 100: 1018–1024.
9. Ochieng JW, Steane DA, Ladiges PY, Baverstock PR, Henry RJ, Shepherd M. Microsatellites retain phylogenetic signals across genera in *Eucalypts* (Myrtaceae). *Genet Mol Biol.* 2007; 30: 1125–1134.
10. Ganai MW, Polley A, Graner EM, Plieske J, Wieseke R, Luerksen H, Durstewitz G. Large SNP arrays for genotyping in crop plants. *Bioscience.* 2012; 37: 821–828.
11. Hebert PDN, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. *Proc R Soc Lond [Biol]*. 2003; 270: 313–321.
12. CBOL, Plant Working Group. A DNA barcode for land plants. *Proc. Natl. Acad. Sci. U.S.A.* 2009; 106: 12794–12797. doi: [10.1073/pnas.0905845106](https://doi.org/10.1073/pnas.0905845106) PMID: [19666622](https://pubmed.ncbi.nlm.nih.gov/19666622/)
13. Elansary HO. Towards a DNA barcode library for Egyptian flora, with a preliminary focus on ornamental trees and shrubs of two major gardens. *DNA Barcodes.* 2013; 1: 46–55.
14. Gere J, Yessoufou K, Maurin O, Mankga LT, Daru BH, Van der Bank M. Incorporating *trnH-psbA* to core DNA barcodes improves discrimination of species within southern African Combretaceae. *Zookeys* 2013; 365: 127–147.
15. Yessoufou K, Davies JT, Maurin O, Kuzmina M, Schaefer H, Van der Bank M, Savolainen V. Large herbivores favour species diversity but have mixed impacts on phylogenetic community structure in an African savanna ecosystem. *Ecology.* 2013; 101: 614–625.
16. Ashfaq M, Asif M, Anjum ZI, Zafar Y. Evaluating the capacity of plant DNA barcodes to discriminate species of cotton (*Gossypium*: Malvaceae). *Mol Ecol Resour.* 2013; 13: 574–582.

17. Maurin O, Davies TJ, Burrows JE, Daru BH, Yessoufou K, Muasya MA, Van der Bank M, Bond W. Savanna fire and the origins of “underground forests” of Africa. *New Phytologist*. 2014; 204: 201–214. doi: [10.1111/nph.12936](https://doi.org/10.1111/nph.12936) PMID: [25039765](https://pubmed.ncbi.nlm.nih.gov/25039765/)
18. Casiraghi M, Labra M, Ferri E, Galimberti A, De Mattia F. DNA barcoding: a six-question tour to improve users’ awareness about the method. *Brief. Bioinformatics*. 2010; 11: 440–453. doi: [10.1093/bib/bbq003](https://doi.org/10.1093/bib/bbq003) PMID: [20156987](https://pubmed.ncbi.nlm.nih.gov/20156987/)
19. Janzen DH, Hallwachs W, Blandin P, Burns JM, Cadiou JM, Chacon I, Dapkey T, et al. Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Mol. Ecol. Resour.* 2009; 9: 1–26.
20. Williamson J, Maurin O, Shiba S, Van der Bank H, Pfab M, Pilusa M, Kabongo R, Van der Bank M. Exposing the illegal trade in cycad species (*Cycadophyta: Encephalartos*) at two traditional medicine markets in South Africa using DNA barcoding. *Genome*. 2016.
21. Kress WJ, Wurdack KL, Zimmer EA, Weigt LA, Janzen DH. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U.S.A.* 2005; 102: 8369–8374. doi: [10.1073/pnas.0503123102](https://doi.org/10.1073/pnas.0503123102) PMID: [15928076](https://pubmed.ncbi.nlm.nih.gov/15928076/)
22. Chase MW, Cowan RS, Hollingsworth PM, van den Berg C, Madriñán S, Petersen G, et al. A proposal for a standardized protocol to barcode all land plants. *Taxon*. 2007; 56: 295–299.
23. Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, et al. Power and limitations of the chloroplast *trnL*(UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* 2007; 35: 1–8.
24. Kress WJ, Erickson DL. A two-locus global DNA barcode for land plants: the coding *rbcl* gene complements the non-coding *trnH-psbA* spacer region. *PLoS ONE* 2007; 2, e508. doi: [10.1371/journal.pone.0000508](https://doi.org/10.1371/journal.pone.0000508) PMID: [17551588](https://pubmed.ncbi.nlm.nih.gov/17551588/)
25. China Plant BOL Group. Comparative analysis of a large dataset indicates that ITS should be incorporated into the core barcode for seed plants. *Proc. Natl. Acad. Sci. USA* 2011; 108: 19641–19646. doi: [10.1073/pnas.1104551108](https://doi.org/10.1073/pnas.1104551108) PMID: [22100737](https://pubmed.ncbi.nlm.nih.gov/22100737/)
26. Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S. Plant DNA barcoding: from gene to genome *Biol. Rev.* 2015; 90, 157–166.
27. Daru BH, van der Bank M, Bello A, Yessoufou K. Testing the reliability of the standard and complementary DNA barcodes for the monocot subfamily Alooideae from South Africa. *Genome* 2016;
28. Parmentier I, Duminil J, Kuzmina M, Philippe M, Thomas DW, Kenfack D, et al. How effective are DNA barcodes in the identification of African rainforest trees? *PLoS ONE* 2013; 8, e54921. doi: [10.1371/journal.pone.0054921](https://doi.org/10.1371/journal.pone.0054921) PMID: [23565134](https://pubmed.ncbi.nlm.nih.gov/23565134/)
29. Kress WJ, Erickson DL, Jones FA, Swenso NG, Perez R, Sanjur O, Bermingham E. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proc. Natl. Acad. Sci. U.S.A.* 2009; 106: 18621–18626. doi: [10.1073/pnas.0909820106](https://doi.org/10.1073/pnas.0909820106) PMID: [19841276](https://pubmed.ncbi.nlm.nih.gov/19841276/)
30. Costion C, Ford A, Cross H, Crayn D, Harrington M, Lowe A. Plant DNA barcodes can accurately estimate species richness in poorly known floras. *PLoS ONE* 2011; 6, e26841. doi: [10.1371/journal.pone.0026841](https://doi.org/10.1371/journal.pone.0026841) PMID: [22096501](https://pubmed.ncbi.nlm.nih.gov/22096501/)
31. Christina VLP, Annamalai A. Nucleotide based validation of *Ocimum* species by evaluating three candidate barcodes of the chloroplast region. *Mol. Ecol. Resour.* 2014; 14: 60–68. doi: [10.1111/1755-0998.12167](https://doi.org/10.1111/1755-0998.12167) PMID: [24164957](https://pubmed.ncbi.nlm.nih.gov/24164957/)
32. Li HQ, Chen JY, Wang S, Xiong SZ. Evaluation of six candidate DNA barcoding loci in *Ficus* (Moraceae) of China. *Mol. Ecol. Resour.* 2012; 12: 783–790. doi: [10.1111/j.1755-0998.2012.03147.x](https://doi.org/10.1111/j.1755-0998.2012.03147.x) PMID: [22537273](https://pubmed.ncbi.nlm.nih.gov/22537273/)
33. Liu Y, Zhang L, Liu Z, Luo K, Chen S, Chen K. Species identification of *Rhododendron* (Ericaceae) using the chloroplast deoxyribonucleic acid *psbA-trnH* genetic marker. *Pharmacogn. Mag.* 2012; 8: 29–36. doi: [10.4103/0973-1296.93311](https://doi.org/10.4103/0973-1296.93311) PMID: [22438660](https://pubmed.ncbi.nlm.nih.gov/22438660/)
34. Liu Z, Zeng X, Yang D, Chu G, Yuan Z, Chen S. Applying DNA barcodes for identification of plant species in the family Araliaceae. *Gene*. 2012; 499: 76–80. doi: [10.1016/j.gene.2012.02.016](https://doi.org/10.1016/j.gene.2012.02.016) PMID: [22406497](https://pubmed.ncbi.nlm.nih.gov/22406497/)
35. Chen X, Liao B, Song J, Pang X, Han J, Chen S. A fast SNP identification and analysis of intraspecific variation in the medicinal *Panax* species based on DNA barcoding. *Gene*. 2013; 530: 39–43. doi: [10.1016/j.gene.2013.07.097](https://doi.org/10.1016/j.gene.2013.07.097) PMID: [23933277](https://pubmed.ncbi.nlm.nih.gov/23933277/)
36. Rikken M. The global competitiveness of the Kenyan flower industry. Fifth Video Conference on the Global Competitiveness of the Flower Industry in Eastern Africa, December, 2011. Available from: <http://www.kenyaflowercouncil.org>.
37. Elansary HO, Mahmoud EA, Shokralla S, Yessoufou K. Diversity of plants, traditional knowledge and practices in local cosmetics: A case study from Alexandria, Egypt. *Econ. Bot.* 2015; 69: 114–126.

38. Bezeng BS, Savolainen V, Yessoufou K, Papadopulos AST, Maurin O, Van der Bank M. A phylogenetic approach towards understanding the drivers of plant invasiveness on Robben Island, South Africa. *Bot. J. Linn. Soc.* 2013; 172: 142–152.
39. Daru BH, van der Bank M, Maurin O, Yessoufou K, Schaefer H, Slingsby JA, Davies TJ. A novel phylogenetic regionalization of the phytogeographic zones of southern Africa reveals their hidden evolutionary affinities. *J. Biogeogr.* 2016; 43: 155–166.
40. Levin RA, Wagner WL, Hoch PC, Nepokroeff M, Pires JC, Zimmer EA, Sytsma KA. Family-level relationships of Onagraceae based on chloroplast *rbcl* and *ndhF* data. *Am. J. Bot.* 2003; 90: 107–115. doi: [10.3732/ajb.90.1.107](https://doi.org/10.3732/ajb.90.1.107) PMID: [21659085](https://pubmed.ncbi.nlm.nih.gov/21659085/)
41. Ross HA, Murugan S, Li WLS. Testing the reliability of genetic methods of species identification via simulation. *Syst. Biol.* 2008; 57: 216–230. doi: [10.1080/10635150802032990](https://doi.org/10.1080/10635150802032990) PMID: [18398767](https://pubmed.ncbi.nlm.nih.gov/18398767/)
42. Gao T, Yao H, Song J, Liu C, Zhu Y, Ma X, et al. Identification of medicinal plants in the family Fabaceae using a potential DNA barcode *ITS2*. *Ethnopharmacol.* 2011; 130: 116–121.
43. Meier R, Shiyang K, Vaidya G, Ng PKL. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Syst. Biol.* 2006; 55: 715–728. doi: [10.1080/10635150600969864](https://doi.org/10.1080/10635150600969864) PMID: [17060194](https://pubmed.ncbi.nlm.nih.gov/17060194/)
44. Ratnasingham S, Hebert PDN. BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Mol. Ecol. Notes.* 2007; 7: 355–364. doi: [10.1111/j.1471-8286.2007.01678.x](https://doi.org/10.1111/j.1471-8286.2007.01678.x) PMID: [18784790](https://pubmed.ncbi.nlm.nih.gov/18784790/)
45. Kimura M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Mol. Evol.* 1980; 16: 111–120.
46. Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis. Version 6.0. *Mol. Biol. Evol.* 2013; 30: 2725–2729. doi: [10.1093/molbev/mst197](https://doi.org/10.1093/molbev/mst197) PMID: [24132122](https://pubmed.ncbi.nlm.nih.gov/24132122/)
47. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004; 32: 1792–1797. doi: [10.1093/nar/gkh340](https://doi.org/10.1093/nar/gkh340) PMID: [15034147](https://pubmed.ncbi.nlm.nih.gov/15034147/)
48. Nei M. *Molecular Evolutionary Genetics*. Columbia University Press, New York; 1987.
49. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 2003; 19: 2496–2497. PMID: [14668244](https://pubmed.ncbi.nlm.nih.gov/14668244/)
50. Theodoridis S, Stefanaki A, Tezcan M, Aki C, Kokkini S, Vlachonassios KE. DNA barcoding in native plants of the Labiatae (Lamiaceae) family from Chios Island (Greece) and the adjacent Cesme-Karaburun Peninsula (Turkey). *Mol. Ecol. Resour.* 2012; 12: 620–633. doi: [10.1111/j.1755-0998.2012.03129.x](https://doi.org/10.1111/j.1755-0998.2012.03129.x) PMID: [22394710](https://pubmed.ncbi.nlm.nih.gov/22394710/)
51. Meyer CP, Paulay G. DNA barcoding: error rates based on comprehensive sampling. *PLoS Biol.* 2005; 3: 2229–2238.
52. Little DP, Knopf P, Schulz C. DNA barcode identification of Podocarpaceae—the second largest conifer family. *PLoS ONE* 2013; 8, e81008. doi: [10.1371/journal.pone.0081008](https://doi.org/10.1371/journal.pone.0081008) PMID: [24312258](https://pubmed.ncbi.nlm.nih.gov/24312258/)
53. Saarela JM, Sokoloff PC, Gillespie LJ, Consaul LL, Bull RD. DNA Barcoding the Canadian Arctic Flora: Core Plastid Barcodes (*rbcl* + *matK*) for 490 Vascular Plant Species. *PLoS ONE* 2013; 8, e77982. doi: [10.1371/journal.pone.0077982](https://doi.org/10.1371/journal.pone.0077982) PMID: [24348895](https://pubmed.ncbi.nlm.nih.gov/24348895/)
54. Davis AP, Govaerts R, Bridson DM, Ruhsam M, Moat J, Brummitt NA. A global assessment of distribution, diversity, endemism, and taxonomic effort in the Rubiaceae. *Ann. Mo. Bot. Gard.* 2009; 96: 68–78.
55. Frodin DG. History and concepts of big plant genera. *Taxon* 2004; 53: 753–776.
56. Monro AK. Revision of species-rich genera: a phylogenetic framework for the strategic revision of *Pilea* (Urticaceae) based on cpDNA, nrDNA, and morphology. *Am. J. Bot.* 2006; 93: 426–441. doi: [10.3732/ajb.93.3.426](https://doi.org/10.3732/ajb.93.3.426) PMID: [21646202](https://pubmed.ncbi.nlm.nih.gov/21646202/)
57. Clement WL, Donoghue MJ. Barcoding success as a function of phylogenetic relatedness in Viburnum, a clade of woody angiosperms. *BMC Evol. Biol.* 2012; 12: 73. doi: [10.1186/1471-2148-12-73](https://doi.org/10.1186/1471-2148-12-73) PMID: [22646220](https://pubmed.ncbi.nlm.nih.gov/22646220/)
58. Rheindt FE, Edwards SV. Genetic introgression: an integral but neglected component of speciation in birds. *Auk* 2011; 128: 620–632.
59. Stoeckle MY, Thaler DS. DNA Barcoding Works in Practice but Not in (Neutral) Theory. *PLoS ONE* 2014; 9: e100755. doi: [10.1371/journal.pone.0100755](https://doi.org/10.1371/journal.pone.0100755) PMID: [24988408](https://pubmed.ncbi.nlm.nih.gov/24988408/)
60. Nilsson R.H., Ryberg M., Kristiansson E., Abarenkov K., Larsson K.H., Kõljalg U. Taxonomic reliability of DNA sequences in public sequence databases: A fungal perspective. *PLoS ONE* 2006; 1:e59. doi: [10.1371/journal.pone.0000059](https://doi.org/10.1371/journal.pone.0000059) PMID: [17183689](https://pubmed.ncbi.nlm.nih.gov/17183689/)
61. Griffiths AJF, Miller JH, Suzuki DT, Lewontin RC, Gelbart WM. *An Introduction to Genetic Analysis*. 7th edition. New York: W. H. Freeman. Inheritance of organelle genes and mutations. 2006; Available from: <https://www.ncbi.nlm.nih.gov/books/NBK22059/>.

62. Guo Y-Y, Huang L-Q, Liu Z-J, Wang X-Q. Promise and Challenge of DNA Barcoding in Venus Slipper (*Paphiopedilum*). PLoS ONE 2016; 11: e0146880. doi: [10.1371/journal.pone.0146880](https://doi.org/10.1371/journal.pone.0146880) PMID: [26752741](https://pubmed.ncbi.nlm.nih.gov/26752741/)
63. Kress WJ, Erickson DL, Swenson NG, Thompson J, Uriarte M, Zimmerman JK. Advances in the use of DNA barcodes to build a community phylogeny for tropical trees in a Puerto Rican Forest Dynamics Plot. PLoS ONE 2010; 5, e15409. doi: [10.1371/journal.pone.0015409](https://doi.org/10.1371/journal.pone.0015409) PMID: [21085700](https://pubmed.ncbi.nlm.nih.gov/21085700/)
64. Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, van der Bank M, et al. A DNA barcode for land plants. Proc. Natl. Acad. Sci. U.S.A. 2009; 106: 12794–12797. doi: [10.1073/pnas.0905845106](https://doi.org/10.1073/pnas.0905845106) PMID: [19666622](https://pubmed.ncbi.nlm.nih.gov/19666622/)
65. Roy S, Tyagi A, Shukla V, Kumar A, Singh UM, Chaudhary LB, et al. Universal plant DNA barcode loci may not work in complex groups: A case study with Indian *Berberis* species. PLoS ONE 2010; 5, e13674. doi: [10.1371/journal.pone.0013674](https://doi.org/10.1371/journal.pone.0013674) PMID: [21060687](https://pubmed.ncbi.nlm.nih.gov/21060687/)
66. Ran JH, Wang PP, Zhao HJ, Wang XQ. A test of seven candidate barcode regions from the plastome in *Picea* (Pinaceae). J. Integr. Plant Biol. 2010; 52: 1109–1126. doi: [10.1111/j.1744-7909.2010.00995.x](https://doi.org/10.1111/j.1744-7909.2010.00995.x) PMID: [21106009](https://pubmed.ncbi.nlm.nih.gov/21106009/)
67. Kuzmina ML, Johnson KL, Barron HR, Hebert PDN. Identification of the vascular plants of Churchill, Manitoba, using a DNA barcode library. BMC Ecol. 2012; 12, 25. doi: [10.1186/1472-6785-12-25](https://doi.org/10.1186/1472-6785-12-25) PMID: [23190419](https://pubmed.ncbi.nlm.nih.gov/23190419/)
68. Dong W, Cheng T, Li C, Xu C, Long P, Zhou CCS. Discriminating plants using the DNA barcode *rbcLb*: an appraisal based on a large data set. Mol. Ecol. Resour. 2014; 14:336–343. doi: [10.1111/1755-0998.12185](https://doi.org/10.1111/1755-0998.12185) PMID: [24119263](https://pubmed.ncbi.nlm.nih.gov/24119263/)
69. Elansary HO, Müller K, Olson MS, Štorchová H. Transcription profiles of mitochondrial genes correlate with mitochondrial DNA haplotypes in natural population of *Silene vulgaris*. BMC Plant Biol. 2010; 10, 11. doi: [10.1186/1471-2229-10-11](https://doi.org/10.1186/1471-2229-10-11) PMID: [20070905](https://pubmed.ncbi.nlm.nih.gov/20070905/)
70. Kress WJ, Wurdack KJ, Zimmer EA, Weigt LE, Janzen DH. Use of DNA barcodes to identify flowering plants. Proc. Natl. Acad. Sci. USA 2005; 102: 8369–8374. doi: [10.1073/pnas.0503123102](https://doi.org/10.1073/pnas.0503123102) PMID: [15928076](https://pubmed.ncbi.nlm.nih.gov/15928076/)
71. Hollingsworth PM, Graham SW, Little DP. Choosing and using a plant DNA barcode. PLoS ONE 2011; 6(5): e19254. doi: [10.1371/journal.pone.0019254](https://doi.org/10.1371/journal.pone.0019254) PMID: [21637336](https://pubmed.ncbi.nlm.nih.gov/21637336/)