A test of *psbK-psbI* and *atpF-atpH* as potential plant DNA barcodes using the flora of the Kruger National Park as a model system (South Africa)

Renaud Lahaye[1,*], Vincent Savolainen[2,3], Sylvie Duthoit[1], Olivier Maurin[1] and Michelle van der Bank[1]

Author affiliation: [1]Department of Botany and Plant Biotechnology, APK Campus, University of Johannesburg, P. O. Box 524, Auckland Park 2006, Johannesburg, South Africa; [2]Royal Botanic Gardens, Kew, Richmond TW9 3DS, UK; [3]Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot SL5 7PY, UK.

We thank the South African National Research Foundation, the University of Johannesburg, SASOL, the UK Darwin Initiative, and The Royal Society (UK) for funding. We also thank the Kruger National Park, South African National Parks, H. Eckhardt, I. Smit, G. Zambatis, T. Khosa, for granting access to the park and sharing data; Stephen Boatwright for proofreading the manuscript; and T. Rikombe, R. Bryden, T. Mhlongo, H. van der Bank for fieldwork.

[*]**To whom correspondence should be addressed:** lahaye@cict.fr

Introduction

DNA barcoding is a new technique that uses short, standardized DNA sequences (400-800 bp) of an organism to determine its identity. Because this sequence has to be variable enough to identify individual species, but not too variable within the same species so that a clear threshold can be defined between intra- and inter-specific diversities, it is very challenging to apply this technique to all species on the planet [1]. A DNA barcode has been identified for animals, i.e. the mitochondrial gene *cox1* [2-5], which shows strong abilities in identifying cryptic species, accelerating biodiversity inventories and helping to identify species from degraded material (e.g. to control trade of threatened) [3, 6-9]. For plants, the identification of a suitable DNA barcode is more problematic. Cho et al. [10] showed that mitochondrial DNA evolves too slowly in plants to provide a region variable enough to discriminate between species. Then the quest for the best suitable barcode started [11] and is still ongoing [12].

Kress et al. [13] opened the debates and suggested the use of multiple genes to identify plant species quickly and accurately. At the Second International Barcode of Life Conference in Tapei (September 2007), at least five different plant DNA barcodes were proposed [12], but no consensus reached. Among those, both *atpF-atpH* and *psbK-psbI* suggested by Kim et al. [12] have not yet been tested. Here, we evaluate the use of these loci as DNA barcodes for plants by applying them to a wide range of plant species. The two new intergenic loci *atpF-atpH* and *psbK-psbL* are both localized in the large single copy (LSC) of the plastid genome. The genes *atpF* and *atpH* encode ATP synthase subunits CFO I and CFO III, respectively [14]. Both genes *psbK* and *psbI* encode two low molecular mass polypeptides, K and I, respectively, of the photo-system II [15]. These two new loci are conservative from algae to land plants and even in parasitic plants [16-18]. In this study, we focus on the trees and shrubs from the Kruger National Park (hereafter KNP), part of the Maputaland-Pondoland-Albany hotspot [19] in southern Africa. On a selected sampling from the 2,700 taxa surveyed in the area, we applied several metrics following Lahaye et al. [20] to evaluate the efficiency of combining *matK* either to *trnH-psbA* and/or *atpF-atpH* and/or *psbK-psbI* [12] for DNA barcoding purposes.

Material and Methods

Sampling. In total 101 taxa from the KNP were sampled, covering 18 families from the monocotyledons to the euasterids II. This dataset included 31 species of trees and shrubs in which we had more than one representative per species, 3 species of Orchids, one of which with 2 representatives, and 3 parasitic plants, one of which is achlorophyllous. Parasitic plants have been sampled to test the universality of the potential DNA barcodes. We used *Amborella trichopoda* Baill. (complete genome GenBank accession AJ506156) as outgroup for the phylogenetic analyses. All specimens were collected in different ecosystems when possible (Figure 1) and voucher specimens are available as detailed in Table 1.

Collection and preservation. Collection of plant material was done in the KNP with the assistance of the park's rangers. Plants were sampled and pressed for herbarium voucher specimens in triplicate, one for the herbarium of the KNP, one for Kew Herbarium (K; United Kingdom), and one for the herbarium at Pretoria (PRE; South Africa). Information about the locality and habit of collected plants were entered on a palmtop-GPS to facilitate their further treatment, and also noted on hard copy for security. For each plant collected, leaf material was stored in silica for molecular studies, and flowers and fruit stored in ethanol when available.

DNA sequencing. Total DNA was extracted from dried leaf material using the standard method of Doyle and Doyle [21] and cleaned with QIAquick silica columns (Qiagen, Helden, Germany). Sequences of *matK* and *trnH-psbA* for each taxa were published in Lahaye et al. [20] and their accession numbers are available from GenBank (Table 1). We amplified *atpF-atpH* and *psbK-psbI* using PCR as follows: 35 cycles, 30 sec denaturation at 94ºC, 40 sec annealing at 51ºC, and 40 sec extension at 72ºC. Primers were kindly provided by Kim Ki-Joong: *atpF-atpH*- *atpF* 5'-ACTCGCACACACTCCCTTTCC-3', *atpH* 5'-GCTTTTATGGAAGCTTTAACAAT-3'; and *psbK-psbI*: *psbK*- 5'-TTAGCCTTTGTTTGGCAAG-3', *psbI*- 5'-AGAGTTTGAGAGTAAGCAT-3'. After cycle sequencing using Big Dye terminator v3.1 and sequencing on a 3130*xl* genetic analyzer (Applied Biosystems, UK), electropherograms were edited using SEQUENCER 4.6 software (Genes Codes Corporation, USA) and DNA sequences aligned by eye in

PAUP4b10* [22] (incomplete sequences at both ends were excluded from the analyses). Taxa with missing data (amplification or sequencing failed) were removed from the combined matrix in order to analyze complete matrices.

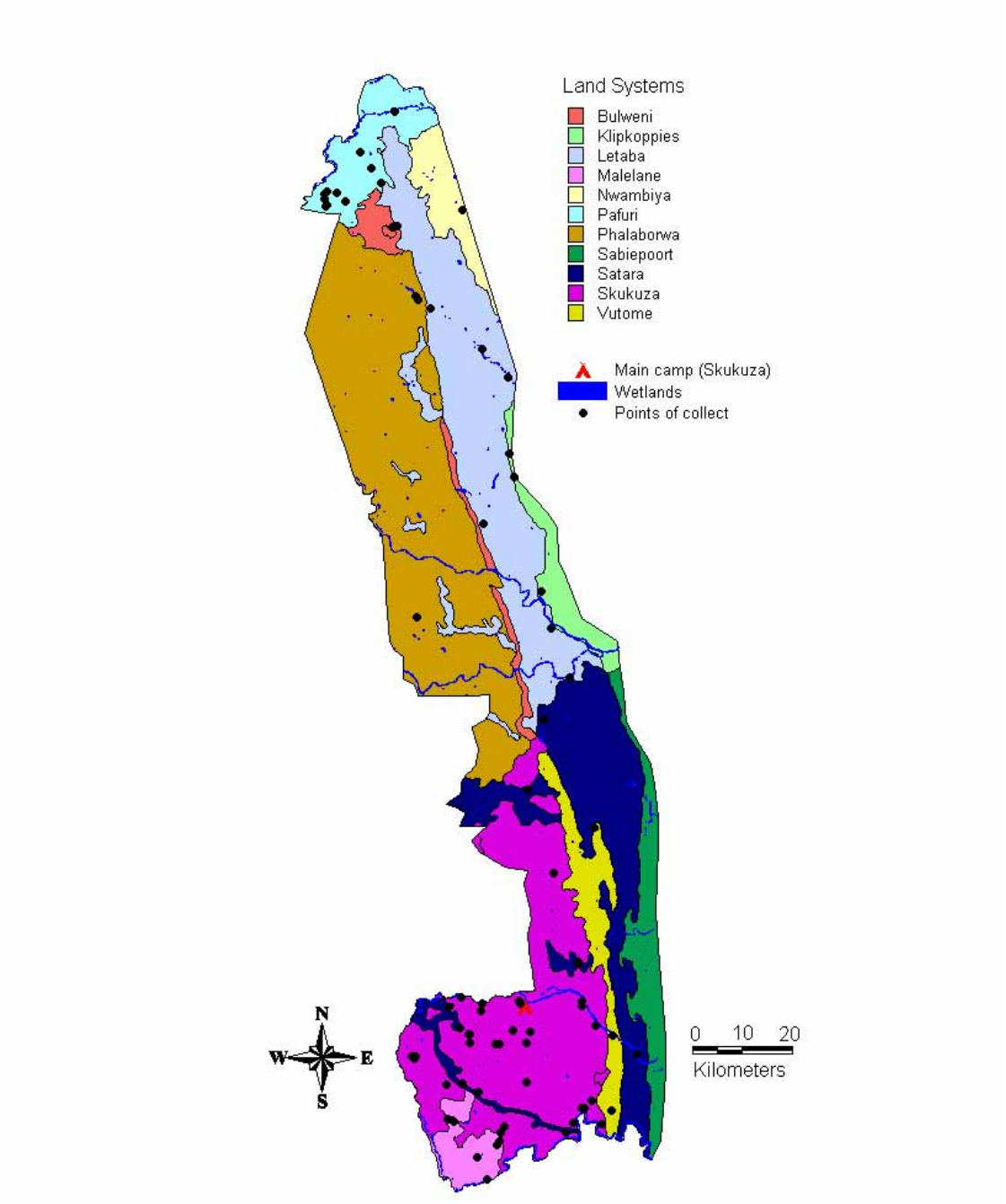

Figure 1. Map of the KNP with landsystems following Venter (1990) and collecting points from this study

.

Plant Family	**name Checked on IPNI**	**Voucher**	**Location**	**GPS**	**Altitude**	***matK***	***trnH-psbA***	***atpF-atpH***	***psbK-psbI***
Fabaceae	*Acacia exuvialis* Verdoorn	OM260	KNP	S24 58 54.3 E31 34 26.3	284 m	EU214205	EU213781	-	EU626889
Fabaceae	*Acacia exuvialis* Verdoorn	RL1204	KNP	S25 29 35.4 E31 28 12.3	319 m	EU214206	EU213782	EU626796	EU626890
Fabaceae	*Acacia exuvialis* Verdoorn	RL1412	KNP	S25 21 41.5 E31 30 56.5	320 m	EU214207	EU213783	-	EU626891
Fabaceae	*Acacia nigrescens* Oliver	RL1111	KNP	S25 06 26.4 E31 30 24.5	452 m	EU214208	-	EU626797	EU626892
Fabaceae	*Acacia nigrescens* Oliver	RL1205	KNP	S25 29 35.4 E31 28 12.3	319 m	EU214209	EU213784	EU626798	EU626893
Fabaceae	*Acacia nigrescens* Oliver	RL1656	KNP	S22 41 29.6 E31 01 37.2	439 m	EU214210	EU213785	EU626799	EU626894
Fabaceae	*Acacia tortilis* Hayne	OM261	KNP	S24 59 20.9 E31 34 34.5	266 m	EU214213	EU213788	EU626800	EU626895
Fabaceae	*Acacia tortilis* Hayne	RL1483	KNP	S24 36 53.6 E31 40 51.4	333 m	EU214211	EU213786	EU626801	EU626896
Fabaceae	*Acacia tortilis* Hayne	RL1608	KNP	S22 57 38.1 E31 14 50.5	302 m	EU214212	EU213787	EU626802	EU626897
Orchidaceae	*Acampe praemorsa* (Roxb.) Blatt. & McCann	RBN203	KNP	S22 42 06.1 E30 58 14.4	504 m	EU214214	EU213789	EU626803	EU626898
Amborellaceae	*Amborella trichopoda* Baill.	-	-	-	-	AJ506156	AJ506156	AJ506156	AJ506156
Orchidaceae	*Ansellia africana* Lindl.	OM1163	KNP	S25 12 54.8 E31 35 36.0	280 m	EU214215	-	EU626804	EU626899
Orchidaceae	*Ansellia africana* Lindl.	OM531	KNP	S25 19 54.3 E31 44 28.5	225 m	EU214216	-	EU626805	EU626900
Orchidaceae	*Bonatea speciosa* Willd.	RL1158	KNP	S25 13 11.4 E31 23 41.8	472 m	EU214217	EU213790	EU626806	EU626901
Asteraceae	*Brachylaena huillensis* O.Hoffm.	OM1281	KNP	S23 28 54.6 E31 33 27.0	421 m	EU214218	EU213791	EU626807	EU626902
Asteraceae	*Brachylaena huillensis* O.Hoffm.	OM247	KNP	S25 06 12.7 E31 35 44.2	276 m	EU214219	EU213792	EU626808	EU626903
Asteraceae	*Brachylaena huillensis* O.Hoffm.	RBN360	KNP	S22 42 51.4 E31 23 46.3	507 m	EU214220	EU213793	EU626809	EU626904
Combretaceae	*Combretum apiculatum* Sond.	RL1100	KNP	S25 06 24.7 E31 30 41.4	389 m	EU214221	EU213794	EU626810	EU626905
Combretaceae	*Combretum apiculatum* Sond.	RL1185	KNP	S25 23 11.2 E31 30 42.1	391 m	EU214222	EU213795	EU626811	EU626906
Combretaceae	*Combretum apiculatum* Sond.	RL1355	KNP	S25 20 11.4 E31 49 48.0	213 m	EU214223	EU213796	EU626812	EU626907
Combretaceae	*Combretum collinum* Fresen.	OM722	KNP	S25 00 07.4 E31 21 07.0	378 m	EU214224	EU213797	EU626813	EU626908
Combretaceae	*Combretum collinum* Fresen.	RL1164	KNP	S25 14 44.5 E31 26 39.8	419 m	EU214225	EU213798	EU626814	EU626909
Combretaceae	*Combretum collinum* Fresen.	RL1392	KNP	S25 25 45.2 E31 26 26.4	334 m	EU214226	EU213799	EU626815	EU626910
Combretaceae	*Combretum hereroense* Schinz	RL1120	KNP	S25 06 28.6 E31 29 58.5	383 m	EU214227	EU213800	EU626816	EU626911
Combretaceae	*Combretum hereroense* Schinz	RL1183	KNP	S25 23 11.2 E31 30 42.1	391 m	EU214228	EU213801	EU626817	EU626912
Combretaceae	*Combretum hereroense* Schinz	RL1364	KNP	S25 17 18.5 E31 46 34.6	235 m	EU214229	EU213802	EU626818	EU626913
Euphorbiaceae	*Croton gratissimus* Burch	OM785	KNP	S23 48 24.9 E31 38 27.2	285 m	EU214230	EU213803	EU626819	EU626914
Euphorbiaceae	*Croton gratissimus* Burch	RL1619	KNP	S22 45 43.6 E31 10 50.8	379 m	EU214231	EU213804	EU626820	EU626915
Euphorbiaceae	*Croton gratissimus* Burch	RL1621	KNP	S22 45 52.1 E31 10 29.1	414 m	EU214232	EU213805	EU626821	EU626916

Plant Family	**name Checked on IPNI**	**Voucher**	**Location**	**GPS**	**Altitude**	***matK***	***trnH-psbA***	***atpF-atpH***	***psbK-psbI***
Euphorbiaceae	*Croton megalobotrys* Müll.Arg.	OM774	KNP	S24 03 13.4 E31 43 50.0	211 m	EU214233	EU213806	EU626822	EU626917
Euphorbiaceae	*Croton megalobotrys* Müll.Arg.	RL1540	KNP	S23 54 53.6 E31 40 18.7	201 m	EU214234	EU213807	EU626823	EU626918
Euphorbiaceae	*Croton megalobotrys* Müll.Arg.	RL1574	KNP	S23 11 37.5 E31 32 16.5	246 m	EU214235	EU213808	EU626824	EU626919
Euphorbiaceae	*Croton pseudopulchellus* Pax	RBN186	KNP	S22 39 57.7 E30 59 57.6	468 m	EU214236	EU213809	EU626825	EU626920
Euphorbiaceae	*Croton pseudopulchellus* Pax	RBN262	KNP	S22 26 00.7 E31 10 57.6	291 m	EU214237	EU213810	EU626826	EU626921
Euphorbiaceae	*Croton pseudopulchellus* Pax	RL1650	KNP	S22 40 09.2 E30 57 39.6	451 m	EU214238	EU213811	-	EU626922
Orchidaceae	*Eulophia* R.Br.	OM473	KNP	S25 03 40.0 E31 23 11.2	351 m	EU214239	EU213812	EU626827	EU626923
Proteaceae	*Faurea rochetiana* Chiov. ex Pic.Serm.	OM1413	KNP	S25 08 43.0 E31 14 33.4	726 m	EU214240	EU213813	EU626828	EU626924
Proteaceae	*Faurea rochetiana* Chiov. ex Pic.Serm.	OM1461	KNP	S25 08 43.6 E31 14 32.6	722 m	EU214241	EU213814	EU626829	EU626925
Proteaceae	*Faurea rochetiana* Chiov. ex Pic.Serm.	OM1463	KNP	S25 08 43.1 E31 14 33.1	727 m	EU214242	EU213815	EU626830	EU626926
Proteaceae	*Faurea saligna* Harv.	OM1438	KNP	S25 19 31.7 E31 21 42.3	486 m	EU214243	EU213816	EU626831	EU626927
Proteaceae	*Faurea saligna* Harv.	OM1443	KNP	S25 19 16.9 E31 20 59.5	523 m	EU214244	EU213817	EU626832	EU626928
Proteaceae	*Faurea saligna* Harv.	OM1445	KNP	S25 19 39.5 E31 22 08.8	473 m	EU214245	EU213818	EU626833	EU626929
Moraceae	*Ficus abutilifolia* Miq.	OM557	KNP	S25 04 41.4 E31 24 54.5	414 m	EU214248	EU213821	EU626834	EU626930
Moraceae	*Ficus abutilifolia* Miq.	RL1471	KNP	S24 52 32.9 E31 45 21.9	256 m	EU214246	EU213819	EU626835	EU626931
Moraceae	*Ficus abutilifolia* Miq.	RL1501	KNP	S24 22 39.3 E31 35 51.8	369 m	EU214247	EU213820	EU626836	EU626932
Moraceae	*Ficus glumosa* Delile	OM564	KNP	S25 04 36.8 E31 25 03.7	473 m	EU214249	EU213822	EU626837	EU626933
Moraceae	*Ficus glumosa* Delile	RL1407	KNP	S25 23 41.1 E31 30 02.4	466 m	EU214250	EU213823	EU626838	EU626934
Moraceae	*Ficus glumosa* Delile	RL1429	KNP	S25 08 29.6 E31 14 42.6	665 m	EU214251	EU213824	EU626839	-
Moraceae	*Ficus sycomorus* L.	RBN197	KNP	S22 40 53.4 E30 57 43.2	445 m	EU214252	EU213825	EU626840	EU626935
Moraceae	*Ficus sycomorus* L.	RL1598	KNP	S23 06 46.1 E31 27 16.5	264 m	EU214253	EU213826	EU626841	EU626936
Moraceae	*Ficus sycomorus* L.	RL1614	KNP	S22 45 43.1 E31 11 18.4	356 m	EU214254	EU213827	EU626842	EU626937
Malvaceae	*Grewia bicolor* Juss.	OM329	KNP	S25 04 18.8 E31 36 29.5	363 m	EU214255	EU213828	EU626843	EU626938
Malvaceae	*Grewia bicolor* Juss.	RL1545	KNP	S23 36 52.2 E31 27 36.5	290 m	EU214256	EU213829	EU626844	EU626939
Malvaceae	*Grewia bicolor* Juss.	RL1658	KNP	S22 41 29.6 E31 01 37.2	439 m	EU214257	EU213830	EU626845	EU626940
Malvaceae	*Grewia flavescens* Juss.	OM323	KNP	S25 04 18.8 E31 36 29.5	363 m	EU214258	EU213831	-	EU626941
Malvaceae	*Grewia flavescens* Juss.	RL1472	KNP	S24 52 32.9 E31 45 21.9	256 m	EU214259	EU213832	EU626846	EU626942
Malvaceae	*Grewia flavescens* Juss.	RL1604	KNP	S22 58 18.8 E31 15 13.5	305 m	EU214260	-	-	-
Malvaceae	*Grewia villosa* Willd.	RL1342	KNP	S24 58 56.5 E31 46 02.3	208 m	EU214261	EU213833	-	EU626943
Malvaceae	*Grewia villosa* Willd.	RL1523	KNP	S24 10 31.8 E31 38 53.8	255 m	EU214262	EU213834	EU626847	EU626944

Plant Family	**name Checked on IPNI**	**Voucher**	**Location**	**GPS**	**Altitude**	***matK***	***trnH-psbA***	***atpF-atpH***	***psbK-psbI***
Malvaceae	*Grewia villosa* Willd.	RL1569	KNP	S23 24 48.9 E31 32 21.8	363 m	EU214263	EU213835	-	EU626945
Apiaceae	*Heteromorpha arborescens* Cham. & Schltdl.	OM1430	KNP	S25 13 27.0 E31 20 34.3	655 m	EU214264	EU213836	EU626848	EU626946
Apiaceae	*Heteromorpha arborescens* Cham. & Schltdl.	OM1488	KNP	S24 59 58.3 E31 21 04.3	359 m	EU214265	EU213837	EU626849	EU626947
Apiaceae	*Heteromorpha arborescens* Cham. & Schltdl.	OM1516	KNP	S25 20 29.0 E31 31 25.8	426 m	EU214266	EU213838	EU626850	EU626948
Hydnoraceae	*Hydnora johannis* Becc.	OM534	KNP	S25 21 37.5 E31 43 11.1	241 m	EU214267	-	-	EU626949
Arecaceae	*Hyphaene coriacea* Gaertn.	OM1184	KNP	S25 08 03.4 E31 56 37.7	167 m	EU214268	EU213775	EU626851	EU626950
Arecaceae	*Hyphaene coriacea* Gaertn.	OM1187	KNP	S25 17 45.4 E31 51 44.5	185 m	EU214269	EU213776	EU626852	EU626951
Arecaceae	*Hyphaene coriacea* Gaertn.	OM236	KNP	S25 03 08.3 E31 48 38.6	201 m	EU214271	EU213778	EU626853	EU626952
Arecaceae	*Hyphaene coriacea* Gaertn.	OM755	KNP	S24 29 10.7 E31 48 29.4	259 m	EU214270	EU213777	EU626854	EU626953
Arecaceae	*Hyphaene petersiana* Klotzsch ex Mart	OM1296	KNP	S22 38 18.4 E31 08 25.1	382 m	EU214272	EU213779	EU626855	EU626954
Arecaceae	*Hyphaene petersiana* Klotzsch ex Mart	OM908	KNP	S22 32 55.9 E31 04 25.5	347 m	EU214273	EU213780	EU626856	EU626955
Myrothamnaceae	*Myrothamnus flabellifolia* Welw.	OM1137	KNP	S25 06 15.4 E31 24 58.6	452 m	EU214275	EU213840	EU626857	EU626956
Myrothamnaceae	*Myrothamnus flabellifolia* Welw.	OM1209	KNP	S25 04 03.5 E31 33 04.7	485 m	EU214276	EU213841	EU626858	EU626957
Myrothamnaceae	*Myrothamnus flabellifolia* Welw.	OM285	KNP	S25 04 01.2 E31 33 04.8	577 m	EU214274	EU213839	EU626859	EU626958
Anacardiaceae	*Rhus gueinzii* Sond.	OM265	KNP	S24 59 25.4 E31 27 19.6	268 m	EU214277	EU213842	EU626860	EU626959
Anacardiaceae	*Rhus gueinzii* Sond.	RL1366	KNP	S25 17 23.1 E31 46 06.3	208 m	EU214278	EU213843	EU626861	EU626960
Anacardiaceae	*Rhus gueinzii* Sond.	RL1474	KNP	S24 52 08.3 E31 45 22.4	283 m	EU214279	EU213844	EU626862	EU626961
Anacardiaceae	*Rhus leptodictya* Diels	RBN205	KNP	S22 42 13.5 E30 57 56.4	487 m	EU214280	EU213845	EU626863	EU626962
Anacardiaceae	*Rhus leptodictya* Diels	RL1645	KNP	S22 42 06.5 E30 58 10.5	499 m	EU214281	EU213846	EU626864	EU626963
Anacardiaceae	*Rhus leptodictya* Diels	RL1655	KNP	S22 41 29.1 E31 01 38.4	448 m	EU214282	EU213847	EU626865	EU626964
Anacardiaceae	*Rhus transvaalensis* Engl.	OM282	KNP	S25 08 53.2 E31 14 38.3	664 m	EU214283	EU213848	EU626866	EU626965
Anacardiaceae	*Rhus transvaalensis* Engl.	OM943	KNP	S25 08 30.6 E31 14 07.8	610 m	-	EU213849	EU626867	EU626966
Anacardiaceae	*Rhus transvaalensis* Engl.	RL1427	KNP	S25 08 59.4 E31 14 35.0	630 m	EU214284	EU213850	EU626868	EU626967
Solanaceae	*Solanum panduriforme* Drège ex Dunal	OM1115	KNP	S25 00 44.2 E31 27 13.7	341 m	EU214285	EU213851	EU626869	EU626968
Solanaceae	*Solanum panduriforme* Drège ex Dunal	OM326	KNP	S25 04 18.8 E31 36 29.5	363 m	EU214286	EU213852	EU626870	EU626969
Solanaceae	*Solanum panduriforme* Drège ex Dunal	OM350	KNP	S25 04 17.5 E31 36 29.2	354 m	EU214287	EU213853	EU626871	EU626970
Apiaceae	*Steganotaenia araliacea* Hochst.	OM1350	KNP	S23 52 55.8 E31 15 00.9	422 m	EU214288	EU213854	EU626872	EU626971
Apiaceae	*Steganotaenia araliacea* Hochst.	OM1517	KNP	S23 52 56.3 E31 15 06.4	420 m	EU214289	EU213855	EU626873	EU626972
Apiaceae	*Steganotaenia araliacea* Hochst.	OM566	KNP	S25 04 36.8 E31 25 03.7	473 m	EU214290	EU213856	EU626874	EU626973
Orobanchaceae	*Striga elegans* Benth.	OM683	KNP	S25 04 02.4 E31 33 06.1	383 m	EU214291	-	EU626875	EU626974

Plant Family	**name Checked on IPNI**	**Voucher**	**Location**	**GPS**	**Altitude**	***matK***	***trnH-psbA***	***atpF-atpH***	***psbK-psbI***
Loganiaceae	*Strychnos decussata* (Pappe) Gilg	OM900	KNP	S22 35 35.0 E31 06 37.5	329 m	EU214292	EU213857	EU626876	EU626975
Loganiaceae	*Strychnos decussata* (Pappe) Gilg	RL1560	KNP	S23 24 53.0 E31 32 29.7	379 m	EU214293	EU213858	EU626877	EU626976
Loganiaceae	*Strychnos decussata* (Pappe) Gilg	RL1561	KNP	S23 24 53.0 E31 32 29.7	379 m	EU214294	EU213859	EU626878	EU626977
Loganiaceae	*Strychnos madagascariensis* Spreng. ex Baker	RL1433	KNP	S25 08 24.1 E31 14 51.5	641 m	EU214295	EU213860	EU626879	EU626978
Loganiaceae	*Strychnos madagascariensis* Spreng. ex Baker	RL1460	KNP	S24 58 21.4 E31 23 21.8	342 m	EU214296	EU213861	EU626880	EU626979
Loganiaceae	*Strychnos madagascariensis* Spreng. ex Baker	RL1559	KNP	S23 24 53.0 E31 32 29.7	379 m	EU214297	EU213862	EU626881	EU626980
Loganiaceae	*Strychnos spinosa* Lam.	OM220	KNP	S24 59 49.9 E31 46 10.3	208 m	EU214298	EU213863	EU626882	EU626981
Loganiaceae	*Strychnos spinosa* Lam.	RL1346	KNP	S25 04 51.2 E31 51 53.2	185 m	EU214299	EU213864	EU626883	EU626982
Loganiaceae	*Strychnos spinosa* Lam.	RL1652	KNP	S22 39 39.3 E30 58 17.4	430 m	EU214300	EU213865	EU626884	EU626983
Loranthaceae	*Tapinanthus* Blume	OM825	KNP	S22 59 46.4 E31 17 32.6	312 m	EU214301	-	EU626885	EU626984
Velloziaceae	*Xerophyta retinervis* Baker	OM1213	KNP	S25 08 32.4 E31 14 23.7	678 m	EU214302	EU213866	EU626886	EU626985
Velloziaceae	*Xerophyta retinervis* Baker	OM516	KNP	S25 16 03.6 E31 47 53.3	267 m	EU214303	EU213867	EU626887	EU626986
Velloziaceae	*Xerophyta retinervis* Baker	OM562	KNP	S25 04 36.8 E31 25 03.7	473 m	EU214304	EU213868	EU626888	EU626987

Table 1. Material sampled for this study, species checked in IPNI, voucher, GPS and altitude information, GenBank accession numbers. All vouchers have been collected in triplicate, one for Kew Herbarium, one for the herbarium of the KNP at Skukuza (South Africa), and one for the National Herbarium at Pretoria (South Africa).

Genetic analyses. Inter- and intra-specific genetic divergences were calculated using each potential DNA barcode following Meyer and Paulay [23]. Three different metrics were used to characterize intra-specific divergence: (i) average pairwise distances between all individuals sampled within those species that had at least two representatives, (ii) 'mean theta', with theta being the average pairwise distances calculated for each species that had more than one representative, thereby eliminating biases associated with uneven sampling among taxa and (iii) average coalescent depth, i.e. the depth of a node linking all sampled extant members of a species, 'book-ending' intra-specific variability. Genetic distances between con-generic species were used to characterize inter-specific divergence. For each barcode, pairwise distances were calculated with the simplest K2P model following Lahaye et al. [20] in which this model showed the best results. This model also utilizes the CBOL advises for distance calculations (barcoding.si.edu/). Wilcoxon Signed Rank Tests were performed to compare intra- and inter-specific variability for every pair of barcodes following Kress and Erickson [24]. We evaluated 'DNA barcoding gaps' [23] by comparing the distribution of intra- versus inter-specific divergences. Median and Wilcoxon Two-Sample Tests were used to evaluate whether these distributions overlapped.

Phylogenetic analyzes. To evaluate whether species were recovered as monophyletic with each barcode, we used standard phylogenetic techniques. Note that this is not to say that barcodes can be used to reconstruct phylogenies, because in this case we are disregarding the recovered inter-specific relationships. Trees were built with PAUP4b10* [22] using Maximum Parsimony (MP) and UPGMA, the two best algorithms in terms of percentages of species correctly identified [20]. UPGMA trees were inferred with PAUP4b10* from K2P distances. MP analyses were performed using tree bisection-reconnection (TBR), branch swapping and 1,000 random addition sequence replicates keeping 10 trees at each step. MP analyses have been performed with and without coding indels as a 5th state in order to assess the impact of keeping this information for barcoding purposes.

Coalescence analyses. For each barcode, we identified those clusters that were derived from an independent coalescence process and asked whether they matched previously recognized taxonomic species, using methods developed by Pons et al. [25] and Fontaneto

et al. [26]. The likelihood of waiting times between successive branching events on a DNA barcode-based tree was calculated under the null model that all terminals were derived from a single coalescence process, and under the alternative model that all taxa derived from a set of two independently evolving populations. With the alternative model, a threshold age T was calculated, at which point the older nodes represented inter-specific diversification events whereas the younger nodes represented separate coalescent processes typical of intra-specific clusters. We used DNA barcode-based trees from MP and transformed branch lengths with nonparametric rate smoothing [27] to produce ultrametric trees, i.e. branch lengths reflecting time only. We also used the ultrametric UPGMA trees. Likelihood models were determined using an R script available from TGB.

Results & Discussion

Molecular characteristics and PCR success. Amplification was generally successful for each potential barcode tested with more than 92% of taxa successfully amplified and sequenced (Table 2). The best percentage was given by *matK* with 99% of taxa sequenced and the lowest percentage was obtained for *trnH-psbA* with 92%. The potential DNA barcode *psbK-psbI* showed PCR and sequencing performances very close to those of *matK* with 98% of taxa successfully amplified. Both *atpF-atpH* and *trnH-psbA* failed to amplify the parasitic/non-chlorophytic plant *Hydnora johanis*. Alignment of sequences was unproblematic for *matK* and *psbK-psbI*, but *trnH-psbA* and *atpF-atpH* presented significant difficulties due to a high level of length variation (225 to 758 bp and 218 to 847 bp, respectively). Because its alignment was not reliable by Clustal X, we performed a first visual alignment between congeneric species and then aligned all taxa by adding as many gaps as necessary to keep the homology between congeneric species for inter- and intraspecific calculations. The alignment of *trnH-psbA* revealed a highly conservative intron only for the Orchidaceae and Amaryllidaceae which has been identified previously [20, 28]. Combining *matK* to one of the other potential barcodes allowed building a matrix including sequences for all taxa (Table 2).

matK	99%
psbK-psbI	98%
trnH-psbA	92.1%
atpF-atpH	93.1%
matK+trnH-psbA	100%
matK+trnH-psbA+atpF-atpH	100%
matK+trnH-psbA+psbK-psbI	100%
matK+atpF-atpH	100%
matK+psbK-psbI	100%
matK+atpF-atpH+psbK-psbI	100%
4 loci	100%

Table 2. Percentages of taxa represented in each matrix by at least one sequence.

Intra- and Inter-specific diversities. Performances of each DNA barcode was assessed by means of inter- and intra-specific diversity calculated from K2P (Kimura's two parameters) pairwise distance matrices (barcoding.si.edu/; Table 3). The highest inter-specific diversity was reached by *atpF-atpH* (3.45%) followed by *trnH-psbA* (2.55%) and the lowest was given by *psbK-psbI* (1.06%) with *matK* between these (1.34%). Regarding the different metrics to infer the intra-specific differences, the mean theta was in most cases similar to the average of overall intra-specific distances because there is no bias associated with species over-sampled in our study with the majority of the species represented by three specimens. The mean coalescent depth was slightly superior to the average of overall interspecific distances because it takes into consideration only the highest distance between specimens sampled for a species. Results showed the highest mean of intraspecific differences for *trnH-psbA* regardless of the metric used (Table 3). The lowest values were obtained for both *atpF-atpH* and *psbK-psbI*. Wilcoxon rank tests performed on the different distance matrices showed with very high significance that *trnH-psbA* had by far the highest inter-specific variability, followed by *matK* and *atpF-atpH,* which had a similar divergence (Table 4). The highest intra-specific distances were also significantly reached by *trnH-psbA* whereas the three other loci presented almost similar values (Table 5).

	matK	trnH-psbA	atpF-atpH	psbK-psbI	4 loci	matK+trnH-psbA	matK+atpF-atpH+trnH-psbA	matK+psbK-psbI+trnH-psbA	matK+atpF-atpH	matK+psbK-psbI	matK+psbK-psbI+atpF-atpH
Mean of all interspecific distances	0.0134	0.0255	0.0345	0.0106	0.0172	0.0175	0.0189	0.0157	0.0168	0.0118	0.0150
St. deviation +/-	0.0127	0.0227	0.0665	0.0096	0.0151	0.0154	0.0180	0.0121	0.0201	0.0092	0.0159
Mean of all intraspecific distances	0.0012	0.0017	0.0004	0.0005	0.0009	0.0012	0.0009	0.0011	0.0007	0.0009	0.0007
St. deviation +/-	0.0040	0.0041	0.0015	0.0012	0.0015	0.0026	0.0017	0.0021	0.0020	0.0026	0.0016
Mean Theta	0.0012	0.0015	0.0007	0.0005	0.0008	0.0012	0.0009	0.0010	0.0007	0.0009	0.0007
St. deviation +/-	0.0037	0.0032	0.0023	0.0010	0.0013	0.0023	0.0015	0.0018	0.0018	0.0024	0.0015
Mean coalescent depth	0.0017	0.0023	0.0008	0.0008	0.0013	0.0017	0.0014	0.0016	0.0012	0.0013	0.0011
St. deviation +/-	0.0050	0.0047	0.0023	0.0016	0.0018	0.0032	0.0021	0.0026	0.0026	0.0033	0.0021
Number of measurements for all intraspecific distances	93	90	84	91	95	95	95	95	95	95	95
Number of measurements for all interspecific distances	200	194	168	194	206	206	206	206	206	206	206

Table 3. Measures of inter- and intra-specific K2P distances for four potential barcodes and different combinations applied to a selective sampling from the KNP.

Wilcoxon Signed-Ranks Test – Interspecific pair-distances		
matK vs trnH-psbA	W+ = 1462, W- = 14648, N = 179, p <= 2.216e-21	**matK<<trnH-psbA**
matK vs atpF-atpH	W+ = 4977, W- = 5608, N = 145, p <= 0.5341	**matK = atpF-atpH**
matK vs psbK-psbI	W+ = 8655, W- = 6051, N = 171, p <= 0.0447	**matK > psbK-psbI**
trnH-psbA vs atpF-atpH	W+ = 8482, W- = 3608, N = 155, p <= 1.345e-05	**trnH-psbA > atpF-atpH**
trnH-psbA vs psbK-psbI	W+ = 13538, W- = 2572, N = 179, p <= 2.88e-15	**trnH-psbA >> psbK-psbI**
atpF-atpH vs psbK-psbI	W+ = 7663, W- = 2922, N = 145, p <= 2.902e-06	**atpF-atpH > psbK-psbI**
4 loci vs matK+trnH-psbA	W+ = 7286, W- = 12217, N = 197, p <= 0.002095	**4 loci < matK+trnH-psbA**
4 loci vs matK+trnH-psbA+atpF-atpH	W+ = 5244, W- = 14259, N = 197, p <= 1.859e-08	**4 loci < matK+trnH-psbA+atpF-atpH**
4 loci vs matK+trnH-psbA+psbK-psbI	W+ = 6661, W- = 12060, N = 193, p <= 0.0005137	**4 loci < matK+trnH-psbA+psbK-psbI**
4 loci vs matK+atpF-atpH	W+ = 14310, W- = 5193, N = 197, p <= 1.284e-08	**4 loci > matK+atpF-atpH**
4 loci vs matK+psbK-psbI	W+ = 15830, W- = 3673, N = 197, p <= 3.333e-14	**4 loci > matK+psbK-psbI**
4 loci vs matK+psbK-psbI+atpF-atpH	W+ = 15351, W- = 4152, N = 197, p <= 2.807e-12	**4 loci < matK+atpF-atpH+psbK-psbI**
matK+trnH-psbA vs matK+trnH-psbA+atpF-atpH	W+ = 12287, W- = 6434, N = 193, p <= 0.0001661	**matK+trnH-psbA > matK+trnH-psbA+atpHF**
matK+trnH-psbA vs matK+trnH-psbA+psbK-psbI	W+ = 13374, W- = 6129, N = 197, p <= 6.174e-06	**matK+trnH-psbA > matK+trnH-psbA+psbK-psbI**
matK+trnH-psbA vs matK+atpF-atpH	W+ = 13379, W- = 6124, N = 197, p <= 5.995e-06	**matK+trnH-psbA > matK+atpF-atpH**
matK+trnH-psbA vs matK+psbK-psbI	W+ = 16218, W- = 3285, N = 197, p <= 7.1e-16	**matK+trnH-psbA >> matK+psbK-psbI**
matK+trnH-psbA vs matK+atpF-atpH+psbK-psbI	W+ = 13179, W- = 6324, N = 197, p <= 1.894e-05	**matK+trnH-psbA > matK+atpF-atpH+psbK-psbI**

Table 4. Wilcoxon signed rank tests of inter-specific divergence among loci.

	Wilcoxon Signed-Ranks Test - Intraspecific pair-distances	
matK vs trnH-psbA	W+ = 298, W- = 605, N = 42, p <= 0.05574	**matK < trnH-psbA**
matK vs atpF-atpH	W+ = 334, W- = 162, N = 31, p <= 0.09384	**matK = atpF-atpH**
matK vs psbK-psbI	W+ = 299, W- = 229, N = 32, p <= 0.5189	**matK = psbK-psbI**
trnH-psbA vs atpF-atpH	W+ = 340, W- = 95, N = 29, p <= 0.008339	**trnH-psbA > atpF-atpH**
trnH-psbA vs psbK-psbI	W+ = 375, W- = 121, N = 31, p <= 0.01318	**trnH-psbA > psbK-psbI**
atpF-atpH vs psbK-psbI	W+ = 89, W- = 142, N = 21, p <= 0.3662	**atpF-atpH = psbK-psbI**
4 loci vs matK+trnH-psbA	W+ = 450, W- = 981, N = 53, p <= 0.01898	**4 loci < matK+trnH-psbA**
4 loci vs matK+trnH-psbA+atpF-atpH	W+ = 486, W- = 945, N = 53, p <= 0.04263	**4 loci < matK+trnH-psbA+atpF-atpH**
4 loci vs matK+trnH-psbA+psbK-psbI	W+ = 319, W- = 1007, N = 51, p <= 0.001283	**4 loci < matK+trnH-psbA+psbK-psbI**
4 loci vs matK+atpF-atpH	W+ = 923, W- = 508, N = 53, p <= 0.06687	**4 loci = matK+atpF-atpH**
4 loci vs matK+psbK-psbI	W+ = 901, W- = 530, N = 53, p <= 0.1015	**4 loci = matK+psbK-psbI**
4 loci vs matK+psbK-psbI+atpF-atpH	W+ = 906, W- = 525, N = 53, p <= 0.09256	**4 loci = matK+atpF-atpH+psbK-psbI**
matK+trnH-psbA vs matK+trnH-psbA+atpF-atpH	W+ = 810, W- = 271, N = 46, p <= 0.003294	**matK+trnH-psbA > matK+trnH-psbA+atpHF**
matK+trnH-psbA vs matK+trnH-psbA+psbK-psbI	W+ = 833, W- = 392, N = 49, p <= 0.02864	**matK+trnH-psbA > matK+trnH-psbA+psbK-psbI**
matK+trnH-psbA vs matK+atpF-atpH	W+ = 924, W- = 252, N = 48, p <= 0.0005795	**matK+trnH-psbA > matK+atpF-atpH**
matK+trnH-psbA vs matK+psbK-psbI	W+ = 854, W- = 371, N = 49, p <= 0.01652	**matK+trnH-psbA > matK+psbK-psbI**
matK+trnH-psbA vs matK+atpF-atpH+psbK-psbI	W+ = 1068, W- = 363, N = 53, p <= 0.001832	**matK+trnH-psbA > matK+atpF-atpH+psbK-psbI**

Table 5. Wilcoxon signed rank tests of intra-specific difference among loci.

In a multi loci approach for DNA barcoding purposes, the highest mean of inter-specific variability was achieved by *matK* combined with *trnH-psbA* and *atpF-atpH* whereas the highest mean of intra-specific distances were given by combining *matK* with *trnH-psbA* (Table 3). Wilcoxon statistical rank tests showed the combination *matK* + *trnH-psbA* having the highest inter-specific pair-distances (Table 4). They revealed also that all the combinations including *trnH-psbA* had a higher intra-specific variability than combinations without it (Table 5).

Distribution of distances. Accuracy of each DNA barcode was assessed by looking at the distribution of inter- and intraspecific K2P distances to infer the barcoding gap [23]. Distributions were similar for each single potential barcode with two peaks of inter- and intraspecific variability that could be distinguished (Figure 2).

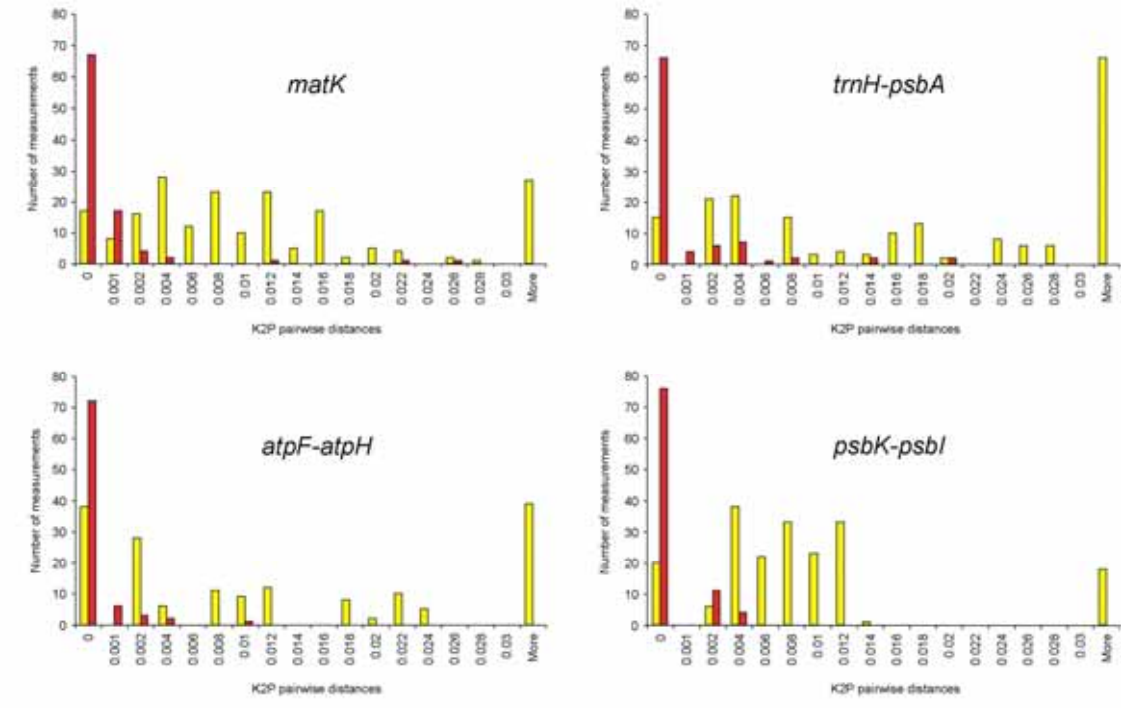

Figure 2. Relative distributions of inter-specific divergence between con-generic species (yellow) and intra-specific K2P distances (red) for four single loci: *matK*, *trnH-psbA*, *psbK-psbI* and *atpF-atpH*. Barcoding gaps were assessed with Median tests and Wilcoxon Two-Sample tests, and all were highly significant ($p<0.0001$).

Each distribution also showed a slight overlap between intra- and inter-specific distances, but to a lesser extent for *matK* and *trnH-psbA*. Combining the different loci showed distributions with a slight decrease of this overlap (Figure 3).

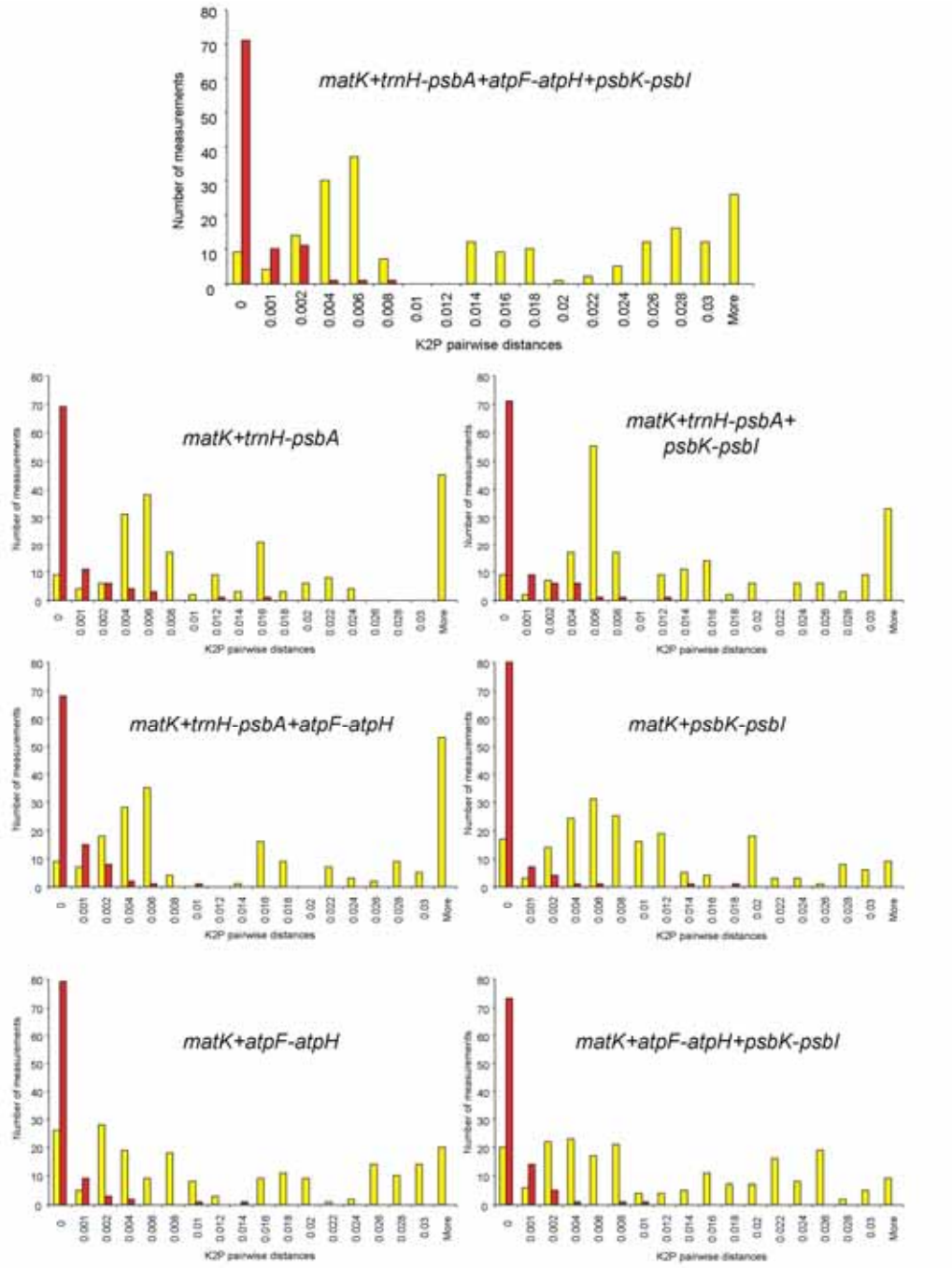

Figure 3. Relative distributions of inter-specific divergence between con-generic species (yellow) and intra-specific K2P distances (red) for 7 different combinations keeping *matK* for each. Barcoding gaps were assessed with Median tests and Wilcoxon Two-Sample tests, and all were highly significant ($p<0.0001$).

Two clear peaks were still distinguishable and a slight overlap still occurred between low classes of intra- and inter-specific distances, but the overlap observed was less than that for the single locus approach. These observations were confirmed by median and Wilcoxon two samples statistical tests differentiating the medians for the former and the medians plus the distributions between the inter- and intra-specific distances for the latter. For each distribution, Median and Wilcoxon two sample tests were significant (Table 6). In a single locus approach, the highest significances were given by *matK* and *psbK-psbI*. Combining the loci made the significance increasing with the highest significance given by the combination *matK+trnH-psbA+psbK-psbI*.

K2P distributions	median test	Wilcoxon Two Sample Test
matK	#A = 199 #B = 93, Median = 0.00524, p <= 1.11e-26	#A = 200 #B = 93, W = 6020.5, p <= 9.314e-30
trnH-psbA	#A = 194 #B = 90, Median = 0.00799, p <= 1.11e-22	#A = 194 #B = 90, W = 5634, p <= 6.125e-29
atpF-atpH	#A = 168 #B = 84, Median = 0.00216, p <= 1.52e-23	#A = 168 #B = 84, W = 5526, p <= 8.996e-21
psbK-psbI	#A = 194 #B = 91, Median = 0.00509, p <= 1.44e-29	#A = 194 #B = 91, W = 5333, p <= 2.524e-32
4 loci	#A = 206 #B = 95, Median = 0.00608, p <= 1.23e-28	#A = 206 #B = 95, W = 5507, p <= 2.394e-36
matK+trnH-psbA	#A = 206 #B = 95, Median = 0.00648, p <= 8.07e-28	#A = 206 #B = 95, W = 5675, p <= 4.825e-35
matK+trnH-psbA+atpF-atpH	#A = 206 #B = 95, Median = 0.00574, p <= 5.11e-29	#A = 206 #B = 95, W = 5642.5, p <= 2.711e-35
matK+trnH-psbA+psbK-psbI	#A = 206 #B = 95, Median = 0.00676, p <= 5.11e-29	#A = 206 #B = 95, W = 5540, p <= 4.338e-36
matK+atpF-atpH	#A = 206 #B = 95, Median = 0.00401, p <= 1.2e-26	#A = 206 #B = 95, W = 6318, p <= 2.802e-30
matK+psbK-psbI	#A = 206 #B = 95, Median = 0.00607, p <= 8.07e-28	#A = 206 #B = 95, W = 6064, p <= 4.064e-32
matK+atpF-atpH+psbK-psbI	#A = 206 #B = 95, Median = 0.00493, p <= 2.92e-28	#A = 206 #B = 95, W = 6026.5, p <= 2.151e-32

Table 6. Median and Wilcoxon two sample statistical tests applied to the distributions of intra- and inter-specific K2P distances for each potential DNA barcode.

Species identification. The performance of each DNA barcode in identifying and delineating species was assessed by the percentage of monophyletic species recovered by MP and UPGMA analyses (Table 7). Because *trnH-psbA* and *atpF-atpH* were highly variable and their alignment showed many indels, MP analyses were performed with and without coding the gaps as 5^{th} state to infer whether this information could be useful for barcoding purposes. The highest values of species monophyly were obtained from UPGMA reconstruction. The UPGMA analysis of *trnH-psbA* gave 90.3% of species monophyletic but only 77.4% supported by BS>50%. Although *matK* and *psbK-psbI* grouped 87.5% of the species under UPGMA reconstruction, they gave 78.1% of monophyletic species with a BS>50%, a value higher than *trnH-psbA*. *MatK* showed the

best percentage of species correctly identified using MP reconstruction. Coding the gaps as 5th state in the MP analysis did not significantly affect the results obtained for *matK* and *psbK-psbI*, but it increased the percentages of species correctly identified by 6% and 7% given by the more variable *atpF-atpH* and *trnH-psbA*, respectively. In a multi-loci approach, it is noteworthy that combining all potential barcodes did not result in 100% monophyly for species whatever the reconstruction method. Each barcode failed in grouping the two different species of *Faurea*. That can be done by using the intergenic locus *atpF-atpH* and by coding the gaps in the matrix as 5th state of character, but this decreases the total percentage of monophyletic species. In a multi-loci approach, combining *matK* and *psbK-psbI* gave the highest percentage of monophyletic species (Table 7).

	UPGMA	MP	MP+5th state character
trnH-psbA	90.3 (77.4)	71 (71)	77.4 (74.2)
matK	87.5 (78.1)	75 (75)	75 (75)
psbK-psbI	87.5 (78.1)	62.5 (68.8)	53.1 (53.1)
atpF-atpH	82.8 (69)	65.5 (65.5)	72.4 (69)
matK+psbK-psbI	93.8 (87.5)	81.3 (81.3)	59.4 (56.3)
matK+trnH-psbA+psbK-psbI	93.5 (90.3)	87.1 (87.1)	80.6 (80.6)
matK+atpF-atpH+psbK-psbI	93.1 (86.2)	86.2 (86.2)	82.8 (82.8)
matK+trnH-psbA+atpF-atpH+psbK-psbI	92.9 (89.3)	85.7 (85.7)	82.1 (82.1)
matK+trnH-psbA	90.3 (87.1)	83.9 (83.9)	77.4 (77.4)
matK+atpF-atpH	89.7 (82.8)	79.3 (79.3)	79.3 (79.3)
matK+trnH-psbA+atpF-atpH	89.3 (85.7)	82.1 (82.1)	82.1 (82.1)

Table 7. Proportion (%) of monophyletic species (with BS > 50% in brackets) recovered with UPGMA and MP analyses with gaps not coded and coded as a fifth character state.

Coalescence. The accuracy of the DNA barcode can be assessed by evaluating the ability of each candidate to give genetic clusters that are derived from an independent coalescence process and that corresponds to a recognized taxonomic species [25, 26]. The highest number of genetic clusters corresponding to taxonomic species was given using the UPGMA trees. Transforming MP trees by NPRS for coalescence analysis gave half the genetic clusters corresponding to taxonomic species compared to the UPGMA trees

(Table 7). In a single barcode approach, *matK* gave the highest numbers of genetic clusters corresponding to taxonomic species (Table 8). When *matK* was combined with *psbK-psbI* the value increased from 22 to 23 genetic clusters corresponding to recognized species. Molecular evolutionary rates of both *matK* and *psbK-psbI* showed higher abilities to differentiate independently evolving entities corresponding to taxonomic species than the high variable *trnH-psbA* and *atpF-atpH*.

	UPGMA	MP	Nos. of potential genetic clusters
matK	22	11	32
psbK-psbI	20	15	32
atpF-atpH	18	12	29
trnH-psbA	16	12	31
matK+psbK-psbI	23	8	32
matK+atpF-atpH+psbK-psbI	20	4	29
matK+atpF-atpH	20	6	29
matK+trnH-psbA+psbK-psbI	3	7	31
matK+trnH-psbA+atpF-atpH+psbK-psbI	3	1	28
matK+trnH-psbA	3	8	31
matK+trnH-psbA+atpF-atpH	3	5	28

Table 8. Coalescence analyses indicating the number of independent genetic clusters corresponding to taxonomically recognized species.

Our results showed that combining *matK* to *trnH-psbA* and *psb-psbI* can slightly increase its performance in identifying species. However we still support the conclusion of Lahaye et al. [20], i.e. that *matK* should be used for DNA barcoding of plants in a single locus approach and that case-by-case additional barcodes are developed for problematic groups.

Literature Cited

1. Savolainen, V., Cowan, R. S., Vogler, A. P., Roderick, G. K. & Lane, R. Towards writing the encyclopaedia of life: an introduction to DNA barcoding. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1805-1811 (2005).

2. Hebert, P. D. N., Cywinska, A., Ball, S. L. & deWaard, J. R. Biological identifications through DNA barcodes. *Proc. R. Soc. Biol. Sci. Ser. B* 270, 313-321 (2003).
3. Smith, M. A., Fisher, B. L. & Hebert, P. D. N. DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1825-1834 (2005).
4. Vences, M., Thomas, M., Bonett, R. M. & Vieites, D. R. Deciphering amphibian diversity through DNA barcoding: chances and challenges. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1859-1868 (2005).
5. Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R. & Hebert, P. D. N. DNA barcoding Australia's fish species. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1847-1857 (2005).
6. Janzen, D. H. et al. Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1835-1845 (2005).
7. Hajibabaei, M., Janzen, D. H., Burns, J. M., Hallwachs, W. & Hebert, P. D. N. DNA barcodes distinguish species of tropical Lepidoptera. *Proc. Natl. Acad. Sci. U. S. A.* 103, 968-971 (2006).
8. Smith, M. A., Woodley, N. E., Janzen, D. H., Hallwachs, W. & Hebert, P. D. N. DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera : Tachinidae). *Proc. Natl. Acad. Sci. U. S. A.* 103, 3657-3662 (2006).
9. Rubinoff, D. Utility of mitochondrial DNA barcodes in species conservation. *Conserv. Biol.* 20, 1026-1033 (2006).
10. Cho, Y., Mower, J. P., Qiu, Y.-L. & Palmer, J. D. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proc. Natl. Acad. Sci. U. S. A.* 101, 17741-17746 (2004).
11. Rubinoff, D., Cameron, S. & Will, K. Are plant DNA barcodes a search for the Holy Grail? *Trends Ecol. Evol.* 21, 1-2 (2006).
12. Pennisi, E. Wanted: A Barcode for Plants. *Science* 318, 190-191 (2007).
13. Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A. & Janzen, D. H. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U. S. A.* 102, 8369-8374 (2005).
14. Drager, R. G. & Hallick, R. B. A Novel Euglena-Gracilis Chloroplast Operon Encoding 4 Atp Synthase Subunits and 2 Ribosomal-Proteins Contains 17 Introns. *Curr. Genet.* 23, 271-280 (1993).
15. Meng, B. Y., Wakasugi, T. & Sugiura, M. 2 Promoters within the *Psbk-Psbl-Trng* Gene-Cluster in Tobacco Chloroplast DNA. *Curr. Genet.* 20, 259-264 (1991).
16. Knauf, U. & Hachtel, W. The genes encoding subunits of ATP synthase are conserved in the reduced plastid genome of the heterotrophic alga *Prototheca wickerhamii*. *Mol. Genet. Genom.* 267, 492-497 (2002).
17. Funk, H. T., Berg, S., Krupinska, K., Maier, U. G. & Krause, K. Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii*. *BMC Plant Biol.* 7 (2007).

18. McNeal, J. R., Kuehl, J. V., Boore, J. L. & de Pamphilis, C. W. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. *BMC Plant Biol.* 7 (2007).
19. Myers, N. Biodiversity hotspots revisited. *Bioscience* 53, 916-917 (2003).
20. Lahaye, R. et al. DNA Barcoding the Floras of Biodiversity Hotspots. *Proc. Natl. Acad. Sci. U. S. A.* 105, 2923-2928. (doi/10.1073/pnas.0709936105) (2008).
21. Doyle, J. J. & Doyle, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytoch. Bull.* 19, 11-15 (1987).
22. Swofford, D. L. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. (Sinauer Associates, Sunderland, Massachusetts, 2002).
23. Meyer, C. P. & Paulay, G. DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biol.* 3, 2229-2238 (2005).
24. Kress, W. J. & Erickson, D. L. A Two-Locus Global DNA Barcode for Land Plants: The Coding *rbcL* Gene Complements the Non-Coding *trnH-psbA* Spacer Region. *PLoS one* 2, e508 (2007).
25. Pons, J. et al. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst. Biol.* 55, 595-609 (2006).
26. Fontaneto, D. et al. Independently evolving species in asexual bdelloid rotifers. *PLoS Biol.* 5, 914-921 (2007).
27. Sanderson, M. J. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* 14, 1218-1231 (1997).
28. Chang, C. C. et al. The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): Comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Mol. Biol. Evol.* 23, 279-291 (2006).