UPPSALA
UNIVERSITET

DNA barcoding as a tool for the identification of unknown plant material

A case study on medicinal roots traded in the medina of Marrakech

Anders Rydberg

Degree project in biology, Master of science (1 year), 2010
Examensarbete i biologi 30 hp till magisterexamen, 2010
Biology Education Centre and Department of Systematic Biology, Uppsala University
Supervisors: Anneleen Kool and Hugo de Boer

Contents

Summary

DNA barcoding, which involves using a short, agreed-upon region of the genome for sequence comparison, has been proposed as a global standard for identifying unknown biological material to species level. Here I present a case study on the use of barcoding for identifying medicinal root material sold by herbalists in Marrakech, Morocco. In total, 111 roots samples bought in Marrakech markets were included in the study, and were sequenced for three proposed barcode regions in the plastid genome. The BLAST algorithm was then used to compare root sequences to an expanded reference database for Moroccan medicinal plants.

Sequencing success was relatively high for *rpoC1* and *psbA-trnH*, but low (<30%) for *matK*, perhaps indicating that a modified protocol is necessary for it to be used as a molecular barcode. A majority of the market samples were identified to a high level of accuracy when combining *rpoC1* and *psbA-trnH* BLAST results, although *rpoC1* searches alone resulted in a number of ambiguous identifications. For around 25% of the market samples, the barcoding identification differed significantly from previous hypotheses based on the vernacular name of the plant. Many of these samples are associated with species complexes, i.e. a group of (potentially unrelated) species sold under the same name. Samples with the local names „Bougoudz" and „Ndkhir" were identified as *Tamus communis* L., constituting apparently new knowledge for the Moroccan herbal pharmacopoeia.

The results suggest that DNA barcoding can provide accurate identifications for plant samples that are not distinguishable by morphology. A workflow for the identification of unknown ethnobiological material, based on the methods used in this study, is presented.

Introduction

DNA barcoding

DNA barcoding is a method for identifying living organisms to species. It makes use of a short (<1000 bp), agreed-upon region of the genome (a "barcode") that evolves fast enough to differ between closely related species. When a barcode sequence has been retrieved from an unknown sample, an algorithm is used to compare it to a reference database containing barcodes from identified museum specimens, thus enabling it to be identified. In other words, DNA barcodes function as molecular identifiers for each species, in the same way as the machine-readable black-and-white barcodes are used in the retail industry to identify commercial products.

In effect, barcoding in its modern form was popularized in a paper by Hebert *et al.* (2003), who proposed to use the mitochondrial gene CO1 as the standard barcode for all animals. This was readily adopted by the scientific community, and assessments have since shown that CO1 can be used to distinguish over 90% of species in most animal groups (e.g. Kerr et al. 2007, M.A. Smith et al. 2008). In recent years the barcoding movement has grown substantially, and worldwide efforts coordinated by CBOL (the Consortium for the Barcode of Life) are now being put into retrieving barcode sequences from all organisms (CBOL, 2009).

Barcoding of plants, however, has developed at a markedly slower pace. Early on, it became clear that the mitochondrial genome evolves far too slowly in plants to allow it to distinguish between species. Various genes and non-coding regions in the plastid genome have been put forward as alternatives. In addition to being sufficiently fast-evolving, a molecular barcode must also be flanked (surrounded) by conserved regions of the genome that can function as primer binding sites for PCR reactions. The ideal plant barcode needs to be amplifiable with only a single set of primers, so that it can be efficiently retrieved from any of the over 200,000 species of plants. A single barcoding locus combining these two traits has not been found, and a combination of two or more, probably plastid, loci will almost certainly be required to approach the level of species discrimination and universality that CO1 confers for animals (Kress & Erickson, 2007).

DNA barcoding has been applied to a broad range of subjects, including taxonomic studies of "cryptic" taxa or species complexes, e.g. skipper butterflies (Burns *et al.* 2008). Barcoding has also been used in ecological studies to survey animal diets through the analysis of plant remains in feces (Valentini *et al.* 2009) and in identifying smoked fish products sold under ambiguous product names (P.J. Smith *et al.* 2008). Barcoding lends itself particularly well to these "forensic" applications in that it requires only a small tissue sample from the organism in need of identification, and can be used in cases where the sample is degraded, has been processed, or when it consists of only non-characteristic parts of the organism.

The Moroccan traditional pharmacopoeia

Traditional medicine has played an important role in many North African societies, and continues to do so until this day. This is evident not least in the Moroccan city of Marrakech,

situated at a crossroads of trade routes between the High Atlas Mountains and surrounding coastal plains.

The traditional equivalent of the doctor in Moroccan medicine is the herbalist – a profession that continues to be practiced in Marrakech, manifested by the herbalist-owned drug stores that line the market districts of the medina, or old town. In these shops, Marrakech herbalists stock a variety of plant parts and plant-derived products, sold either separately or in mixtures. In general, these plant parts are harvested in the wild by specialized collectors and reach the herbalist via wholesalers (Bellakhdar, 1997).

An important part of the plant inventory of Moroccan herbalists consists of barks and roots, which overall possess few characteristic features that enable them to be identified by morphology (Fig. 1). As a matter of course, any herbalist will be able to provide information about the local name of a plant product, its medicinal uses and origins. However, this information may be imprecise or insufficient for species identification purposes, considering the fact that herbalists often do not possess knowledge of medicinal plants in the wild (Bellakhdar, 1997). In other words, confirming the origin of a root sample bought from these herbalists has so far presented somewhat of a challenge.

The identity of the plants being sold in these markets has conservational as well as medical implications. For example, rare or endangered species could inadvertently be collected if they are easily confused with their more abundant relatives. Likewise, increasing demands for medicinal products may lead to the overharvesting and local extinction of otherwise non-threatened plant species. Misidentified collections could also lead to the introduction of toxic or otherwise unsuitable species to the market, with potential health risks to end-users. In all of these cases, appropriate measures could be taken if a reliable method for identifying medicinal plant products to species existed.

The reference database

By nature, a molecular approach such as DNA barcoding would seem ideally suited to the task of identifying plant samples indistinguishable by morphology. A few studies have in fact demonstrated the potential of barcoding in an ethnobiological context, e.g. in discriminating new grass species used by hill tribes of the Western Ghats (Ragupathy *et al.* 2009) and for identifying medicinal leeches to species (Siddall *et al.* 2007). In general, however, these studies are rather limited in scope in that they focus on only one or a few species groups. Studies on identifying large amounts of ethnobotanical material (e.g. entire market inventories) have not yet been undertaken. This is perhaps a consequence of the fact that relatively few plant barcodes have been sequenced, which in turn stems from the lack of agreement on a standard barcode for plants. Given the current state of affairs, researchers intending to identify unknown plant material will need to sample their own reference databases.

In order to be of any use, a reference database must contain sequences retrieved from reliably identified museum specimens, and all species deemed candidates to the identity of a particular sample must be included. An existing reference database of Moroccan medicinal plants was compiled in a previous barcoding study on Marrakech root products (Krüger, 2008). The plant

species included in this database were chosen based mainly on the vernacular names of the market samples to be identified, which were translated to scientific names by means of the comprehensive Moroccan pharmacopoeia by Bellakhdar (1997). In Krüger's study many market samples could not be accurately identified, suggesting that the reference database was incomplete – perhaps due to inaccurate translations.

Aims

The current study, in effect a continuation and expansion of the study by Krüger (2008), intends to test DNA barcoding as a method to identify medicinal plant roots used in the traditional Moroccan pharmacopoeia. Specifically, the material to be identified is a selection of 111 plant samples bought from various herbalists in the markets of Marrakech. Emphasis will be placed on identifying samples of unclear affinity, and samples potentially involved in species complexes. A new approach will be used to complement the existing barcode reference database for Moroccan medicinal plants. The possible implications of the identifications from a conservational and ethnobiological perspective will be discussed. An evaluation of the methods used in this study with respect to their suitability in similar applications of barcoding will also be included, with focus on testing an actual approach which ethnobiologists could use to identify plants included in traditional medicinal products.

Fig. 1. A selection of dried root samples acquired from herbalists in the medina of Marrakech.

Materials and methods

Overview

To address previous problems owing to an incomplete reference database, this study used an integrated approach to barcoding and reference sampling, with DNA sequencing conducted in several steps. Firstly, previously unidentified market samples were sequenced for the nuclear ITS (internal transcribed spacer) region. These samples were then given preliminary identifications by comparing the sequences to online gene repositories, using the BLAST algorithm. On the basis of these BLAST searches, reference samples of species that were judged to be missing from the reference database were collected and sequenced for the barcoding regions *rpoC1*, *matK* and *psbA-trnH*. These loci are the same as in Krüger (2008), and constitute one of two options recommended for plants by the Royal Botanic Gardens at Kew (2007). All of the market samples were subsequently sequenced for these barcode regions as well. As a final step in the process, BLAST queries against the expanded reference database were conducted to conclusively identify the market samples.

Market samples

A total of 111 market samples of medicinal plant parts bought from herbalists in the medina of Marrakech comprised the basis of this study (mostly roots and barks, see Fig. 1). Ninety-six of these samples were collected in October and November 2007, and an additional 15 samples of products that were found difficult to sequence were collected in November 2008. The vernacular name for each sample (as communicated by the herbalist) was recorded, along with the herbalist‟s name and the place and date of purchase. In most cases several samples were collected per vernacular name (= "product"), such that the resulting collection comprises 46 different medicinal plant products. Some products are further divided into subtypes specified by modifiers placed after the main noun (e.g. „Tigndizt lghlida‟ vs. „Tigndizt rqiqa‟). Putative scientific names have been assigned to the material based on the Moroccan vernacular names, using the herbal pharmacopoeia of Morocco by Bellakhdar (1997). Where cited in the text, these proposed names are given in simple quotation marks, so as to distinguish them from barcoding identifications.

DNA extraction from market samples

The DNA isolation procedure used for the root material is a modification of the Carlson/Yoon method (Yoon *et al.* 1991), and is presented below:

A small amount of the sample was fragmented into coarse grains, if necessary using a scalpel. The sample fragments were transferred to a mortar and ground with grinding sand until homogenized. No more than 500 µl of the ground material was transferred to a 2 ml tube. To this was added 750 µl of CTAB (hexadecyl trimethyl ammonium bromide) and 20 µl mercaptoethanol. The tube was incubated at 65°C for 45 min, and mixed intermittently by inverting. Subsequently, 750 µl of chloroform/isoamylalchohol (24:1) was added and the tube put horizontally in a shaker and shaken at 100 rpm for 30 min. The tube was then centrifuged at 12000 x g for 10 min. The upper aqueous phase (containing the DNA) was transferred to a new 1.5 ml tube. This chloroform/isoamylalchohol extraction step was repeated and the tube centrifuged for 5 min. The aqueous phase was again transferred to a new 1.5 ml tube. The DNA was then precipitated with 0.1 vol. of 3M NaAc (pH 4.6) and 2 vol. of 95% EtOH and incubated at -20°C for 1 hour or overnight.

The precipitate was centrifuged at 12000 x g for 10 min, after which the liquid was discarded (the DNA pellet remaining at the bottom of the tube), and 750 µl of 70% EtOH added. This was mixed and centrifuged for 5 min. Again, the ethanol was poured off, the tube centrifuged for a few seconds, and the remaining liquid was removed with a pipette. The tube was placed horizontally in the fume hood (with the cap open) for 30 min. or until the remaining liquid had evaporated. 100 µl of EB buffer was added. This was left for 30 min. or more (mixed gently if necessary) until the pellet had dissolved. The dissolved DNA was subsequently purified using the GE illustra GFX™ PCR DNA and Gel Band Purification Kit following the standard protocol.

Extending the reference database

In this study, a sampling approach based on an intermediate ITS-sequencing step was used to complement the existing reference database: Most market samples that could not be identified by Krüger (2008) were sequenced for the nuclear ITS region (see below for sequencing/amplification procedure and Table 1 for the primers used). These ITS sequences were then queried using the BLASTn algorithm (via the NCBI web-interface, http://blast.ncbi.nlm.nih.gov/Blast.cgi) to the online nr-database, in general using default parameters. The highest-scoring hits from these queries were used as preliminary identifications and provided a basis of judgment for the further sampling of reference material.

In choosing which species to add to the reference database, only species known to occur in Morocco (as per Fennane *et al.* 1999-2007) were considered. All genera considered candidates to the identity of a certain market sample were sampled with at least one species each, while larger genera were sampled with up to three or four species. Comprehensive sampling e.g. of entire genera with five or more species was not deemed practical in most cases at this stage of the project.

DNA extraction from reference samples

In most cases, total DNA from leaf material of the reference samples was extracted and purified in the same way as for the market samples, though omitting the manual grinding process in favor of a Mini-Beadbeater (BioSpec Products) step:

Silica beads and ca. 0.02 g of plant material, 750 µl of CTAB (hexadecyl trimethyl ammonium bromide) and 20 µl mercaptoethanol were added to a 2 ml tube. The tube was put into the Mini-Beadbeater and shaken for 40 seconds or more, and subsequently incubated at 65°C for 45 min, intermittently mixed by inverting. The solution was transferred to a new 2 ml tube, whereupon chloroform/isoamylalchohol extraction, etc. was performed as detailed above under "DNA extraction from market samples".

PCR and sequencing

PCR amplification of purified total DNA was performed in 200 µl reaction tubes with a total volume of 50 µl. Each tube contained a mixture of 5 µl reaction buffer (ABgene, 10x), 3 µl $MgCl_2$ (25mM), 1 µl dNTP"s (10 µM), 0.25 µl Taq-polymerase (ABgene; 5U/µl), 0.25 µl BSA (Roche Diagnostics), 12.5 µl of each primer (2mM) and 1 µl template DNA.

The primers used at various stages of this study are listed in Table 1.

Table 1. Primers used.

Name	Sequence 5′-3′	Reference
ITS-5	GGAAGGAGAAGTCGTAACAAGG	Sang *et al.* (1995)
ITS-4	TCCTCCGCTTATTGATATGC	White *et al.* (1990)
rpoC1-2F	GGCAAAGAGGGAAGATTTCG	Royal Botanic Gardens, Kew (2007)
rpoC1-4R	CCATAAGCATATCTTGAGTTGG	Royal Botanic Gardens, Kew (2007)
psbA-F	GTTATGCATGAACGTAATGCTC	Sang *et al.* (1997)
trnH-R	CGCGCATGGTGGATTCACAAATC	Sang *et al.* (1997)
matK-2.1aF	ATCCATCTGGAAATCTTAGTTC	Royal Botanic Gardens, Kew (2007)
matK-5R	GTTCTAGCACAAGAAAGTCG	Royal Botanic Gardens, Kew (2007)
matK-1018R	GTACYACYGAAKRATYBAGYCSCAC	Zhang *et al.* (2006)
matK-390F	CGATCTATTCATTCAATATTTC	Cuénoud *et al.* (2002)
matK-1326R	TCTAGCACACGAAAGTCGAAGT	Cuénoud *et al.* (2002)

Sequencing was performed by Macrogen Inc. (Seoul, South Korea) on an ABI3730XL automated sequencer (Applied Biosystems). The same primers as used in PCR were used for the sequencing reactions (Table 1).

Trace files were aligned with the programs Gap4 and Pregap4 (Bonfield *et al.* 1995), both modules in the Staden package (Staden, 1996).

Identification process

The BLASTn algorithm as implemented in the stand-alone executables formatdb and blastall (Altschul *et al.* 1997) was used in this study to assess the identification of unknown sequences. The program formatdb (v. 2.2.18) was used for creating searchable databases from text files containing all reference sequences. For this purpose parameters were set to their default values, although it was specified that the database consisted of nucleotide sequences. Separate databases were made for *rpoC1* and *psbA-trnH* sequences, and barcode sequences from the market samples were subsequently queried against these databases only, using the blastall (v. 2.2.18) program. It was necessary to disable the "low-complexity filter" in order to avoid misleading interpretations of a run of 7-9 A's in the *rpoC1* gene of most samples. Otherwise, default parameters were generally used. The highest-scoring hit from each query on the basis of E-value and maximum identity was taken as the barcoding identification. If a query resulted in two or more equally high-scoring hits, the identification was considered „ambiguous".

Results

Reference database

Plant material from 33 herbarium specimens kept at the Reading University Herbarium (RNG), UK, was sampled for this study (see Appendix). Most of these specimens were collected in close range of the High Atlas Mountains, which is the main site of origin for medicinal plant products traded in Marrakech. The expanded reference database thus comprises 136 samples.

DNA extraction

There were a number of samples from which the extraction methods consistently failed to yield PCR products. In total these make up 24 of 111 market samples (20.7%; corresponding to 12 products) and 10 of 136 reference samples (7.4%).

PCR and sequencing success

Amplification of *matK* using all combinations of the primers above (Table 1) yielded PCR products for less than 30 % of the reference specimens. This was deemed insufficient for barcoding purposes, and a decision was subsequently made to exclude *matK* as a potential barcode in this study. PCR success rates for the other two barcoding loci (*psbA-trnH* and *rpoC1*) are summarized in Table 2. As can be seen, most roots were successfully sequenced for at least one of the two barcoding regions. The reference samples (which were extracted from herbarium vouchers) were consistently easier to sequence than the market samples.

Table 2. PCR/sequencing success.

Region	Market samples	Reference samples
psbA-trnH	71.3% (62 of 87 samples)	83.3% (105 of 126 samples)
rpoC1	83.9% (73 of 87 samples)	93.7% (118 of 126 samples)
psbA-trnH **and** *rpoC1*	62.1% (54 of 87 samples)	77.0% (97 of 126 samples)
psbA-trnH **or** *rpoC1*	93.1% (81 of 87 samples)	100% (126 of 126 samples)

Note: Only samples from which total DNA was successfully extracted are included here.

The new additions to the reference database are presented in the Appendix.

Barcoding identifications

The results from BLAST queries against the *rpoC1* and *psbA-trnH* subsets of the reference database are summarized in Table 3.

Table 3. Barcoding identifications.

Coll. no.	Vernacular name	Proposed scientific name (as based on Bellakhdar, 1997)	*rpoC1*			*psbA-trnH*		
			Barcoding identification	E-value	Max. id.	Barcoding identification	E-value	Max. id.
EM429	3oud amskhsr [highest]	*Polygonum aviculare (P. maritimum)*	MU19 *Daucus crinitus*	0	99.8%	MU19 *Daucus crinitus*	1E-131	99.6%
EM453	3oud amskhsr [highest]	*Polygonum aviculare (P. maritimum)*	MU19 *Daucus crinitus*	0	99.8%	MU19 *Daucus crinitus*	1E-130	99.6%
EM417	3oud amskhsr [secondary]	*Polygonum aviculare (P. maritimum)*	MU17 *Thapsia platycarpa* MU14 *Thapsia garganica* MU16 *Thapsia transtagana*	0	100.0%			
EM451	3oud amskhsr [secondary]	*Polygonum aviculare (P. maritimum)*	MU17 *Thapsia platycarpa* MU14 *Thapsia garganica* MU16 *Thapsia transtagana*	0	100.0%	MU16 *Thapsia transtagana*	1E-111	99.5%
EM449	3qrqrha [good]	*Anacyclus pyrethrum*	MU34 *Anacyclus homogamus* MU32 *Anacyclus radiatus* ssp. *radiatus* MU31 *Anacyclus pyrethrum* var. *depressus*	0	100.0%	MU32 *Anacyclus radiatus* ssp. *radiatus*	0	100.0%
EM408	3qrqrha [highest]	*Anacyclus pyrethrum*	MU34 *Anacyclus homogamus*	0	100.0%	MU32 *Anacyclus radiatus* ssp. *radiatus*	0	100.0%
EM444	3qrqrha [highest]	*Anacyclus pyrethrum*	MU34 *Anacyclus homogamus* MU32 *Anacyclus radiatus* ssp. *radiatus* MU31 *Anacyclus pyrethrum* var. *depressus*	0	100.0%	MU32 *Anacyclus radiatus* ssp. *radiatus*	0	100.0%
EM416	3qrqrha [secondary]	*Anacyclus pyrethrum*	MU39 *Catananche montana* MU38 *Catananche caerulea* MU37 *Catananche arenaria*	0	99.3%	MU37 *Catananche arenaria* MU38 *Catananche caerulea*	0	98.9%
EM450	3qrqrha [secondary]	*Anacyclus pyrethrum*	MU34 *Anacyclus homogamus* MU32 *Anacyclus radiatus* ssp. *radiatus* MU31 *Anacyclus pyrethrum* var. *depressus*	0	100.0%	MU32 *Anacyclus radiatus* ssp. *radiatus* MU34 *Anacyclus homogamus*	0	99.8%
EM371	3rouq frifra	*Magydaris panacifolia*	AR26 *Kundmannia sicula*	0	100.0%	AR26 *Kundmannia sicula*	7E-76	99.3%
EM412	3rouq frifra	*Magydaris panacifolia*	MU18 *Foeniculum vulgare*	0	99.6%	MU18 *Foeniculum vulgare*	6E-99	94.4%
EM365	3rouq lbsbas lbldi	*Foeniculum vulgare*	AR14 *Anethum foeniculoides* MU18 *Foeniculum vulgare*	0	100.0%			

EM387	3rouq lbsbas lbldi	*Foeniculum vulgare*	MU18 *Foeniculum vulgare* AR14 *Anethum foeniculoides*	0	100.0%	AR14 *Anethum foeniculoides*	1E-111	95.1%
EM366	3rouq lbsbas lboustani	*Foeniculum vulgare*	MU18 *Foeniculum vulgare* AR14 *Anethum foeniculoides*	0	100.0%	AR14 *Anethum foeniculoides*	1E-132	100.0%
EM372	3rouq lbsbas lboustani	*Foeniculum vulgare*	MU18 *Foeniculum vulgare* AR14 *Anethum foeniculoides*	0	100.0%	AR14 *Anethum foeniculoides*	1E-132	100.0%
EM402	3rouq lfijel	*Ruta montana (R. chalepensis)*	MU97 *Ruta montana*	0	100.0%			
EM439	3rouq lgsb	*Arundo donax*				MU90 *Arundo donax*	0	100.0%
EM443	3rouq lgsb	*Arundo donax*				MU90 *Arundo donax*	0	100.0%
EM442	3rouq lhrml	*Peganum harmala*	MU50 *Carlina brachylepis*	0	100.0%	MU50 *Carlina brachylepis*	0	100.0%
EM430	3rouq lkhrchouf	*Cynara cardunculus*	MU43 *Cynara humilis* MU42 *Cynara baetica* ssp. *maroccana*	0	99.8%	MU43 *Cynara humilis* MU42 *Cynara baetica* ssp. *maroccana*	0	100.0%
EM395	3rouq lkhrchouf rroumi	*Cynara cardunculus*	MU48 *Echinops spinosus* AR21 *Echinops fontqueri*	0	99.8%	MU48 *Echinops spinosus*	0	100.0%
EM388	3rouq njm lbori	*Cynodon dactylon*				MU92 *Cynodon dactylon*	0	99.3%
EM389	3rouq njm lmawi	*Cynodon dactylon*				MU91 *Arundo plinii* MU90 *Arundo donax*	0	98.6%
EM440	3rouq sdr	*Ziziphus lotus*	AR3 *Ziziphus lotus*	0	100.0%	AR3 *Ziziphus lotus*	0	100.0%
EM363	3rouq tlh dkr	*Acacia sp.*	AR19 *Acacia gummifera*	0	100.0%	AR19 *Acacia gummifera*	0	100.0%
EM364	3rouq tlh ntwa	*Acacia sp.*	AR19 *Acacia gummifera*	0	100.0%	AR19 *Acacia gummifera*	0	100.0%
EM399	3rq sous	*Glycyrrhiza glabra*	MU57 *Glycyrrhiza glabra*	0	100.0%	MU57 *Glycyrrhiza glabra*	0	98.3%
EM409	3rq sous lbldi	*Glycyrrhiza glabra*	MU57 *Glycyrrhiza glabra*	0	100.0%	MU57 *Glycyrrhiza glabra*	1E-179	98.2%
EM378	3rq sous lhchich	*Glycyrrhiza glabra*	MU57 *Glycyrrhiza glabra*	0	99.8%	MU57 *Glycyrrhiza glabra*	0	99.7%
EM373	3rq sous lqash	*Glycyrrhiza glabra*	MU57 *Glycyrrhiza glabra*	0	100.0%	MU57 *Glycyrrhiza glabra*	0	100.0%
EM437	Addad	*Atractylis gummifera*	AR2 *Carlina lanata* MU30 *Chamaeleon gummifer*	0	100.0%			
EM374	Addad bjlftou	*Atractylis gummifera*	AR2 *Carlina lanata* MU30 *Chamaeleon gummifer*	0	100.0%	MU30 *Chamaeleon gummifer*	0	100.0%
EM397	Addad dkr	*Atractylis gummifera*	AR2 *Carlina lanata* MU30 *Chamaeleon gummifer*	0	100.0%	MU30 *Chamaeleon gummifer*	0	100.0%
EM380	Addad lmjllaf	*Atractylis gummifera*				MU30 *Chamaeleon gummifer*	0	100.0%
EM396	Addad ntwa	*Atractylis gummifera*				AR6 *Rhaponticum acaule*	0	93.1%
EM404	Bid lghoul	*Mandragora autumnalis*	MU7 *Withania frutescens*	0	99.1%	MU107 *Mandragora officinarum*	0	100.0%
EM420	Blalouz	*Asphodelus microcarpus*	MU73 *Asparagus aphyllus*	0	97.3%			
EM436	Boughlam ssahraoui	*Spergularia marginata*				AR29 *Silene vulgaris*	7E-31	85.8%
EM377	Bougoudz	*????????*	AR16 *Tamus communis*	0	100.0%			

EM452	Bougoudz	*????????*	AR16 *Tamus communis*	0	100.0%			
OA1	Bougoudz	*????????*	AR16 *Tamus communis*	0	100.0%			
OA2	Bougoudz	*????????*	AR16 *Tamus communis*	0	100.0%			
OA4	Bougoudz	*????????*	AR16 *Tamus communis*	0	100.0%			
EM447	Bouzfour	*Daucus crinitus*	AR26 *Kundmannia sicula*	0	100.0%	AR26 *Kundmannia sicula*	8E-85	98.8%
EM405	Brztm	*Aristolochia longa (A. baetica)*	MU29 *Aristolochia pistolochia* MU28 *Aristolochia paucinervis* MU26 *Aristolochia longa*	0	100.0%			
EM414	Dryas	*Thapsia garganica*	MU17 *Thapsia platycarpa* MU14 *Thapsia garganica* MU16 *Thapsia transtagana*	0	100.0%			
OA3	Harmel	*Peganum harmala*	MU102 *Populus nigra*	0	94.6%	MU54 *Bryonia dioica*	2E-32	88.4%
EM435	L3chba	*Smilax aspera*	MU106 *Smilax aspera*	0	99.8%			
EM431	L3nsla	*Urginea maritima*	MU78 *Urginea maritima* var. *maura*	0	100.0%	MU78 *Urginea maritima* var. *maura*	0	99.8%
EM410	Lbkbouka	*Colchicum autumnale*	AR25 *Bunium bulbocastanum* AR4 *Bunium alpinum* ssp. *atlanticum* AR23 *Bunium pachypodum* AR24 *Bunium bulbocastanum*	0	100.0%	AR25 *Bunium bulbocastanum*	1E-90	98.3%
EM438	Lfouwwa	*Rubia peregrina (R. tinctorum)*	MU95 *Rubia peregrina*	0	100.0%			
EM379	Lfouwwa lfrouguiyya	*Rubia peregrina (R. tinctorum)*	MU95 *Rubia peregrina*	0	100.0%	MU96 *Rubia tinctorum*	1E-102	90.6%
EM390	Lfouwwa lfrouguiyya	*Rubia peregrina (R. tinctorum)*	MU95 *Rubia peregrina*	0	100.0%	MU96 *Rubia tinctorum*	1E-118	91.1%
EM391	Lfouwwa rqiqa (jbal nawahi mrrakch)	*Rubia peregrina (R. tinctorum)*	MU95 *Rubia peregrina*	0	99.1%	MU96 *Rubia tinctorum*	1E-60	85.3%
EM398	Lfwilia	*Astragalus lusitanicus*	MU59 *Astragalus lusitanicus*	0	100.0%			
EM432	Lghzghaz	*Carlina involucrata*	MU50 *Carlina brachylepis*	0	100.0%	MU50 *Carlina brachylepis*	0	100.0%
EM433	Lklkh	*Ferula communis*	AR7 *Ferula communis*	0	100.0%	AR7 *Ferula communis*	1E-136	100.0%
EM382	Lmghizla	*Eryngium triquetrum*	AR17 *Eryngium tricuspidatum*	0	100.0%	AR17 *Eryngium tricuspidatum*	0	99.8%
EM424	Lmghizla	*Eryngium triquetrum*	AR17 *Eryngium tricuspidatum*	0	100.0%	AR17 *Eryngium tricuspidatum*	0	99.3%
EM422	Ndkhir	*???????*	AR16 *Tamus communis*	0	100.0%			
EM427	Oudn lhllouf	*Pulicaria arabica*				MU44 *Pulicaria odora*	0	100.0%
EM403	Srghina	*Corrigiola telephiifolia*	AR9 *Corrigiola litoralis* ssp. *litoralis*	0	100.0%	AR9 *Corrigiola litoralis* ssp. *litoralis*	0	100.0%
EM368	Srghina l3adia	*Corrigiola telephiifolia*	AR29 *Silene vulgaris*	0	98.4%	AR29 *Silene vulgaris*	3E-30	85.8%
EM376	Srghina l3adia	*Corrigiola telephiifolia*	AR12 *Corrigiola telephiifolia*	0	100.0%	AR12 *Corrigiola telephiifolia*	1E-129	99.2%
EM367	Srghina lmsouwsa	*Corrigiola telephiifolia*	AR12 *Corrigiola telephiifolia*	0	100.0%	AR12 *Corrigiola telephiifolia*	1E-140	99.6%

EM421	Srghina lmsouwsa	*Corrigiola telephiifolia*	AR12 *Corrigiola telephiifolia*	0	100.0%	AR12 *Corrigiola telephiifolia*	1E-139	99.2%
EM423	Srghina rrahmania	*Corrigiola telephiifolia*	AR29 *Silene vulgaris*	0	96.0%	AR9 *Corrigiola litoralis* ssp. *litoralis*	2E-25	85.5%
EM446	Ssousban	*Iris germanica*	MU75 *Asparagus albus*	0	96.6%	MU75 *Asparagus albus*	0	95.7%
EM413	Tafgha	*Centaurea chamaerhaponticum*	AR33 *Carthamus pinnatus*	0	99.3%	AR6 *Rhaponticum acaule*	0	94.9%
OA10	Tafgha	*Centaurea chamaerhaponticum*	AR33 *Carthamus pinnatus*	0	99.3%	AR6 *Rhaponticum acaule*	0	94.8%
EM407	Taskra	*Echinops spinosus*	AR21 *Echinops fontqueri* MU48 *Echinops spinosus*	0	100.0%	MU48 *Echinops spinosus*	0	98.5%
EM356	Terta	*Withania frutescens*	AR26 *Kundmannia sicula*	0	100.0%	AR26 *Kundmannia sicula*	1E-99	98.9%
OA11	Terta	*Withania frutescens*	MU7 *Withania frutescens*	0	100.0%	MU107 *Mandragora officinarum*	1E-138	92.5%
OA8	Terta	*Withania frutescens*	MU7 *Withania frutescens*	0	100.0%	MU107 *Mandragora officinarum*	1E-138	92.5%
EM428	Tighcht	*Saponaria vaccaria*	AR29 *Silene vulgaris*	0	99.8%	AR29 *Silene vulgaris*	1E-152	96.2%
EM361	Tigndizt lghlida [highest]	*Anacyclus pyrethrum*	MU39 *Catananche montana* MU38 *Catananche caerulea* MU37 *Catananche arenaria*	0	99.6%	MU35 *Catananche caespitosa*	0	100.0%
EM362	Tigndizt rqiqa [secondary]	*Anacyclus pyrethrum*	MU37 *Catananche arenaria*	0	99.8%			
EM411	Tlh	*Acacia* sp.	AR19 *Acacia gummifera*	0	100.0%	AR19 *Acacia gummifera*	0	100.0%
EM425	Zziyata	*Limoniastrum guyonianum*	AR26 *Kundmannia sicula*	0	100.0%			

The highest scoring hits returned from BLASTn searches against the reference database (see Appendix) are taken to be the barcoding identification. Ambiguous results, i.e. multiple high-scoring hits (with the same E-value and maximum identity) are given where these occur. The herbalist"s quality rating of the product (where given) is indicated in square brackets following the vernacular name. Samples in blue were only sequenced and/or analyzed for one of the two barcoding regions. Maximum identity values below 99% for *rpoC1* and below 98% for *psbA-trnH*, as well as E-values above 1E-75 for the latter, are highlighted in red (see Discussion).

In general, sequence variation was lower among closely related species for the coding gene *rpoC1* than for the intergenic spacer region *psbA-trnH*. This resulted in a larger amount of ambiguous species identifications for *rpoC1* (samples receiving multiple equally high-scoring hits; 20 of 73 = 27.4% of all cases). In many cases, however, the corresponding *psbA-trnH* query is unambiguous in identifying the sample to species, e.g. EM361, EM366, EM410.

It is notable that for around 25% of the market samples, the barcoding identification differed significantly from the proposed scientific name in terms of evolutionary relatedness. For example, EM410 „Lbkbouka" („*Colchicum autumnale*", Colchicaceae) is unambiguously identified as a *Bunium* sp. (Apiaceae), and EM442 „3rouq lhrml" („*Peganum harmala*", Nitrariaceae) is identified as *Carlina brachylepis* (Asteraceae) with support from both barcoding regions.

A significant finding is the identification of the root products „Bougoudz" and „Ndkhir" as *Tamus communis* (Dioscoreaceae), with very high certainty for *rpoC1* (all six sequences were 100% identical to the reference sequence). Neither the species itself nor the local names were recorded in Bellakhdar (1997), or in Bellakhdar *et al*. (1991). However, *Tamus communis* was recorded in Boulos (1983) under other local names, and the name „bû-godz" was mentioned in Bellakhdar *et al.* (1982), there tentatively labeled as „*Dahlia variabilis*".

Discussion

DNA extraction

Some of the reference samples that could not be sequenced here are from plants that are known to contain secondary metabolites that inhibit DNA extraction (e.g. tannins in *Quercus*). It is likely that specialized protocols would aid in the process of extracting DNA from these (e.g. Csaikl *et al.* 1998). On the other hand, sequencing failure for many of the market samples may be due to post-harvest processing resulting in complete degradation of DNA, such as drying at high temperatures, slow drying under moist conditions or storage in alcohol. In these cases the only possible solution is to search for herbalists that supply the product in an unprocessed state. Considering that herbalists do not normally store plant products in a way conductive to the preservation of DNA, it is hardly surprising that extraction success rates in this study were lower for market samples than for reference samples.

The potential barcoding loci used in this study

ITS. Due mainly to its popularity in phylogenetic studies, ITS is by far the most widely sequenced locus for angiosperms (99,123 accessions in GenBank/EBI as of Sept. 2009, compared to 30,325 entries for *rbcL*, which is the most frequently sequenced plastid gene; see Chase *et al.* 2005). Arguably, this should make it the most suitable barcoding region if "quick and dirty" identifications are desired (e.g. in angiosperm-wide studies), as the chances of finding a high-scoring BLAST-hit are maximized.

The main reason to discredit ITS as an official barcode is its documented non-linear pattern of evolution in some groups of plants, whereby in extreme cases multiple divergent copies may occur within the same individual (Zheng *et al.* 2008). On the other hand, the presence of highly universal primers for the ITS region and its high evolutionary divergence rate (Kress *et al.* 2005) suggest that its use as a barcode should perhaps not be discredited entirely, at least for those groups in which "problematic" evolutionary patterns have not been observed.

In this study, the ITS region was used to provide preliminary identifications for samples that were not identified previously. In general this approach was highly successful: In most cases ITS-blasting against the online nr-database resulted in quite precise identifications and allowed for more accurate sampling of reference specimens than previously possible. Thus ITS contributed significantly to the barcoding identifications, even though it was not used as an actual barcode.

matK. Although attempts were made to amplify this region using several pairs of primers claimed by some to be universal in angiosperms (e.g. Lahaye *et al.* 2008), it was not possible to retrieve sequences for more than about 30% of attempted reference specimens. Furthermore, a certain primer pair did not always yield a PCR product in all members of a group of seemingly closely related taxa, indicating that the primer regions themselves are not conserved. As the presence of conserved flanking regions and universal primers is such a key asset for a molecular barcode (especially when dealing with material of completely unknown affinity), I decided not to use *matK* in the remainder of the study. Similarly low success rates have been reached by others (Kress & Erickson, 2007, Fazekas *et al.* 2008). Surprisingly, an

official proposal to adopt *matK* as a standard barcode for plants (in conjunction with *rbcL*) has recently been put forward by a leading group of plant barcoding researchers (CBOL Plant Working Group 2009; at the time of writing, this proposal was being reviewed by a specialist committee at CBOL). The conclusions drawn in that report are not supported by the initial attempts conducted in this study, although apparently better-performing primers were designed to support their proposal.

rpoC1. The main advantage of this chloroplast region is its very high amplification success rate, as confirmed here (94% of all reference samples were successfully sequenced) and in many other studies this locus typically scores the highest in this aspect (Sass et al. 2007). On the other hand *rpoC1* exhibits a slower rate of evolution compared to non-coding plastid regions and some plastid genes (e.g. *matK*; Newmaster *et al.* 2008). In this study, 27.4% of all root sequences received two or more equally high-scoring hits for *rpoC1* (see Table 3), due to these hits having identical sequences. Such cases would probably increase in frequency if the reference database were larger and contained more species in diverse genera. Evidently, interspecific variation for this locus is too low in some plant groups to allow for it to be used as a single barcode.

psbA-trnH. PCR success for this locus, although lower than that of *rpoC1*, was relatively high for reference sequences (83.3%) and moderate for root sequences (71.3%). Discriminatory power was substantially superior to that of *rpoC1* as only 6.6% of all samples received multiple highest-scoring hits. Thus in terms of these two factors the two loci seem to complement each other rather well.

Assembling the *psbA-trnH* trace files into contigs was not always straightforward. In many cases mononucleotide repeats (i.e. strings of 10 or more consecutive A‟s or T‟s) had induced Taq-polymerase errors, which made it difficult to accurately read the trace files. This resulted in a number of unreliable sequences that could not be used. It has been suggested that this feature of *psbA-trnH* and other non-coding regions prevent their use in future large-scale barcoding projects, in which manual editing of sequences is necessarily kept to a minimum (Devey *et al.* 2009).

Concluding remarks. It seems evident that *psbA-trnH* and *rpoC1* need to be combined, if one is to maximize success in terms of both discriminatory power and sequencing coverage. Overall, all reference samples and 93.1% of all market samples were successfully sequenced for at least one of the two loci (Table 2). The combined discriminatory power of *psbA-trnH*/*rpoC1* resulted in ambiguous identifications for only 7 out of 79 samples (8.9%). Whether or not these cases can be resolved through the addition of a third plastid locus, e.g. *rbcL* or perhaps *matK* (using the primers laid out by the CBOL Plant Working Group), is a possible topic for a future study.

Accuracy of the identifications

Currently, there are no standardized methods for comparing unknown sequences to reference sequences. The BLAST algorithm (Altschul *et al.* 1997) used here is in essence a local alignment tool, and was not specifically designed for barcoding. Nevertheless, it has repeatedly been used for this purpose in recent years (e.g. Ford *et al.* 2009), and comparisons

based on test datasets show that it does not perform worse than other methods (Little & Stevenson, 2007).

A disadvantage of BLAST is that there are no statistical methods that can give a measure of the accuracy of identifications (Munch *et al.* 2008). However, the E-value and maximum identity are two statistics that can be used as an informal measure of the likeliness of an identification being correct. In general, one can assume that the closer a hit approaches 100% in sequence identity (and an E-value of 0), the more likely it is to have been correctly identified to species as well. However, there is a possibility that hits scoring 100% in sequence identity may be incorrect, if there are closely related species in the target geographical area that were not included in the reference database. An extreme example of a group for which this is possible is *Silene* (Caryophyllaceae), with 70 species in Morocco (Fennane *et al.* 1999-2007), of which only 3 species were sampled here.

Constraints prevented the complete sampling of all groups/genera in this study. Nonetheless, the support for high-scoring hits being the correct identifications can be considered robust for smaller groups that were completely sampled, and especially for phylogenetically distinctive taxa. Such is the case for e.g. *Corrigiola* and *Tamus*.

A number of samples in this study received relatively low-scoring highest hits, indicating that they are incorrect. A clear distinction cannot objectively be made between "reliable" identifications and less reliable identifications. However, for the sake of discussion a maximum identity threshold of 99% for *rpoC1* and 98% for *psbA-trnH*[1] is posited. Hits below these thresholds (highlighted in blue in Table 3) have probably been incorrectly identified. In these cases the correct ID (or at least a close relative of the correct ID) was either sampled but not successfully sequenced, or was missed altogether when the reference database was created. Overall, these low-scoring highest hits account for a comparatively small amount of all identifications (especially when *rpoC1* and *psbA-trnH* identifications are collated), suggesting that the methods used here to extend the reference database were largely successful.

Ethnobotanical implications

The apparent discrepancy between the barcoding identifications and the vernacular names can probably largely be explained by the lack of a one-to-one correspondence between the vernacular names of plants (or plant products) and biological species. This phenomenon is a feature of virtually all folk classifications systems of living organisms (Cotton, 2002). A few example cases of this discrepancy and the potential underlying causes are given here.

Under-differentiation and product qualities. The results seem to indicate the presence of species complexes, i.e. a group of species to which the same vernacular name is applied, possibly because of taxonomic under-differentiation (i.e. failure to distinguish between closely-related species). In some instances, species affinity for a particular root sample seems to correlate with the "quality" assigned to the root product by the herbalist. The most clear-cut

[1] These rather conservative values are based on perceived rates of divergence between species of the same genus, and between closely related genera, for the two loci (cf. "mean percent sequence divergence" for the two regions in Kress & Erickson, 2007).

case is „3oud amskhsr", of which the samples designated as the highest in quality (EM429, EM453) are identified as *Daucus crinitus*, whereas those of secondary quality (EM417, EM451) are identified as *Thapsia* sp. The various types of „Srghina" (6 samples tested, see Table 3) constitute another species complex, although here it is less clear how the types actually relate to biological entities, if at all they do. In both of these examples, the fact that these species are treated as subtypes of the same vernacular name suggests that they are believed to share the same medicinal properties and are used to treat the same ailments.

Taxonomic ***over-differentiation*** also occurs in a few instances, i.e. where one biological species is referred to by several vernacular names. For example, EM371 „3rouq frifra", EM447 „Bouzfour", EM356 „Terta" and EM425 „Zziyata" were all identified as *Kundmannia sicula* (Apiaceae). The most common vernacular for this species is apparently „Zziyata", while „3rouq frifra" and „Bouzfour" usually refer to other members of the family (Bellakhdar, 1997). The latter two cases might therefore have resulted from a misidentification by the *collector*. „Terta", on the other hand normally applies to the unrelated *Withania frutescens* (Solanaceae; OA8, OA11), which in the wild is very unlikely to be confused for any of the other three species. Human error on the part of the *herbalist* (i.e. a mix-up of similar-looking root products) might explain this particular case.

Adulteration. The trade in medicinal plants provides the main source of income for herbalists, and economic constraints may provide incentive for herbalists to substitute rare ingredients with cheaper and more readily available species, misleadingly selling these under the same name. Such cases of deliberate adulteration of intended ingredients are difficult to distinguish from cases of genuine under- or over-differentiation; e.g. any of the cases mentioned in the previous sections could have resulted either inadvertently (by mistake or misidentification), or purposefully.

A sample of „Lbkbouka" (EM410), which translates as „*Colchicum autumnale*", provides a clearer example of suspected adulteration. This plant has traditionally been used to treat acute arthritis and renal disorders (Boulos, 1983) although Bellakhdar (1997) states that it is no longer traded in Morocco, owing to its extreme toxicity. Perhaps expectedly, barcoding analysis showed the vernacular name specified by the herbalist to be misleading. Instead the sample was identified as *Bunium* sp. (for which „Lbkbouka" does not apply), a plant with similar bulbous underground parts, but non-toxic and entirely unrelated to *Colchicum*. If Bellakhdar's claim that *Colchicum* is no longer used in the Moroccan pharmacopoeia is correct, then the usage of the name „Lbkbouka" is probably intentionally deceptive.

Other findings. Finally, the assignment of the local names „Ndkhir" and „Bougoudz" to *Tamus communis* represents new or previously unrecorded knowledge for the Moroccan pharmacopoeia. This "discovery" is further proof that DNA barcoding can provide the means for studying and identifying medicinal plant products that are unidentifiable by morphology alone.

Conclusions

The results presented in this report demonstrate the utility of DNA barcoding as a method for identifying ethnobotanical material. There is considerable potential for the use of barcoding in

applied cases, such as studying the interactions between market processes underlying the trade in medicinal plant products and the status of wild plant populations.

The methods tested and undertaken in this study can be summarized in a simple workflow. As previously stated, this approach to barcoding is especially suitable in cases for which knowledge of the material in need of identification is lacking, or suspected to be inaccurate. It assumes the need for researchers to compile their own reference databases, as the online availability of barcode sequences for angiosperms is as yet fragmentary and is likely to remain so in the near future. To a large extent, this approach also circumvents the current methodological limitations inherent to mainstream forms of plant barcoding (e.g. the lack of agreement upon which regions to use, and which algorithm to compare sequences with).

1. Collect the material in need of identification. (These are the „market samples".)
2. Extract total DNA from the market samples.
3. Amplify and sequence the ITS region for all market samples.
4. Run BLAST queries for the ITS sequences against online databases.
5. Compile a list of probable ID candidates for the samples, based on the results in the previous step.
6. Collect identified material of all listed candidates, either in the field or from herbarium specimens. (These are the „reference samples".)
7. Extract total DNA from the reference samples.
8. Amplify and sequence agreed-upon barcode loci for the reference samples, and for the market samples.
9. Use the program "formatdb" to create a database of all reference barcodes.
10. BLAST each of the unknown barcode sequences against the reference database, using "blastall". The highest-scoring hit from each query is taken as the barcoding identification.

Acknowledgements

I would like to thank my supervisors Anneleen Kool and Hugo de Boer for their support and firm optimism throughout the various stages of this project. I am also grateful to Abderrahim Ouarghidi, Gary Martin and Abdelaziz Abbad in Marrakech, Ronnie Rutherford and Stephen Jury in Reading, Nahid Heidari for guidance in the lab, Sanja Savić for the good advice, and many others who have helped me in one way or another during the course of this project.

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25: 3389-3402.

Bellakhdar, J. 1997. La pharmacopée marocaine traditionnelle. Ibis Press, Paris.

Bellakhdar, J., Honda, G. & Miki, W. 1982. Herb drugs and herbalists in the Maghrib. Studia Culturae Islamicae 19. Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo.

Bellakhdar, J., Claisse, R., Fleurentin, J. & Younos, C. 1991. Repertory of standard herbal drugs in the Moroccan pharmacopoea. Journal of Ethnopharmacology 35: 123-143.

Bonfield, J.K., Smith, K.F. & Staden, R. 1995. A new DNA sequence assembly program. Nucleic Acids Research 24: 4992-4999.

Boulos, L. 1983. Medicinal plants of North Africa. Reference Publications, Inc., Algonac, Michigan.

Burns, J.M., Janzen, D.H., Hajibabaei, M., Hallwachs, W. & Hebert, P.D.N. 2008. DNA barcodes and cryptic species of skipper butterflies in the genus *Perichares* in Area de Conservacion Guanacaste, Costa Rica. Proceedings of the National Academy of Sciences of the United States of America 105: 6350-6355.

CBOL Plant Working Group 2009. A DNA barcode for land plants. Proceedings of the National Academy of Sciences of the United States of America 106: 12794-12797.

Chase, M.W., Salamin, N., Wilkinson, M., Dunwell, J.M., Kesanakurthi, R.P., Haidar, N. & Savolainen, V. 2005. Land plants and DNA barcodes: short-term and long-term goals. Philosophical Transactions of the Royal Society 360: 1889-1895.

Consortium for the Barcode of Life. http://www.barcoding.si.edu/index.html. Accessed August 27, 2009.

Cotton, C.M. 2002. Ethnobotany: Principles and Applications. John Wiley & Sons, Chichester.

Csaikl, U.M., Bastian, H., Brettschneider, R., Gauch, S., Meir, A., Schauerte, M., Scholz, F., Sperisen, C., Vornam, B. & Ziegenhagen, B. 1998. Comparative Analysis of Different DNA

Extraction Protocols: A Fast, Universal Maxi-Preparation of High Quality Plant DNA for Genetic Evaluation and Phylogenetic Studies. Plant Molecular Biology Reporter 16: 69-86.

Cuénoud, P., Savolainen, V., Chatrou, L.W., Powel, M., Grayer, R.E.J. & Chase, M.W. 2002. Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid rbcL, atpB, and matK DNA sequences. American Journal of Botany 89: 132-144.

Devey, D.S., Chase, M.W. & Clarkson, J.J. 2009. A stuttering start to plant DNA barcoding: microsatellites present a previously overlooked problem in non-coding plastid regions. Taxon 58: 7-15.

Fazekas, A.J., Burgess, K.S., Kesanakurti, P.R., Graham, S.W., Newmaster, S.G., Husband, B.C., Percy, D.M., Hajibabaei, M., Barrett, S.C.H. 2008. Multiple Multilocus DNA Barcodes from the Plastid Genome Discriminate Plant Species Equally Well. PLoS ONE, doi:10.1371/journal.pone.0002802.

Fennane, M., Ibn Tattou, M., Mathez, J., Ouyahya, A. & El Oualidi, J. (eds.) 1999-2007. Flore pratique du Maroc. Institut Scientifique, Rabat.

Ford, C.S., Ayres, K.L., Toomey, N., Haider, N., Van Alphen Stahl, J., Kelly, L.J., Wikström, N., Hollingsworth, P.M., Duff, R.J., Hoot, S.B., Cowan, R.S., Chase, M.W. & Wilkinson, M.J. 2009. Selection of candidate coding DNA barcoding regions for use on land plants. Botanical Journal of the Linnean Society 159: 1-11.

Hebert, P.D.N., Cywinska, A., Ball, S.L. & deWaard, J.R. 2003. Biological identifications through DNA barcodes. Proceedings of the Royal Society B 270: 313-322.

Kerr, K.C.R., Stoeckle, M.Y., Dove, C.J., Weigt, L.A., Francis, C.M. & Hebert, P.D.N. 2007. Comprehensive DNA Barcode coverage of North American Birds. Molecular Ecology Notes 7: 535-543.

Kress, J.W., Wurdack, K.J., Zimmer, E.A., Weigt, L.A. & Janzen, D.H. 2005. Use of DNA barcodes to identify flowering plants. Proceedings of the National Academy of Sciences 102: 8369-8374.

Kress, J.W. & Erickson, D.L. 2007. A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. PLoS ONE, doi:10.1371/journal.pone.0000508.

Krüger, Å. 2008. DNA-Barcoding identification of medicinal roots from Morocco. MSc-thesis. Uppsala University.

Lahaye, R., van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., Maurin, O., Duthoit, S., Barraclough, T.G. & Savolainen, V. 2008. DNA barcoding the floras of biodiversity hotspots. Proceedings of the National Academy of Sciences 105: 2923-2928.

Little, D.P. & Stevenson, D.W. 2007. A comparison of algorithms for the identification of specimens using DNA barcodes: examples from gymnosperms. Cladistics 23: 1–21.

Munch, K., Boomsma, W., Huelsenbeck, J.P., Willerslev, E. & Nielsen, R. 2008. Statistical Assignment of DNA Sequences Using Bayesian Phylogenetics. Systematic Biology 57: 750-757.

Newmaster, S.G., Fazekas, A.J., Steeves, R.A.D. & Janovec, J. 2008. Testing candidate plant barcode regions in the Myristicaceae. Molecular Ecology Resources 8: 480-490.

Ragupathy, S., Newmaster, S.G., Murugesan, M. & Balasubramaniam, V. 2009. DNA barcoding discriminates a new cryptic grass species revealed in an ethnobotany study by the hill tribes of the Western Ghats in southern India. Molecular Ecology Resources 9 (Suppl. 1): 164-171.

Royal Botanic Gardens, Kew, 2007 onwards. DNA Barcoding. http://www.kew.org/barcoding/protocols.html. Accessed July 16, 2009.

Sang, T., Crawford, D.J. & Stuessy, T.F. 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). American Journal of Botany 84: 1120-1136.

Sass, C., Little, D.P., Stevenson, D.W., Specht, C.D. 2007. DNA Barcoding in the Cycadales: Testing the Potential of Proposed Barcoding Markers for Species Identification of Cycads. PLoS ONE, doi:10.1371/journal.pone.0001154.

Siddall, M.E., Trontelj, P., Utevsky, S.Y., Nkamany, M. & Macdonald, K.S. III 2007. Proceedings of the Royal Society B 274: 1481-1487.

Smith, M.A., Poyarkov, N.A. Jr. & Hebert, P.D.N. 2008. CO1 DNA barcoding amphibians: take the chance, meet the challenge. Molecular Ecology Resources 8: 235-246.

Smith, P.J., Mcveagh, S.M. & Steinke, D. 2008. DNA barcoding for the identification of smoked fish products. Journal of Fish Biology 72: 464-471.

Staden, R. 1996. The Staden sequence analysis package. Molecular biotechnology 5: 233-241.

Valentini, A., Miquel, C., Nawaz, M.A., Bellemain, E., Coissac, E., Pompanon, F., Gielly, L., Cruaud, C., Nascetti, G., Wincker, P., Swenson, J.E. & Taberlet, P. 2009. Molecular Ecology Resources 9: 51-60.

White, T.J., Bruns, T., Lee, S. & Taylor, J.W. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pp. 315-322 In: Innis, M.A., Gelfand, D.H., Sninsky, J.J. & White, T.J. (eds.), PCR Protocols: A Guide to Methods and Applications. Academic Press, New York.

Yoon, C.S., Glawe, D.A. & Shaw, P.D. 1991. A method for rapid small-scale preparation of fungal DNA. Mycologia 83: 835-838.

Zhang, L., Simmons, M.P., Kocyan, A. & Renner, S.S. 2006. Phylogeny of the Cucurbitales based on DNA sequences of nine loci from three genomes: Implications for morphological and sexual system evolution. Molecular Phylogenetics and Evolution 39: 305-322.

Zheng, X., Cai, D., Yao, L. & Teng, Y. 2008. Non-concerted ITS evolution, early origin and phylogenetic utility of ITS pseudogenes in *Pyrus*. Molecular Phylogenetics and Evolution 48: 892-903.

Appendix

Reference database

New collections to the reference database for Moroccan medicinal plants added in Oct. 2008. The fourth and fifth columns indicate sequencing success for the two barcode regions used in the study. For previous additions to the database see Krüger (2008).

Ref. no.	Scientific name	Collector, collection no.	*rpoC1*	*psbA-trnH*
AR1	*Scolymus hispanicus*	S.L. Jury 11671	X	X
AR2	*Carlina lanata*	S.L. Jury 15064	X	X
AR3	*Ziziphus lotus*	S.L. Jury 15149	X	X
AR4	*Bunium alpinum* ssp. *atlanticum*	OPTIMA ITER V 1064	X	X
AR5	*Magydaris panacifolia*	H.J.M. Bowen 8412	X	X
AR6	*Rhaponticum acaule*	S.L. Jury 15826		X
AR7	*Ferula communis*	S.L. Jury 16346	X	X
AR8	*Eryngium campestre*	S.L. Jury 17523	X	X
AR9	*Corrigiola litoralis* ssp. *litoralis*	S.L. Jury 18530	X	X
AR10	*Acacia tortilis* ssp. *raddiana*	S.L. Jury 19064	X	X
AR11	*Ornithogalum narbonense*	S.L. Jury 19343		
AR12	*Corrigiola telephiifolia*	S.L. Jury 19483	X	X
AR13	*Ammoides pusilla*	M. Ait Lafkih, M. A. Carine, S. L. Jury, F. J. Rumsey 245	X	X
AR14	*Anethum foeniculoides*	S.L. Jury & T.M. Upson 20571	X	X
AR15	*Elaeoselinum meioides*	S.L. Jury & T.M. Upson 20572	X	X
AR16	*Tamus communis*	C. Harrouni, S.L. Jury & T.M. Upson 20616	X	
AR17	*Eryngium tricuspidatum*	S.L. Jury & R. Shkwa 20881	X	X
AR18	*Eryngium triquetum*	J.A. Mejías & **S. Silvestre** 308	X	
AR19	*Acacia gummifera*	Aparaicio, Rowe & Silvestre	X	X
AR20	*Carlina involucrata*	P. Kennedy	X	
AR21	*Echinops fontqueri*	J. Lambinon & G. van den Sande 94/Ma/454	X	
AR22	*Armeria simplex*	E. Bayón, C. Oberprieler, R. Vogt		
AR23	*Bunium pachypodum*	M.A. Mateos & **J.M. Montserrat** 5835/3	X	X
AR24	*Bunium incrassatum*	Davis 52337	X	X
AR25	*Bunium bulbocastanum*	J.M. Montserrat & B. Valdés	X	X
AR26	*Kundmannia sicula*	T. Abdelkader, J. Molero, **J.M. Montserrat**, J. Pallàs, J. Vicens & M. Veny 3604/5	X	X
AR27	*Eryngium ilicifolium*	M.A. Mateos & **B. Valdés** 641/93	X	
AR28	*Quercus rotundifolia*	A. Achhal, F. Bombardó & **Romo** 6761/4		
AR29	*Silene vulgaris*	M.A. Mateos, E. Ramos & J. Villarreal 5726/95	X	X
AR30	*Silene portensis*	M.A. Mateos & **J.M. Montserrat** 5651/4	X	X
AR31	*Asphodelus macrocarpus* ssp. *rubescens*	M.A. Mateos & **J.M. Montserrat** 5869/2		
AR32	*Silene filipetala*	Reading Univ./BM. Exped. 547	X	X
AR33	*Carthamus pinnatus*	A. Kool 1018	X	X