UPPSALA
UNIVERSITET

DNA-Barcoding identification of medicinal roots from Morocco

Åsa Krüger

Degree project in biology, Master of science (1 year), 2008
Examensarbete i biologi 30 hp till magisterexamen, 2008
Biology Education Centre
Supervisors: Hugo de Boer and Anneleen Kool

Abstract

Barcoding is a recent DNA based identification method. The purpose of barcoding is to identify unknown species with small fragments of DNA. This will enable extraction and comparison of DNA from all stages of life and even degraded DNA. This approach has already shown success in identification of animal species. To apply this method on identification of plants alteration of genes is necessary due to different inheritance patterns. There are different proposals of suitable gene sequences and there is a probability that more than one locus are needed for a reliable determination. The sequence suitability can alter between families since their evolution progress varies. In this study the regions *matK, psbA/trnH* and *rpoC1* were used. The regions have shown different suitability as universal barcode regions, *rpoC1* primers have shown capability of amplifying sequences from 86% of the voucher specimens, the *psbA/trnH* primers 40% but the *matK* primers only amplified 18% successfully. This result indicates that *rpoC1* might be suitable as a universal barcode region. The regions need to be further analyzed for sequence variability, there is a possibility that two or more regions combined could give a proper determination.

Keywords: DNA barcoding, *matK*, *psbA/trnH*, *rpoC1*.

Acknowledgments

I give my sincerest thanks to my supervisors Hugo de Boer and Anneleen Kool for giving me the opportunity to work with this interesting project. I would also like to thank Gary Martin, Mohamed El Haouzi and everyone involved in the project in Morocco. Furthermore, I give many thanks to Steven Jury and Ronald Rutherford from the Reading herbarium for sharing their expertise on the Moroccan flora. Thanks also to Nahid Heidari and everybody in the lab at Systematic Botany for your help. Niklas Wikström thank you for helping me with the Staden package. Many thanks to my friends and family for all your support during this process. Finally, I would like to thank Henrik Svengren for valuable comments.

Uppsala, May 2008,
Åsa Krüger

Contents

1 Introduction **1**
1.1 Background 1
1.2 Reference sequence database 2
1.2.1 Basic Local Alignment Search Tools (BLAST) 3
1.3 The Moroccan medicinal roots 3
1.4 The objective of the thesis 4

2 Material and Methods **5**
2.1 Material collecting 5
2.1.1 Reference specimens 5
2.1.2 Moroccan specimens 5
2.2 Extraction protocol 6
2.2.1 Procedure 6
2.3 PCR Protocol 7
2.3.1 Material 7
2.3.2 PCR cycle 8
2.4 Sequencing and submission to Genbank 8

3 Results **9**
3.1 Reference library 9
3.2 Moroccan roots 9
3.2.1 The *Polygonum* group 9
3.2.2 The *Anacyclus* group 10
3.2.3 The *Catananche* group 10
3.2.4 The unknown species group 10
3.2.5 The selection of interesting roots group 10

4 Discussion **12**
4.1 Reference library 12
4.2 Identification of the Moroccan roots 13
4.2.1 The *Polygonum* group 13
4.2.2 The *Anacyclus* group 14
4.2.3 The *Catananche* group 14
4.2.4 The unknown roots group 14
4.2.5 The selection of interesting roots group 15
4.3 Method evaluation 15

A Tables **20**

List of Tables

1 Reference library 21
2 The *Polygonum* group 24
3 The *Anacyclus* group 25
4 The *Catananche* group 26
5 The unknown species group 27
6 The selection of interesting roots group 28

1 Introduction

The introductory chapter will give an overview of the theory and objectives of this study.

1.1 Background

Plants are one of the most important resources on earth and have been used for food, medicine, building material and much more throughout history. The diversity of plants is enormous and its width unknown, so far. It is of great importance to identify and separate species from each other, this will provide information about their evolution, discover "new" species and see which role they play in the ecosystem and societies. It can be of interest on a global level, to get a picture of the diversity of life on our planet, especially since the global environment is in a period of rapid change and many species are threatened with extinction.

The identification method of plants has changed throughout history. In the beginning this knowledge was important for survival, e.g. to identify what is poisonous or not. Such knowledge was passed down from generation to generation. In western societies this knowledge developed into a systematic subject; *Taxonomy*. Consequently, a large part of the knowledge is isolated to a small group of people, the taxonomists [1]. In the global society of today this information needs to be accessible and used in cross-scientific research to be part of global diversity projects. The identification of large quantities of plant specimens, as would be the case in a global diversity project, would be a very time consuming task. The molecular identification, and barcoding in particular, can contribute as a supporting method to make the identification process more effective, especially with specimens lacking distinct characters.

DNA Barcoding identification uses small fragments of DNA to tell different species apart. This is a fairly new approach in the aspect of plant identification and a standardized method has not yet been agreed on [2]. The method is under construction and is currently at a fast expanding stage, new barcode sequences are submitted to Genbank every week [3]. The advantages of barcoding have already been seen in projects of e.g. birds and fishes. The gene primarily used for these projects is a mitochondrial gene *cox*1 [3], [4], [5], [6], [7]. This is however not a suitable choice for barcoding in plants due to the

fact that the mitochondrial DNA in plants have not evolved at the same pace as in animaliae, and plants have different inheritance patterns [8], [9]. The challenge with plant barcoding is to find a sequence that is unique for each species and at the same time flanked by conserved regions which will make it possible to construct a universal primer [4]. Another criteria which also need to be fulfilled for a successful barcoding sequence; it should be a short sequence so that it can be used for identification of degraded DNA and it should also be present in all different cell types in the plant [10]. Possibly, two or more loci are needed for a reliable barcode identification of plants [5], [11]. The idea of a plant barcoding system has proven to be a delicate matter since several taxonomist has expressed their critique against the idea. The main issues being; that no such barcode has yet been found and that the method could in the end reduce the importance of morphological identification [12], [13],[14].

1.2 Reference sequence database

To be able to use DNA barcoding for identification of unknown species a reference database is required. An obstacle with using a new DNA identification method without a standardized protocol is that there is no reference database available, thus one has to be made before the identification project can begin. The database consists of voucher material of herbarium specimens, where identification has been based on morphology by skilled taxonomists. A DNA sample will be taken from the herbarium specimens and the different barcoding primers will be used for amplifying a specific region of DNA that will be sequenced. For all species likely to be present, a voucher needs to be sequenced. Other species within the same genus should also be sequenced to be able to detect any mix-up between species. The unknown material will be sequenced using the same primers and these sequences will be run against the database of voucher specimens. When the unknown material is sequenced it will be run against the existing database, now containing the reference material. By doing so there is no need to create a specific database with search engines and algorithms for this specific project. Genbank is used because this project is significantly limited by time and creating this kind of database would probably take a great deal of that time. This study will not aim at creating the actual database but to provide its content.

1.2.1 Basic Local Alignment Search Tools (BLAST)

There are several algorithms used in different databases to calculate the score, or the level of success, of a match based on points added or removed depending on base replacements, gaps or indels. In this study an existing database called Genbank is used which employs the Basic Local Alignment Search Tools (BLAST). BLAST is a number of algorithms that makes local alignments of the sequence with the records in the database. The scores of these alignments depend on how similar the sequences are and what causes eventual mismatches. The e-value which is generated for each match is an indication of the significance of the similarity, the lower the e-value the higher significance to the match. Another important value that is calculated with each hit in Genbank is the percentage of identity. This percentage refers to how invariant the two aligned sequences are, and thereby gives a estimation on how similar they are. [15], [16]

1.3 The Moroccan medicinal roots

The markets of Marrakech hold many herbalists, or herbal vendors as they will be referred to in this text, whom sells plants amongst other things. Many of these plants are traditionally used for medicinal purposes and many people still use them in this way. The plants are sold by their Arabic or Berber name and sometimes one name refers to many different species, differentiated by e.g. quality. It is unclear what the division of qualities correspond to, are they different species used for the same purpose, or old and new material from the same species? Could it be that the original species is getting rarer and more difficult to find, and that replacement species are sold in the same name but referred to as a lower quality of the species asked for? There is also a question of what is sold on the market, in terms of possible toxic plants and endangered species. In some cases it is impossible to identify the specimen bought in the market by morphological methods because it have been grounded into a powder or only a part of the plant is sold which lacks distinct character to make a positive identification. This study is focused on roots sold on the markets of Marrakech. The roots have been chosen because they are often sold without the flowering part which makes them hard to identify by traditional methods. They are often dried and there are several occasions when they are sold by different qualities.

1.4 The objective of the thesis

The objective of this study is to create a reference database for identification of medicinal plants in Morocco by DNA barcoding. The results will indicate if barcoding and the primers used, *matK, rpoC1, psbA/trnH*, is a good alternative for identification of the roots collected for this study.

2 Material and Methods

This chapter will present the material and methods used in this study.

2.1 Material collecting

2.1.1 Reference specimens

The crucial step of this study is gathering the voucher material used to create a reliable database. The material was collected from the herbarium of Reading University. The reasons for this choice are that the Reading herbarium holds a recent collection from Morocco and are mostly collected and identified by, skilled taxonomists, Steven L. Jury and Ronald W. Rutherford. Both of these factors are important. Firstly, with age the quality of DNA decreases and it will be harder to extract and get a good sequencing of the DNA. Secondly, the importance of the collector and identifier; in this case both Jury and Rutherford have great experience of Moroccan plants and their knowledge is valuable to assure that the samples in the database are correctly identified. When sampling the voucher species of the herbaria the specimens are picked based on how recent they are, if they are in a good condition, the collector of the specimen and if the sampling will affect the voucher specimen.

The species included in the reference collection was based on a free list of species sold on the market [17]. From this list species were chosen based on the level of difficulty to identify, occurrence and species which are suspects of mix-up. To this list species in the same genus which are present in the North-west African area was added [18], [19].

2.1.2 Moroccan specimens

When the database is in use, material from the markets of Marrakech will be run against the sequences from the voucher specimens in an attempt to identify them. The market samples were bought from the herbalists by asking for the root by their Arabic or Berber name [20]. If two or more roots have been offered for the same name all different roots have been purchased. The roots were bought from several herbal vendors to diminish sample bias.

2.2 Extraction protocol

Barcoding of plants does not have any standardized protocol since it is such a new method. The procedure used in this study is based on previous studies done on this subject [5], [21], [22], [23].

The extraction protocol is not specific for barcoding and could be used when extracting DNA for other purposes, such as cloning, PCR and regular sequencing.

2.2.1 Procedure

The extraction procedure is based on the Carlson/Yoon extraction and performed in the following manner [24]. The incubator was set to 60 ℃. The screw-cap tubes were filled with silica beads to a half, and then a small amount, approximately 0.02 g, of dry plant material was added. To start the extraction a volume of 750 μl of Carlson buffer and 20 μl mercaptoethanol were mixed with the samples, which then were placed in the incubator for 10 minutes. After the incubation the tubes were directly put in the Mini-beadbeater *3110BX*. Each run in the beadbeater was 40 seconds long at a speed of 5 000 rpm. If the material was not completely homogenized, the run was repeated. The tubes were incubated for 30-60 minutes at 60 ℃ and mixed a couple of times during this period. After 60 minutes had passed 750 μl chloroform/isoamylalcohol 24:1 were added to the samples. Then the tubes were placed horizontally on a shaker for 30 minutes at low speed, around 100 rpm. To separate the different phases the sample were centrifuged in a table centrifuge for 10 minutes at 10 000 rpm. After the separation the water phase was transferred to a 1.5 ml eppendorf tube. Another 1 volume of chloroform/isoamylalcohol was added and the samples were centrifuged for 5 minutes. The water phase was once again transferred to a new eppendorf tube. The DNA was precipitated with 0.1 volume of 3M NaAc, pH4.6 and 2 volumes of 95% EtOH and incubated at -20 ℃ over night.

After the incubation the samples were centrifuged for 10 minutes at 10 000 rpm to pellet the DNA. After the centrifugation the liquid was poured off and 750 μl 70% EtOH was used to wash the pellet by gentle mixing and centrifugation for 5 minutes. The liquid was carefully removed and the pellet was dissolved in 100 μl of EB buffer. The DNA was purified with the

QIAquick DNA and PCR-purification protocol.

The kit consisted of GFX columns where the DNA was captured by capture buffer, then washed with wash buffer and finally eluted using EB buffer.

2.3 PCR Protocol

2.3.1 Material

Amount	Component
14.625 μl	Autoclaved water
5 μl	Buffer (10x)
3 μl	$MgCl_2$
1 μl	dNTPs (10 μM)
0.125 μl	BSA
12.5 μl	Forward primer (2 μM)
12.5 μl	Reverse primer (2 μM)
0.25 μl	Taq
1 μl	Template DNA

The primers used in the different reactions were;

matK 2.1.a F; 5' ATCCATCTGGAAATCTTAGTTC 3'
matK 5 R; 5' GTTCTAGCACAAGAAAGTCG 3'

psbA; 5' GTTATGCATGAACGTAATGCTC 3'
trnH; 5' CGCGCATGGTGGATTCACAATCC 3'

rpoC1 2 F ; 5' GGCAAAGAGGGAAGATTTCG 3'
rpoC1 4 R ; 5' CCATAAGCATATCTTGAGTTGG 3'

When mixing the PCR mixture the components were kept cool on ice during the entire procedure to prevent a premature start of the reaction. The components were mixed thoroughly, so the PCR tubes would get the same composition of ingredients.

2.3.2 PCR cycle

Temperature	Time	Number of cycles
94℃	1 min	1 cycle
94℃	30 s	38 cycles
53℃	40 s	38 cycles
72℃	40 s	38 cycles
72℃	5 min	1 cycle
8℃	∞	

2.4 Sequencing and submission to Genbank

Once the sequences were amplified and purified they were shipped to Macrogen[1], South Korea, for sequencing. The same primers which were used in the PCR amplification was provided for the sequencing. The reference sequences were sequenced with both primers, the root samples were sequenced with a single primer. Once the sequences were obtained they were edited both automatically and manually with the programs Staden package and FinchTV [25], [26]. The reference sequences were submitted to Genbank by the Sequin program, version 7.90 [27]. The submissions contain name of the species, collector, voucher number, region of sequence and the information that the sequence is a barcode. The sequences from the market samples will be run against the database through a Basic Local Alignment Search Tool (BLAST) [15].

[1]See http://www.macrogen.com/eng/macrogen/macrogen_main.jsp.

3 Results

In this chapter the results from the sequencing and identification will be presented.

3.1 Reference library

The voucher specimens submitted to Genbank are presented with collector information and Genbank accession numbers in Table 1, The reference library.

[Insert Table 1 about here]

The table contains information of the specimen's name and accession number in Genbank for the three different regions. With the *matK* region 18% of the voucher specimens was successfully amplified and sequenced. For the *psbA/trnH* the number was 40%. Finally, the *rpoC1* region had processed 86% of the voucher specimens effectively.

3.2 Moroccan roots

The Moroccan roots have been divided into different groups and the results of the BLAST searches are presented below. The division of the groups are based on their putative scientific name, given based on literature [20] and translation. The tables show the Arabic name, the scientific name, the labID, the Genbank result, E-value and percentage of identity. The later two are as discussed in the introduction a measure on how dependable the match is.

3.2.1 The *Polygonum* group

In Table 2 the results from the BLAST search of the *Polygonum* samples bought on the market are presented. The three regions all indicate that sample MU417 would be a *Thapsia sp.*, both *matK* and *rpoC1* got hits for *Thapsia platycarpa* while *psbA/trnH* resulted in *Thapsia villosa*. The remaining species of the group only got results from the *rpoC1* region was the seconde qualite got hits on *Thapsia sp.* and the premiére qualite showed the result *Daucus crintus*.

[Insert Table 2 about here]

3.2.2 The *Anacyclus* group

Table 3 shows the results for the *Anacyclus* group. The results of the three regions are consistent in samples MU361 and MU362 were they all have *Catananche sp.* as hits. Furthermore both *psbA/trnH* and *rpoC1* point out *Anacyclus homogamus* as the result for MU448.

[Insert Table 3 about here]

3.2.3 The *Catananche* group

The *rpoC1* region shown very various results of the *Catananche* group, as can be seen in table 4. The *psbA/trnH* region had two sequenced samples, MU358 and MU357. Only one of these sequences got a hit in Genbank with the result *Armeria atlantica*. The *rpoc1* region generated products in all samples but only one, MU384, had a *Catananche sp.* as a BLAST result. The other hits were, *Limonium thiniense, Medicago truncatula, Foeniculum vulgare* and *Musa acuminata*. We were unable to amplify the *matK* region for this group.

[Insert Table 4 about here]

3.2.4 The unknown species group

In table 5 the three samples of which no putative Latin name was available. A consensus is shown in the results of MU377, both *psbA/trnH* and *rpoC1* got hits for *Dioscorea* species. The remaining result from the *psbA* region were a hit for sample MU422 on *Sorghum bicolor cultivar*. With the *rpoC1* region all three unknown samples got the result *Dioscorea elephantipes*. In this group we were unable to amplify the *matK* region.

[Insert Table 5 about here]

3.2.5 The selection of interesting roots group

The Genbank results generated from the *matK* region, six out of the fifteen samples of this group are presented in table 6. There were hits on *Rubia tinctorum* in three of the cases were the samples were sold as Lfouwwa Ifrouguiyya, putative scientific name; *Rubia sp.* Furthermore, *matK* had a

hit on *Mandragora* and *Aristolochia* where these were the species asked for. These two hits were on records in Genbank which have been submitted by other authors. Finally, the *matK* generated a match for *Thapsia platycarpa* were the sample has been sold as Zziyata, putative scientific name; *Limoniastrum sp.*.

In the case of *psbA/trnH* only three sequences resulted in Genbank hits, even though seven samples had been successfully sequenced. The three hits were a *Zanthoxylum capense* for a sample sold as *Ruta sp.*, a *Monimia ovalifolia* for a sample sold as a*Aristolochia sp.* and a *Thapsia transtagana* for the sample sold as a *Asparagus sp.*

The last region *rpoC1* generated Genbank results in all cases but one were Genbank could find any results for the sequence BLASTed. In all four samples sold as *Rubia sp.* the results from *rpoC1* in Genbank shows hits on *Rubia sp.* It also shows a hit on *Rubia sp.* when the saple was sold as a *Plumbago europaea.* There were three other hits that concurred with the name the sample was sold by, MU402 *Ruta sp.*, MU405 *Aristolochia sp.* and MU435 Smilax aspera. The remaining samples did not get hits on the species name by which they were sold. A sample bought as a *Mandragora sp.* got a hit on *Polygonum maritimum*, the *Asparargus sp.* sample got a hit on *Medicago truncatula* and so did the a sample bought as *Arundo donax.* The other *Aroundo donax* sample got a hit on *Verbascum dentifolium.* The result *Foeniculum vulgare* was generated for both the sample sold as *Limoniastrum guyonianum* and the one sold as *Daucus crinitus.*

[Insert Table 6 about here]

4 Discussion

The discussion will include an evaluation of the method in general and the suitability for this type of project. Further the discussion will focus on the three different regions used and if the are good candidates as a DNA barcoding regions.

4.1 Reference library

In the construction of the reference library problems were experienced with the PCR amplification of the samples with the *matK* primers, only 18 % of the samples were amplified. Similar problems have been reported in previous studies [5], [9]. This result indicates that the *matK* primers 2.1a and 5 are not suitable as a universal primers in this kind of project. Previous studies discuss the possibility to design the primer for projects in certain families [9]. This is not applicable in this case because this study includes many different families and the unknown plant material could be in anyone of these families. It would be much too time consuming and economically inefficient to proceed in such a manner. However, the *matK* did produce some sequences to the library and have been used while studying the different root problem groups.

The reference library also contains the *psbA/trnH* primers. This is an intergenic spacer and many researchers consider this to be one of the best sequences to use so far [5]. This region showed moderate capability of functioning as an universal DNA barcoding region in this study, in terms of PCR amplification. It gave product in 40% of the cases.

The *rpoC1* showed the best amplification success of the three regions. This is a very important aspect in terms of a possible barcode candidate. This is because the intent of barcoding is to be able to identify unknown plant material in the field. It is therefore important that the barcode used can amplify all plant families. It would be a great inconvenience if the barcode primer needed to be redesigned for every family. This would in the end result in that all unknown plant material would have to be tested with many different primers.

Based on the different success rates in amplification among the regions the most suitable region to work as a barcode region is *rpoC1* since it generated sequences from 86% of the voucher samples. Another important aspect to

take into account when discussing the capability of a primer to be a good candidate as a barcode region is the variability in the sequence. It does not matter if the primer can amplify all families if the region does not vary between species. In terms of suitability as a universal barcode primer only *rpoC1* have shown to be a possible candidate, in terms of amplification success, in this study.

A possible solution to the problem, as has been proposed earlier, is a two or more loci based barcode identification. This would give the opportunity to combine and improve the identification by combining the results.

4.2 Identification of the Moroccan roots

When looking at the root samples the same pattern of amplification capability as in the reference library is shown; *matK* as the region with lowest number of amplified samples and *rpoC1* as the region with the greatest amplification success. The results from this study support the discussion about the need for *matK* primers to be designed to fit a certain family. The same goes for *psbA/trnH* and the opposite is true for *rpoC1*.

4.2.1 The *Polygonum* group

In the *Polygonum* case there is no hit for *Polygonum* in any of the BLAST searches. The *rpoC1* primers identified the first quality *Polygonum* samples as *Daucus crinitus* in both cases and as *Thapsia platycarpa* when it is considered second quality. There is a consensus amongst the regions in the result of MU417, they all get hits on *Thapsia* species. However, even though all three regions indicate a *Thapsia* sp. this cannot be considered a true identification. This is due to that DNA barcoding is a new method and the database does not contain many *matK*, *psbA/trnH* and *rpoC1* sequences. This results in fewer species to match an alignment to and identification results from BLAST need to be put in their right context. In the case of *Polygonum* we can consider none of the samples to be an actual *Polygonum* sp. because voucher specimens have been submitted and would match with other sequences of *Polygonum*, but we can only suggest an alternate identification. It would be of interest to further investigate these samples. A possible way to go about this is to sequence the samples with ITS. The ITS region is a widely used nuclear sequence which has many records in Genbank and could give a hint

on where to look for possible families and genus for positive identification [11], [21].

4.2.2 The *Anacyclus* group

In the *Anacyclus* group there is a consensus in the results of MU361 and MU362 as a *Catananche* sp. Since both *psbA/trnH* and *rpoC1* have reference sequences of *Anacyclus*, see Table 1 in the Appendix, the conclusion that these two samples are not *Anacyclus* can be drawn. As previous discussed about the *Polygonum* group, it is problematic to state an alternative identification by these sequences until more barcode sequences has been submitted. The results of the other samples indicate that the samples really could be *Anacylus* sp. since they have good matches with the reference sequences. In this case the result of *Rubia tinctorum* from the *matK* region is not taken under consideration since *matK* were unable to amplify any of the *Anacyclus* voucher specimens.

4.2.3 The *Catananche* group

In the *Catananche* group only one hit is on *Catananche* and the other resulted in hits on the following species; *Limonium, Medicago, Foeniculum* and *Musa*. This shows a great variation in results and this implies that what is sold as *Catananche* is in fact a variation of different species and those are probably not *Catananche* species. This could be a result of an ambiguos name which could affect the communication between consumer and herbal vendor or herbal vendor and supplier. Other possible reasons for this confusion could be that it was not the head herbal vendor that was selling the root but a vendor in learning. It could be so that the price and availability of *Catananche* affects the market. The last explanation might be less likely if further studies on the *Anacyclus* samples show that the suggested *Catananche* identification is correct.

4.2.4 The unknown roots group

The group which is completely unknown got *Dioscorea* as a BLAST result both with *psbA/trnH* and *rpoC1*. This result is less reliable in this study, this is due to the lack of voucher specimens. These results have been derived from other sequences submitted to Genbank. It may very well be *Dioscorea* but it could also be so that this was the best match of present sequences and

that it is in fact a species that is not sequenced and submitted to Genbank yet. This is an interesting group to keep investigate since they showed no hits on the reference material submitted for this study. The ITS region would be a good candidate for this sequencing which have been discussed above.

4.2.5 The selection of interesting roots group

In the larger group of several different families the aim was to discover possible mix-ups. When examining the results one could see that the *Rubia sp.*, *Mandragora sp.*, *Smilax aspera* and *Aristolochia sp.* have been positively identified. These results were mostly generated from the records submitted from the reference library of this study but a few of the hits were on records submitted by others. Some of the results on the other hand strongly indicates that the samples sold under a certain name is in fact another species. An example of this is the *Daucus crinitus* sequenced by the *rpoC1* region. There is a voucher specimen of *Daucus crinitus* and other *Daucus* sp. submitted to Genbank from the reference library and if the sample in fact were a *Daucus sp.* it would get a hit on one of these sequences. As mention earlier the results from these BLAST searches can only indicate that the sample is sold under the wrong name or not but lacks the capacity of give a reliable suggestion for an alternate identification.

4.3 Method evaluation

In the aspect of testing suitability of DNA barcode regions the method was able to indicate differences among the sequences. An improvement could have been to create a database solely for this purpose with other algorithms and the function of an alignment as in for example clustalW. To be able to look at the variability in the different sequences to see if the regions were unique enough to differ between species in the same genus. A clustalW alignment of the sequences has not been done in this study because there are too few genus groups that have been successfully sequenced. The result of such few sequences would not be of significance. Until a barcode database has been made that meet these criteria and holds many more barcode regions this method is not suitable for a collection of a vast variation of families as this study, which also contained completely unknown material. When creating such a database one could also see to that the submission step was simplified from the Sequin program that both EMBL and Genbank use. This program

is not suitable for large submissions and has a lot of additional information which is not needed for this type of study. The submission part of the project was a bottleneck and delayed the study. Overall the largest shortcoming of the method is the lack of a standardized protocol and fixed regions, it has great potential once these have been determined and a DNA barcoding database for plants exists.

References

[1] Quentin D. Wheeler, Peter H. Raven, Edward O. Wilson. 2004. *Taxonomy: Impediment or Expedient?*. Science 16;303, no. 5656, p. 285

[2] Blaxter, Mark L. 2004. *The promise of a DNA taxonomy*. Phil. Trans. Royal sociaety London. B, 359; p. 669-679

[3] Barcoding of life data systems (BOLD systems), URL: http://www.barcodinglife.org. 2007-09-10

[4] Consortium for the Barcode of life (CBOL), URL: http://www.barcoding.si.edu/. 2007-09-10

[5] Kress, John and Erickson David. 2007. *A two-locus global DNA barcode for land plants: The coding rbcL gene complements the non-coding trnH-psbA spacer region*. PLos ONE,6; e508

[6] Mehrdad Hajibabaei, Daniel H. Janzen, John M. Burns, Winnie Hallwachs and Paul D. N. Hebert. 2005. *DNA barcodes distinguish species of tropical Lepidoptera*. PNAS 103; no. 4, p. 968-971

[7] Paul D. N. Hebert, Erin H. Penton, John M. Burns, Daniel H. Janzen and Winnie Hallwachs. *Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator*. PNAS 11; no 41, p. 14812-14817

[8] Hughes, Monica A. 1996. *Plant Molecular Genetics*. Addison Wesley Longman Limited, Printed in Great Britain by Henry Ling Limited, at Dorset Press, Dorchester, DT1 1HD

[9] Newmaster, S.G., Fazekas, A.J., Steeves, R.A.D and Janovec, J. 2007. *Testing candidate plant barcode regions in the Myristicaceae*. Molecular Ecology Notes doi:10.1111/j.1471-8286.2007.02002.x

[10] Chase, Mark W., Cowan, Robyn S, Hollingsworth, Peter M., van den Berg, Cassio, Madriñán, Santiago, Petersen, Gitte, Seberg, Ole, Jørgsensen, Tina, Cameron, Kenneth M., Carine, Mark, Pedersen, Niklas, Hedderson, Terry A.J., Conrad, Ferozah, Salazar, Gerardo A., Richardson, James E., Hollingsworth, Michelle L., Barraclough, Timothy G., Kelly, Laura, Wilkinson, Mike. 2007. *A proposal for a standardised protocol to barcode all land plants*. TAXON 56(2), May, p. 295-299

[11] Chase Mark W., Salamin Nicolas, Wilkinson Mike, Dunwell James M., Kesanakurthi Rao Prasad, Haidar Nadia, Savolainen Vincent. 2005. *Land plants and DNA barcodes: short-term and long-term goals.* Phil. Trans. R. Soc. B, 360, p. 1889-1895

[12] Daniel Rubinoff, Stephen Cameron and Kipling Will 2006; *Are plant DNA barcodes a search for the Holy Grail.* TRENDS in Ecology and Evolution 21; no. 1

[13] Moritz, Craig and Cicero, Carla. 2004. *DNA Barcoding: Promise and Pitfalls.* PLos Biology 2; issue 10, p. 1529-15231

[14] Will, Kipling and Rubinoff, Daniel. 2004. *Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification.* Cladistics 20; p. 47-55

[15] NCBI homepage for nucleotide BLAST,
URL: http://www.ncbi.nlm.nih.gov/blast/Blast.cgi. 2007/2008

[16] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. *Basic Local Alignment Search Tool.* J. Mol. Biol. 215, p. 403-410

[17] El Haouzi, Mohamed. Muséum d'Histoire Naturelle de Marrakech. Personal communication.

[18] Valdes,B., Rejdali, M., Achhal El Kadmiri, A., Jury, S.L., Montserrat, J.M. *Catalogue des plates vasculares du nord du Maroc, incluant des clés d'ídentification.* Universidad de Sevilla, Institut agronomique et vétérrinarie Hassan II, Rabat, University of Reading, Institut botánic de Barcelona. CSIC-Ajuntament de Barcelona, Volume I and II, 2002 Madrid.

[19] IUCN The World Conservation Union. 2005. *A Guide to Medicinal Plants in North Africa.* Gráficas La Paz, Torredonjimeno, Jaén (Spain) ISBN: 2-8317-0893-1

[20] Bellakhdar, Jamal. 1997. *La pharmacopée marocaine traditionnelle, Médecine arabe ancienne et savoirs populaires.* Ibis Press, ISBN: 2-910728-03-x

[21] W. John Kress, Kenneth J. Wurdack, Elizabeth A. Zimmer, Lee A. Weigt and Daniel H. Janzen. 2005. *Use of DNA barcodes to identify flowering plants*, PNAS, 102, no. 23; p. 8369-8374

[22] Vincent Savolainen, Robyn S Cowan, Alfried P Vogler, George K Roderick, and Richard Lane. 2005. *Towards writing the encyclopaedia of life: an introduction to DNA barcoding.* Phil. Trans. R. Soc. B , 360, p. 1805-1811

[23] Royal botanical gardens, Kew. Phase 2 protocols,
URL: http://www.kew.org/barcoding/protocols.html. 2007-09-04

[24] Carlson, JE, Tulsieram LK, Glaubitz JC, Luk VWK, Kauffeldt C and Rutledge R. 1991. *Segregation of random amplified DNA markers in F1 progeny of conifers.* Theor Appl Genet 83: 194-200

[25] The Staden package homepage for Staden package, version 1.6.0,
URL: http://staden.sourceforge.net/news.2005.html#staden160. 2007-12-02

[26] Geospiza homepage for FinchTV version 1.4.0.,
URL: http://www.geospiza.com/finchtv.html. 2007-11-20

[27] Sequin version 7.90, NCBI homepage for the Sequin program,
URL: http://www.ncbi.nlm.nih.gov/Sequin/index.html. 2008-01-20

A Tables

The appendix contain all tables in this report. The reference library is presented in Table 1, containing the voucher sequences submitted to Genbank. Tables 2, 3, 4, 5 and 6 show the results from the Genbank searches of the root samples from Marrakech.

Table 1: Reference library

LabID	Genus	Species	Collector	Collnr	matK	rpoC1	psbA
Mu33	Anacyclus	clavatus (Desf.) Pers.	*Montserrat, J.M.*	2043/94		EU531594	
Mu34	Anacyclus	homogamus	*Jury, S.L. and Shkwa, R.*	20904		EU531595	EU531693
Mu31	Anacyclus	pyrethrum L.	*Jury, S.L.*	17869		EU531592	EU531691
Mu32	Anacyclus	radiatus Loisel.	*Jury, S.L.*	18550		EU531593	EU531692
Mu27	Aristolochia	baetica L.	*Montserrat, J.M and Valdés, B.*	2002	EU531669	EU531588	
Mu26	Aristolochia	longa L.	*Davis*	51947	EU531668	EU531587	
Mu28	Aristolochia	paucinervis Pomel	*Ait Lafkih, M. et. al*	242	EU531670	EU531589	
Mu29	Aristolochia	pistolochia L.	*Jury, S.L.*	16506	EU531671	EU531590	
Mu88	Armeria	atlantica Pomel	*Jury, S.L.*	18964		EU531638	EU531718
Mu89	Armeria	choulettiana Pomel	*Jury, S.L. and Shkwa, R.*	20912		EU531639	
Mu90	Arundo	donax L.	*Jury, S.L.*	17806		EU531640	EU531719
Mu91	Arundo	plinii Turra	*Husain*	158			EU531720
Mu74	Asparagus	acutifolius L.	*OPTIMA ITER V*	1250		EU531626	
Mu75	Asparagus	albus L.	*Jury, S.L.*	12557		EU531627	EU531712
Mu71	Asparagus	altissimus Munby	*Jury, S.L. and Upson, T.M.*	20506	EU531679	EU531623	
Mu73	Asparagus	aphyllus L.	*OPTIMA ITER V*	14		EU531625	
Mu72	Asparagus	horridus L.	*Bombardo, F. et.al*	4565		EU531624	
Mu60	Astragalus	armatus Willd.	*Romo, A.*	R-8883/5		EU531616	
Mu59	Astragalus	lusitanicus Lam.	*Jury, S.L. and Shkwa, R.*	20966		EU531615	
Mu54	Bryonia	dioica Jacq.	*Jury, S.L. and Upson, T.M.*	20437			EU531705
Mu50	Carlina	brachylepis	*Griffiths, A.J.K*	18069	EU531677	EU531608	EU531703
Mu37	Catananche	arenaria L.	*Jury, S.L. and Upson, T.M.*	20689	EU531674	EU531598	EU531695
Mu38	Catananche	caerulea L.	*Jury, S.L.*	11396	EU531675		EU531696
Mu35	Catananche	caespitosa Desf.	*Jury, S.L.*	18111	EU531672	EU531596	
Mu36	Catananche	lutea L.	*Vogt, R.*	12348	EU531673	EU531597	EU531694
Mu39	Catananche	montana	*Mateos, M.A.*	7039/95	EU531676		
Mu30	Chamaeleon	gummifer (L.) Cass.	*Jury, S.L.*	12348		EU531591	
Mu51	Chamaeleon	macrophyllus (Desf.) Petit	*Jury, S.L.*	11909	EU531678	EU531609	EU531704
Mu875	Corrigola	sp				EU531661	
Mu42	Cynara	baetica (Spreng.) Pau	*Jury, S.L.*	11233		EU531601	EU531698
Mu43	Cynara	humilis L.	*Jury, S.L.*	19316		EU531602	EU531699
Mu92	Cynodon	dactylon (L.)Pers.	*Mateos, M.A. et. Al*	7291		EU531641	EU531721
Mu55	Cyperus	longus L.	*Jury, S.L.*	11493		EU531611	

Table 1 (continued)

LabID	Genus	Species	Collector	Collno.	matK	rpoC1	psbA/trnH
Mu56	Cyperus	rotundus L.	*Jury, S.L.*	12712		EU531612	
Mu23	Daucus	aureus Desf.	*Reading Univ./B. M. Exped.*	1076		EU531584	EU531688
Mu21	Daucus	carota L.	*Jury, S.L.*	17848		EU531582	EU531686
Mu19	Daucus	crinitus	*Ait Lafkih, M. et. al*	70		EU531580	EU531684
Mu22	Daucus	durieua Lange	*Díez, M.J.*	3482/94		EU531583	EU531687
Mu24	Daucus	muricatus (L.) L.	*Jury, S.L.*	16748		EU531585	EU531689
Mu20	Daucus	setifolius Desf.	*Jury, S.L.*	17514		EU531581	EU531685
Mu48	Echinops	spinosus L.	*Jury, S.L.*	18328		EU531606	
Mu93	Elymus	repens (L.) Gould	*Jury, S.L.*	17625		EU531642	EU531722
Mu18	Foeniculum	vulgare Miller	*Jury, S. L.*	17822	EU531667	EU531579	
Mu79	Fraxinus	angustifolia Vahl	*Jury, S.L.*	15547		EU531631	
Mu58	Glycyrrhiza	foetida L.	*Jury, S.L.*	14940		EU531614	EU531706
Mu57	Glycyrrhiza	glabra L.	*Pedrol, J.*	4493		EU531613	
Mu115	Haplophyllum	broussonetianum Coss.	*Jury, S.L.*	14466		EU531658	
Mu49	Inula	montana L.	*Brummitt, R.K.*	18721		EU531607	
Mu67	Juglans	regia L.	*Mateos, M.A.*	7147/95		EU531621	
Mu68	Juncus	fontanesii Gay	*Jury, S.L.*	18967		EU531622	
Mu41	Launaea	arborescens (Batt.) Murb.	*Jury, S.L.*	12745		EU531600	
Mu87	Limoniastrum	feei (de Gir.) Batt.	*Jury, S.L.*	19168	EU531681	EU531637	EU531717
Mu86	Limoniastrum	monopetalum	*Mateos, M.A.*	4825/95		EU531636	
Mu107	Mandragora	officinarum L.	*Jury, S.L.*	12238		EU531651	
Mu25	Nerium	oleander L.	*Mateos, M.A*	7113		EU531586	EU531690
Mu80	Olea	europaea L.	*Stephens, L.*	12		EU531632	
Mu61	Ononis	natrix L.	*Jury, S.L. and Upson, T.M.*	20696		EU531617	EU531707
Mu113	Peganum	harmala L.	*OPTIMA ITER V*	124		EU531656	
Mu82	Pinus	halepensis Mill.	*Carus, A.J et. al*	5266		EU531663	EU531714
Mu83	Pinus	nigra W.Arnold	*Jury, S.L.*	12487		EU531634	EU531715
Mu84	Plumbago	europaea L.	*OPTIMA ITER V*	1026	EU531680	EU531635	EU531716
Mu94	Polygonum	maritimum L.	*Jury, S.L. and H. Rankou*	19958		EU531643	
Mu101	Populus	alba L.	*Ait Lafkih, M.*	713		EU531647	
Mu103	Populus	euphratica Olver	*Jury, S.L.*	17803		EU531648	
Mu102	Populus	nigra L.	*Jury, S.L.*	18124		EU531662	
Mu44	Pulicaria	odora (L.) Rchb.	*Ait Lafkih, M. et. al*	110			EU531700

Table 1 (continued)

LabID	Genus	Species	Collector	Collno.	matK	rpoC1	psbA/trnH
Mu46	Pulicaria	mauritanica Coss.	*Jury, S.L. and Upson, T.M.*	20667		EU531604	EU531702
Mu45	Pulicaria	paludosa Link	*Jury, S.L.*	17773		EU531603	EU531701
Mu47	Pulicaria	undulata (L.) C.A.Mey.	*Jury, S.L.*	19074		EU531605	
Mu62	Retama	monosperma (L.) Boiss.	*Jury, S.L. and Upson, T.M.*	20671		EU531618	EU531708
Mu63	Retama	raetam	*Jury, S. L.*	15876		EU531619	EU531709
MU95	Rubia	peregrina	*Jury, S.L.*	14983		EU531644	
MU96	Rubia	tinctorum	*Jury, S.L.*	19809	EU531682		EU531723
Mu100	Ruta	angustifolia Pers.	*Jury, S.L. and Shkwa, R.*	20963		EU531646	
Mu98	Ruta	chalepensis L.	*Davis*	49757		EU531645	
Mu53	Saponaria	glutinosa M.Bieb.	*Jury, S.L.*	17900		EU531610	
Mu40	Scorzonera	pseudopygmaea Lipsky	*Ait Lafkih, M. et. al*	26		EU531599	EU531697
Mu81	Sesamum	indicum L.	*Nesbitt, R.M.*	1939		EU531633	EU531713
Mu106	Smilax	aspera L.	*Jury, S.L.*	15315		EU531650	
Mu109	Tamarix	africana Poir.	*Jury, S.L.*	16673		EU531652	
Mu112	Tamarix	amplexicaulis Ehrenb.	*Jury, S.L.*	19081		EU531655	EU531725
Mu111	Tamarix	canariensis Willd.	*OPTIMA ITER V*	1765		EU531654	EU531724
Mu110	Tamarix	gallica L.	*Mateos, M.A.*	4301/94		EU531653	
Mu14	Thapsia	garganica L.	*Kennedy, P.*	s.n.		EU531575	
Mu17	Thapsia	platycarpa Pomel	*Jury, S. L.*	15837	EU531666	EU531578	
Mu16	Thapsia	transtagana Brot.	*Jury, S. L.*	16325	EU531665	EU531577	
Mu15	Thapsia	villosa L.	*Lambinon, J.*	94/Ma/350	EU531664	EU531576	EU531683
Mu64	Trigonella	foenum-graecum	*Jury, S.L.*	13715		EU531620	EU531710
Mu65	Trigonella	gladiata M.Bieb.	*OPTIMA ITER V*	459			EU531711
Mu116	Urginea	fugax (Moris) Steinh.	*Ait Lafkih, M.*	19		EU531659	
Mu76	Urginea	maritima (L.) Baker	*Jury, S.L.*	12554		EU531628	
Mu78	Urginea	maritima (L.) Baker	*Jury, S.L.*	9030		EU531630	
Mu77	Urginea	undulata (Desf.) Maire	*Jury, S.L.*	13275		EU531629	
Mu105	Verbascum	dentifolium Delile	*Ait Lafkih, M. et. al*	104		EU531649	
MuA7	Whitania	frutesense	*Bengt Oxelman*			EU531660	
Mu114	Zygophyllum	fontanesii	*S.L Jury and Upson, T.M.*	20493		EU531657	

Comment: The vocher specimens sequenced with *matK*,*rpoC1* and *psbA/trnh* primers are presentet with Genbank accession numbers, collector, collector no. and LabID.
The *matK* region generated 18% successful submissions of the vocher specimens to Genbank, the *rpoC1* 86% and psbA/trnH 40%.

Table 2: The *Polygonum* group

Arabic name	**Scientific name**	**LabId**	**Genbank result**	**BLAST score**	**% ID**
			matK		
3oud amskhsr (seconde qualité)	*P. aviculare/maritimum*	MU417	Thapsia platycarpa EU531666	$6e^{-}94$	89%
3oud amskhsr (première qualité)	*P. aviculare/maritimum*	MU429			
3oud amskhsr (seconde qualité)	*P. aviculare/maritimum*	MU451			
3oud amskhsr (première qualité)	*P. aviculare/maritimum*	MU453	No result from Genbank		
			psbA/trnH		
3oud amskhsr (seconde qualité)	*P. aviculare/maritimum*	MU417	Thapsia villosa EU531683	$1e^{-}121$	95%
3oud amskhsr (première qualité)	*P. aviculare/maritimum*	MU429			
3oud amskhsr (seconde qualité)	*P. aviculare/maritimum*	MU451			
3oud amskhsr (première qualité)	*P. aviculare/maritimum*	MU453			
			rpoC1		
3oud amskhsr (seconde qualité)	*P. aviculare/maritimum**	MU417	Thapsia platycarpa EU531578	0.0	99%
3oud amskhsr (première qualité)	*P. aviculare/maritimum**	MU429	Daucus crinitus EU531580	0.0	99%
3oud amskhsr (seconde qualité)	*P. aviculare/maritimum**	MU451	Thapsia garganica EU531575	0.0	99%
3oud amskhsr (première qualité)	*P. aviculare/maritimum**	MU453	Daucus crinitus EU531580	0.0	99%

*Voucher specimen in reference library, see Table 1.
Comment: Blank fields indicate that the sample could not be amplified and therefore not sequenced.

Table 3: The *Anacyclus* group

Arabic name	**Scientific name**	**LabId**	**Genbank result**	**BLAST score**	**% ID**
			matK		
3qrqrha (mauvaise qualité)	*A. pyrethrum*	MU448			
3qrqrha (bonne qualité)	*A. pyrethrum*	MU449			
3qrqrha (seconde qualité)	*A. pyrethrum*	MU450			
3qrqrha (première qualité)	*A. pyrethrum*	MU444			
3qrqrha (seconde qualité)	*A. pyrethrum*	MU416	Rubia tinctorum EU531682	$4e^{-}115$	90%
Tigndizt lghlida (première qualité)	*A. pyrethrum*	MU361	Catananche caespitosa EU531672	0.0	100%
Tigndizt rqiqa (seconde qualité)	*A. pyrethrum*	MU362	Catananche arenaria EU531674	0.0	99%
			psbA/trnH		
3qrqrha (mauvaise qualité)	*A. pyrethrum**	MU448	Anacyclus homogamus EU531693	0.0	99%
3qrqrha (bonne qualité)	*A. pyrethrum**	MU449			
3qrqrha (seconde qualité)	*A. pyrethrum**	MU450			
3qrqrha (première qualité)	*A. pyrethrum**	MU444			
3qrqrha (seconde qualité)	*A. pyrethrum**	MU416	Catananche arenaria EU531695	0.0	98%
Tigndizt lghlida (première qualité)	*A. pyrethrum**	MU361	Catananche arenaria EU531695	0.0	98%
Tigndizt rqiqa (seconde qualité)	*A. pyrethrum**	MU362	Catananche arenaria EU531695	$5e^{-}152$	88%
			rpoC1		
3qrqrha (mauvaise qualité)	*A. pyrethrum**	MU448	Anacyclus homogamus EU531595	0.0	98%
3qrqrha (bonne qualité)	*A. pyrethrum**	MU449	Anacyclus homogamus EU531595	0.0	99%
3qrqrha (seconde qualité)	*A. pyrethrum**	MU450	Anacyclus homogamus EU531595	0.0	98%
3qrqrha (première qualité)	*A. pyrethrum**	MU444	Anacyclus homogamus EU531595	0.0	99%
3qrqrha (seconde qualité)	*A. pyrethrum**	MU416	Catananche arenaria EU531598	0.0	94%
Tigndizt lghlida (première qualité)	*A. pyrethrum**	MU361	Catananche arenaria EU531598	0.0	98%
Tigndizt rqiqa (seconde qualité)	*A. pyrethrum**	MU362	Catananche arenaria EU531598	0.0	99%

* Voucher specimen in reference library, see Table 1.
Comment: Blank fields indicate that the sample could not be amplified and therefore not sequenced.

Table 4: The *Catananche* group

Arabic name	**Scientific name**	**LabId**	**Genbank result**	**BLAST score**	**% ID**
			matK		
Awdmi rroumi	*C. caespitosa**	MU358			
Awdmi lbldi	*C. caespitosa**	MU357			
3rouq awdmi (première qualité)	*C. caespitosa**	MU381			
Awdmi dial jbal nawahi mrrakch	*C. caespitosa**	MU383			
3rouq awdmi dial chamal wa khenifra	*C. caespitosa**	MU384			
			psbA/trnH		
Awdmi rroumi	*C. caespitosa*	MU358	Armeria atlantica EU531718	$2e^{-}114$	89%
Awdmi lbldi	*C. caespitosa*	MU357	No results in Genbank		
3rouq awdmi (première qualité)	*C. caespitosa*	MU381			
Awdmi dial jbal nawahi mrrakch	*C. caespitosa*	MU383			
3rouq awdmi dial chamal wa khenifra	*C. caespitosa*	MU384			
			rpoC1		
Awdmi rroumi	*C. caespitosa**	MU358	Limonium thiniense AM889900	0.0	98%
Awdmi lbldi	*C. caespitosa**	MU357	Medicago truncatula CU570967	0.0	98%
3rouq awdmi (première qualité)	*C. caespitosa**	MU381	Foeniculum vulgare EU531579	0.0	96%
Awdmi dial jbal nawahi mrrakch	*C. caespitosa**	MU383	Musa acuminata EU017018	$2e^{-}32$	82%
3rouq awdmi dial chamal wa khenifra	*C. caespitosa**	MU384	Catananche caespitosa EU531596	$2e^{-}72$	79%

* Voucher specimen in reference library, see Table 1
Comment: Blank fields indicate that the sample could not be amplified and therefore not sequenced.

Table 5: The unknown species group

Arabic name	**Scientific name**	**LabId**	**Genbank result**	**BLAST score**	**% ID**
			matK		
Bougoudz	*?*	MU452			
Ndkhir	*?*	MU422			
Bougoudz	*?*	MU377			
			psbA/trnH		
Bougoudz	*?*	MU452			
Ndkhir	*?*	MU422	Sorghum bicolor cultivar EF115542	$1e^{-}174$	88%
Bougoudz	*?*	MU377	Dioscorea alata AB331306	$2e^{-}129$	90%
			rpoC1		
Bougoudz	*?*	MU452	Dioscorea elephantipes EF380353	0.0	98%
Ndkhir	*?*	MU422	Dioscorea elephantipes EF380353	0.0	98%
Bougoudz	*?*	MU377	Dioscorea elephantipes EF380353	0.0	98%

Comment: Blank fields indicate that the sample could not be amplified and therefore not sequenced.

Table 6: The selection of interesting roots group

Arabic name	Scientific name	LabId	Genbank result	BLAST score	% ID
			matK		
Lfouwwa lfrouguiyya	*Rubia peregrina/tinctorum**	MU379	Rubia tinctorum EU531682	0.0	98%
Lfouwwa lfrouguiyya	*Rubia peregrina/tinctorum* *	MU390	Rubia tinctorum EU531682	0.0	99%
Lfouwwa rqiqa	*Rubia peregrina/tinctorum* *	MU391			
Swak rra3yan	*Plumbago europaea**	MU400			
3rouq lfijel	*Ruta montana/chalepensis*	MU402			
Bid lghoul	*Mandragora autumnalis*	MU404	Mandragora officinarum AJ585883	0.0	99%
Brztm	*Aristolochia longa/baetica**	MU405	Aristolochia paucinervis DQ296662	0.0	100%
Dryas	*Thapsia garganica*	MU414			
Skkoum	*Asparagus sp.**	MU419			
Zziyata	*Limoniastrum guyonianum*	MU425	Thapsia platycarpa EU531666	$6e^{-}94$	89%
L3chba	*Smilax aspera*	MU435			
Lfouwwa	*Rubia peregrina/tinctoria**	MU438	Rubia tinctorum EU531682	0.0	98%
3rouq lgsb	*Arundo donax*	MU439			
3rouq lgsb	*Arundo donax*	MU443			
Bouzfour	*Daucus crinitus*	MU447			
			psbA/trnH		
Lfouwwa lfrouguiyya	*Rubia peregrina/tinctoria**	MU379			
Lfouwwa lfrouguiyya	*Rubia peregrina/tinctoria**	MU390			
Lfouwwa rqiqa	*Rubia peregrina/tinctoria**	MU391			
Swak rra3yan	*Plumbago europaea**	MU400			
3rouq lfijel	*Ruta montana/chalepensis*	MU402	Zanthoxylum capense AM500915	$2e^{-}26$	88%
Bid lghoul	*Mandragora autumnalis*	MU404			
Brztm	*Aristolochia longa/baetica*	MU405	Monimia ovalifolia AF129065	$1e^{-}28$	87%
Dryas	*Thapsia garganica*	MU414	No result in Genbank		
Skkoum	*Asparagus sp.**	MU419	Thapsia transtagana EU531577	0.0	95%
Zziyata	*Limoniastrum guyonianum*	MU425	No result in Genbank		
L3chba	*Smilax aspera*	MU435	No result in Genbank		
Lfouwwa	*Rubia peregrina/tinctoria**	MU438	No result in Genbank		
3rouq lgsb	*Arundo donax**	MU439			
3rouq lgsb	*Arundo donax**	MU443			
Bouzfour	*Daucus crinitus**	MU447			

* Voucher specimen in reference library, see Table 1.
Comment: Blank fields indicate that the sample could not be amplified and therefore not sequenced.

Table 6 (continued)

Arabic name	**Scientific name**	**LabId**	**Genbank result**	**BLAST score**	**% ID**
			rpoC1		
Lfouwwa lfrouguiyya	*Rubia peregrina/tinctoria**	MU379	Rubia peregrina EU531644	0.0	99%
Lfouwwa lfrouguiyya	*Rubia peregrina/tinctoria**	MU390	Rubia peregrina EU531644	0.0	98%
Lfouwwa rqiqa	*Rubia peregrina/tinctoria**	MU391	Rubia peregrina EU531644	0.0	97%
Swak rra3yan	*Plumbago europaea**	MU400	Rubia peregrina EU531644	0.0	96%
3rouq lfijel	*Ruta montana/chalepensis**	MU402	Ruta chalepensis EU531645	0.0	99%
Bid lghoul	*Mandragora autumnalis**	MU404	Polygonum maritimum EU531643	0.0	99%
Brztm	*Aristolochia longa/baetica**	MU405	Aristolochia pistolochia EU531590	0.0	99%
Dryas	*Thapsia garganica**	MU414	No result in Genbank		
Skkoum	*Asparagus sp.**	MU419	M.truncatula CU570967	0.0	96%
Zziyata	*Limoniastrum guyonianum*	MU425	Foeniculum vulgare EU531579	0.0	98%
L3chba	*Smilax aspera**	MU435	Smilax aspera EU531650	0.0	99%
Lfouwwa	*Rubia peregrina/tinctoria**	MU438	Rubia peregrina EU531644	0.0	96%
3rouq lgsb	*Arundo donax**	MU439	M.truncatula CU570967	0.0	95%
3rouq lgsb	*Arundo donax**	MU443	Verbascum dentifolium EU531649	0.0	95%
Bouzfour	*Daucus crinitus**	MU447	Foeniculum vulgare EU531579	0.0	98%

*Voucher specimen in reference library, see Table 1.