CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Tropical AgriSciences

The use of DNA barcoding methods for identification of tropical trees

MASTER'S THESIS

Prague 2020

Elaborated by: Bc. Barbora Legezová

Supervised by: doc. Ing. Bohdan Lojka, Ph.D.

Co-supervisor: Ing. Marie Kalousová, Ing. Anna Maňourová

Declaration

I hereby declare that I have done this thesis entitled ‘The use of DNA barcoding methods for identification of tropical trees’ independently, all texts in this thesis are original, and all the sources have been quoted and acknowledged by means of complete references and according to Citation rules of the FTA.

In Prague 14.8.2020

.....

Barbora Legezová

Acknowledgements

Firstly, I would like to thank my supervisor doc. Ing. Bohdan Lojka, Ph.D. for his overall help, patience and guidance of my thesis and especially for enabling me to conduct my research in Cameroon. I am very grateful to my co-supervisor Ing. Marie Kalousová for her help not only with the extensive laboratory work and for her overall support. Special thanks belong to my other co-supervisor Ing. Anna Maňourová for helping me with the collection of samples, her suggestions and corrections. I am also very grateful that she guided me through Cameroon and even though the field work was not easy and lot of unexpected complications occurred she showed me the beauty of this African country. My thanks also belong to Mgr. Petr Stiblík for always helping me and guiding me through the problematic of DNA barcoding.

Big thanks must go to my Cameroonian family, the Tchana family. Not only did they accommodate us, but they also accepted us to their family and made us feel at home. I also have to thank to my family and my partner for being patient and supportive, all through my research and writing the thesis. Furthermore, I wish to extend my gratitude to my best friend and my language support, Valentýna Foytlová, B.A., who was available whenever I could not find the right words.

This research could not be conducted without the generous financial support of the foundation "Nadání Josefa, Marie a Zdeňky Hlávkových", mobility support under the auspices of Faculty of Tropical AgriSciences, CZU and by World Agroforestry Centre (ICRAF).

Abstract

DNA barcoding allows fast, accurate, automatable and relatively cheap species identification by using a short, standardized DNA region. This method has proven effective mainly among animal species, however, the method is quickly evolving to be suitable for plant species as well. DNA barcoding of tropical trees is a very specific case, due to the often-inaccessible canopy, there is a need to barcode other plant tissues than the leaf, which may bring several complications. The aim of this thesis was to evaluate the DNA barcoding method for identification of tropical trees by both traditional taxonomic and molecular methods. The reliability of both species recognition methods were compared as well as the results among different tissue types (leave, young wood, old wood). 11 tree species, indigenous to Central and West Africa, were sampled. In total 150 samples were collected both for the morphological and genetic analyses. The DNA extraction was done using the CTAB method and the leaf material yielded the best nucleic acid concentration with the highest purity. For PCR amplification the *trnL* approach was used, the samples were amplified in three different overlapping regions. The 'long' sequence (primer pair *c-d*) appeared fragmented when visualized by electrophoresis. When sent for Sanger sequencing the short region (primer pair *g-h*) returned with low taxonomic discrimination and the sequencing success rate was low as well. Therefore, for the means of sequencing the medium region (primer pair *c-h*) was used. The sequencing success rate was 92%. All the tree species were correctly assigned to their family, 48% to the correct genus and 23% to their species. However, the low recognition to higher taxa may be also given by the small number of listed sequences in the GenBank database. Therefore, the best sequence per species was chosen and submitted to the GenBank database, which now can be found there under their accession numbers and used for future barcode identifications.

Key words: Cameroon, plant barcode, tree molecular identification, *trn*L intron, *trnL* approach, wood barcoding

Contents

List of tables

List of figures

List of the abbreviations used in the thesis

DNA	Deoxyribonucleic acid
GPS	Global Positioning System
PCR	Polymerase Chain Reaction
ICRAF	World Agroforestry Centre
EDTA	Ethylenediamine tetraacetic acid
CTAB	Cetyl trimthylammonium bromide
bp	Base pair
NaCl	Sodium chloride
ddH_2O	Double distilled water
PVP	Polyvinylpyrrolidone
COI	Cytochrome oxidase I
rbcL	Ribulose-bisphosphate carboxylase
matK	Maturase K
ATP	Adenosine triphosphate
BOLD	Barcode of Life Data System
CBOL	The Consortium for the Barcode of Life
DDBJ	DNA DataBank of Japan
ENA	European Nucleotide Archive
NCBI	National Center for Biotechnology Information, U.S.A.
GDP	Gross domestic product
IUCN	The International Union for Conservation of Nature
CZU	Czech University of Life Sciences in Prague
TRIS-HCl	(Hydroxymethyl)aminomethane hydrochloride
RPM	Rotation per minute

1. INTRODUCTION

The tropics harbour approximately 78,800 flowering plant species, which is an over a third of the word's total. However, tropical forests are being degraded at a fast pace, and over half of the tree species may face a direct risk of extinction (Gonzalez et al. 2009). Cameroon's assessed biodiversity reaches to more than nine thousand plant species, of which 156 are endemic. Unfortunately, many species are either threatened, endangered or close to extinction. There are many drivers of biodiversity loss in Cameroon including unsustainable logging and slash-and-burn agriculture (Eyebe et al. 2012). Some species are disappearing at fast pace, before they are identified or discovered. Therefore, large-scale diversity surveys are desparately needed in order to develop informed conservation strategies for these diverse ecosystems (Gonzalez et al. 2009).

However, species identification on the basis of morphological characters often represents a challenging task that requires experienced taxonomists. These morphology-based procedures are usually time consuming and therefore expensive (Pečnikar & Buzan 2013). Moreover, this approach to the task of routine species identification has some limitations, especially the identification of morphologically cryptic taxa, discrimination of different life stages or genders or the use of damaged or incomplete specimens (Hebert et al. 2003).

These limitations and the decreasing number of skilled taxonomists showed that finding a new approach in taxonomical identification was needed (Hebert et al. 2003). DNA barcoding is a relatively new approach aiming to provide rapid, accurate and automatable species identifications by using a standardized DNA region. DNA barcoding is based on three principles: standardization, minimalism and scalability. DNA barcodes are short DNA sequences between 400 and 800 base pairs long, they are amplified using Polymerase Chain Reaction (PCR) and then sequenced. This approach can serve as a new tool for taxonomists and as a device for non-experts to objectively identify species - already described or newly discovered (Taberlet et al. 2007).

This approach is very well developed for animals, where a protein-coding region from the mitochondrial genome - COI (*cytochrome oxidase 1*) is widely used. The COI sequence enables discrimination for more than 98% of animal species (Pečnikar & Buzan

2013). However, finding a plant equivalent to this barcode has proven difficult. There are few limitations concerning barcoding of plant species. Currently, it is recommended to use a core-barcode consisting of portions of two plastid coding regions, *rbcL* (ribulose-bisphosphate carboxylase) and *matK* (maturase K). Nevertheless, this combination could identify species in 72% of the cases and this resolution is unlikely to be evenly distributed across land plant species (Hollingsworth et al. 2011). Although, it has been pointed out that DNA barcoding should be useful in discrimination between forest seedlings or undertaking large-scale biodiversity surveys in situations where taxonomic expertise is limiting. Still, tropical plants present challenges to DNA barcoding, due to the greater abundance of secondary metabolites (Gonzalez et al. 2009).

Therefore, the main objective of this thesis was to evaluate the use of the DNA barcoding method for the identification of tropical trees in humid tropical forests and agroforestry systems of Cameroon.

2. LITERATURE REVIEW

2.1. Taxonomy

The identification and characterisation of living things is essential to biological science. Since the foundation of binomal taxonomy by Carl Linnaeus (1707 – 1778) taxonomists have collected specimens in the field and catalogued and described them according to their morphological and anatomical characteristics. Organisms are then sorted into categories (called taxa) according to these shared features. A taxon, plural taxa, is defined as any unit used in the science of biological classification. Taxa are usually named and arranged in a hierarchical ranking: kingdom, phylum, class, order, family, genus and species (De Queiroz 2007). Reliable species identification is an essential stone for taxonomy, as species are its basic units (Mallet 2006).

2.1.1. Species definition and identification

As stated above, species are considered as the basic unit of classification and fundamental unit of biology (Mayr 1982). However, they remain difficult to define. From several definitions and concepts species concept suggested, only some have found a widespread use, such as the typological, biological, phylogenetic or morphological species concept. The definition of species has been redefined over time in light of new information (Wiley 1978).

The first species concept invented by Linnaeus (1758) was the typological concept (phenotypical), according to this approach species are defined based on phenotypic characteristics of individual organisms that do not occur in other species, such as morphological, anatomical, physiological, biochemical or ethological features.

Phylogenetic concept defines species with regard to the development in time. It says that species is the smallest population of populations which has fixed heritable differences from other such populations. Members of a species are all descendent of a common ancestor (Nixon & Wheller 1990).

Biological species concept defines species as a group of similar living organisms capable of successfully interbreeding, exchanging genes, producing viable offspring and are reproductively isolated from populations of related species (Mayr 1942).

Morphological species concept, which is the most common one, is used to define species based on morphological characteristics. This concept originates from comparative anatomy, where it is possible to visually determine the difference amongst species (Hillis 1987). Morphological features include colour, shape, size and structure. By comparing and distinguishing these features and by using morphological keys, taxonomists are able to determine organisms, identify their species and other higher taxa.

With the development of molecular biology, the newest concept arose, the genetic species concept where DNA sequences are used as determining signs (Herbert & Gregory 2005). Genetic species are defined as a group of natural, genetically compatible, interbreeding populations that are genetically isolated from other such groups (Baker & Bradley 2006). One of the methods used within this frame is DNA barcoding.

2.1.2. Traditional vs. modern taxonomy

In the lastt decades, teaching and financing of taxonomy has declined. However, in the last ten years, taxonomy suddenly became fashionable again, and revolutionary approaches to taxonomy using DNA and Internet technology are now being contemplated. The new excitement about taxonomy is driven partly by advances in technology, and partly by newly perceived needs given by the biodiversity crisis (Mallet & Willmot 2003). Identifying organisms has grown in importance as we monitor the biological effects of global climate change and attempt to preserve species diversity in the face of accelerating habitat destruction. Despite the persistent efforts of taxonomists to map the diversity of all living organisms, it is estimated that most species have not yet been discovered. Estimates of the number of all species of organisms on Earth vary widely, from 3.6 million to possible hundreds of millions. Thousands of plant and animal species are lost each year, most of which have not yet been identified (Wilson 2003).

Several methods of identifying species are commonly used today, the most popular are morphological taxonomy and molecular systematics. Each method has its benefits and downfalls (Friedheim 2016). Morphological systematics originated from comparative anatomy, where species are distinguished visually by their macromorphological features. This method is the foundation for all species identification up till today, it has been used for about 250 years. Therefore, it represents the most reliable method for species that are very well known and described (Hillis 1987).

However, it should not be forgotten that taxonomy is a specialist field. Considering the decreasing number of taxonomists and the increasing number of named species, molecular tools have become a mainstay of modern taxonomic analysis (Pečnikar & Buzan 2013). Molecular systematics come from molecular genetics, where DNA sequences of organism are used for species recognition. Each species is genetically unique, no genome is identical to another (Herbert & Gregory 2005).

There are some cases, where using only traditional taxonomic methods is very difficult or even impossible. For example, determination of small organisms such as microbial species, which are very hard to see with the naked eye. Many species go through different life stages, e.g. the evolution from larvae to adult individual. Also, the existence of sexual dimorphism, where the female and male individuals may vary in size or colour may hinder correct taxonomical identification (Savolainen et al. 2005). Some organisms look morphologically indistinguishable, but are in fact different species, these are called cryptic species. Discrimination of damaged or incomplete specimen is also a complicated case (Duellman & Venegas 2005). The use of morphological keys often demands such high level of expertise that misidentification is a common mistake. These limitations typical for morphology-based identification and the declining number of experts on taxonomy signal the need for a new approach to taxon recognition (Hebert et al. 2003).

Traditional taxonomists argue that the molecular species identification is not sufficient to replace morphological phenotypic characterization (Will & Rubinoff 2004). The main disadvantage of molecular identification is the intraspecific variation. The assumption is, that intraspecific variability should be lower than interspecific variability. However, this variation is not completely specified. Therefore, there is a possibility that some organisms could be mistaken for new species or subspecies (Savolainen et al. 2005). Other arguments such as that morphological approaches to taxonomy are necessarily more accurate or 'richer' than barcoding were refuted by Packer et al. (2009). On the other side, molecular scientists suggest morphological taxonomy is outdated and time-consuming (Hebert et al. 2003). Where one method succeeds, the other has a downfall, which is why a combination of both mentioned methods should be used to ensure the highest possible reliability and precision (Friedheim 2016).

2.2. Plant genome

DNA is the hereditary or genetic material, present in all cells, which carries information for the structure and function of living things. Genetic information of plant cells is carried in DNA molecules, chromosomes, and most of the DNA is stored in the nucleus. The nuclei of plant cells contain linear molecules of DNA, and the number and length of these molecules differ by species. The main purpose of nuclear DNA is to transfer information about structure and features of specific organism throughout generations (Řepková 2013). Beside the chromosomes in the nucleus, chloroplasts and mitochondria have their own DNA. The genome of mitochondria contains DNA for its functions: converting the energy of chemical bonds into the cell energy currency, adenosine triphosphate (ATP), as well as for the mitochondrial protein synthesis. Chloroplast DNA contains genes that are involved in photosynthesis and with components of that protein-synthesizing apparatus which is active within the organelle. Chloroplast genomes usually contain up to 140 genes. Both mitochondrial and chloroplast DNA replicate separately from nuclear DNA (Robinson et al. 2017).

Genetic information is encoded in genes which are composed of exons and introns. An exon is the protein coding region of a gene that contains the information required to encode a protein. In eukaryotes, genes are made up of coding exons interspersed with non-coding introns (Brown 2012). Even though introns are the non-coding part of the genome, they have many other purposes. One of the intronic functions is the increase in protein abundance of intron-bearing genes. They usually guarantee high expression of various genes. Introns are about 10 times longer than the exons. The sequence and length of introns vary rapidly over evolutionary time (Chorev & Carmel 2012).

One of the most important mechanisms of evolution are mutations. DNA mutation is a change of genetic information, usually caused by some physical, chemical or biological factors. The rate of DNA mutation is related to the size of genome. In the animal cell, the mutation in nuclear DNA is relatively slow compared to mitochondrial DNA (10 times slower). In the case of plant species, the plastid and nuclear genomes have a 3 to 10-time greater mutation rate than the mitochondrial genome (Sloane et al. 2012).

2.3. DNA Barcoding

Genomic approaches for species identification exploit diversity among DNA sequences. These sequences can be viewed as genetic ‘barcodes’ that are embedded in every cell. DNA barcoding is a molecular and bioinformatics tool that aims to identify biological species. The role of barcoding is to assign unidentified specimens to an already characterized species. DNA barcoding is a relatively new method where each taxon can be uniquely branded by its genetic information. The novelty of this method is that only a small segment of DNA is needed instead of using the whole genome (Hebert et al. 2003).

The DNA barcoding concept was proposed by Hebert et al. (2003) and it represented a major step forward for DNA-based species identification. However, the origins of the idea of using molecular tools for determination of species differences goes back to the mid-60s (Hubby & Lewontin 1966). Determination of the whole organisms by molecular methods was already mentioned in the 80s (McAndrew & Majumdar 1983). The term DNA barcode was firstly used in a paper published by Arnot et al. (1993), where the authors mention DNA-based genotyping with the use of PCR amplification. The idea of using a small segment of the genome to discriminate organisms was initially accepted by those working with viruses and bacteria, which are the least morphologically traceable groups (Pace 1997).

DNA barcoding is a method developed to identify species using short DNA sequences that should differ between species and at the same time be similar within the same species. According to these differences, individual species can be determined. The purpose of DNA barcoding is to provide a rapid, reliable and relatively cheap identification method. Using this method species can be identified even from small, damaged, or industrially processed material. It should be easy enough to use so that a person without extensive taxonomy knowledge can objectively identify organisms to the level of species. DNA barcode is a unique pattern of DNA sequence that identifies each living organism. The universal DNA barcode is a molecular marker, that should be present in all species and should have enough discriminatory power to distinguish them (Hebert et al. 2003).

One of the main features of DNA barcode is the possibility to easily associate all life stages and genders or to identify organism from parts and pieces, damaged or incomplete specimens, or to distinguish a matrix containing a mixture of biological species (Valentini

et el. 2010). DNA barcoding is suitable for both the molecular identification of already described species (Hebert et al. 2003) as well as for the discovery of undescribed species (Hebert et al. 2004).

2.3.1. Barcoding Gap

An important aspect of DNA barcoding is accuracy. Accuracy depends especially on the extent of, and separation between, intraspecific variation and interspecific divergence in the selected marker. The more overlap there is between genetic variation within species and divergence separating sister species the less effective the barcoding becomes. If interspecific divergence exceeds intraspecific genetic variability, a gap called the 'barcode gap' is formed (Meyer & Paulay 2005). Hebert et al. (2004) defined such 'barcoding gap' as the existence of at least 10 times greater average interspecific distance over the average intraspecific genetic distance. According to the size of the gap, or the extent of the difference between the intraspecific and interspecific variability, it is possible to determine if the samples represent one or more species.

Some researchers state that the barcoding gap does not really exist, as it represents only an artefact due to the insufficient sampling across taxa (Wiemers & Friedler 2007). This hypothesis might be fatal for DNA barcoding and for the approach of identifying using thresholds, because without the existence of a barcoding gap, it would not be possible to detect whether a sample is assigned to the right species.

2.3.2. Databases

The validity of DNA barcoding also depends on establishing reference sequences from taxonomically confirmed species. Extensive reference databases allow an unknown sequence of interest to be matched to the reference database to identify the species from which it came (Stoeckle 2003). Therefore, an assembly of a comprehensive library that links barcodes and organisms is required. The sequence data that are essential for this process are deposited and maintained in publicly funded databases and can be accessed without cost. The database created for the purpose of publishing all data obtained by DNA barcoding is called Barcode of Life Data System (BOLD) (http://www.boldsystems.org/). It is an informatics workspace aiding the acquisition, storage, analysis and publication of DNA barcode records, where all working groups can insert results from their barcoding

projects (Ratnasingham & Hebert 2007). The database was created to be public and freely available, it has records of barcode sequences, vouchers, images, maps, collection coordinates, lists of primers used for DNA barcode amplification and published papers related to DNA barcoding. Barcode sequence data of unknown species can be rapidly and accurately compared using a large suite of online database tools for the collection and management of specimen. It contains over 370,000 plant barcodes representing over 58,510 species of plants (De Boer et al. 2015).

Another vast genetic sequence database is the GenBank, it is a collection of all publicly available DNA sequences. GenBank is part of the International Nucleotide Sequence Database Collaboration, which comprises of the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI (National Center for Biotechnology Information, U.S.A.) and these three organizations exchange data on daily basis.

In the autumn of 2004, the National Center for Biotechnology (NCBI) (www.ncbi.nlm.nih.gov) sealed a partnership with CBOL (The Consortium for the Barcode of Life) whereby barcode standard DNA sequences and relevant supporting data can now be archived in GenBank (Savolainen et al. 2005).

Both BOLD and GenBank contain the same public records, but they offer different options for optimizing their use as reference libraries. Therefore, it depends on the specific user and their needs. Sonet et al. (2013) recommended the use of BOLD Identification System and to search the dataset including early-released sequences. This option optimizes the number of best-matches and allows to verify the quality of the data.

2.3.3. Ideal DNA Barcode

The universal barcode should fulfill these criteria: (i) it should be sufficiently variable to distinghuish among all species but conserved enough to be less variable within than between species. (ii) It should be standardized with the same DNA region; the same DNA region is used for different taxonomic groups. (iii) The target DNA region should contain enough phylogenic information to easily assign species to its higher taxa (genus, family, etc.). (iv) It should be extremely robust, with highly conserved priming sites, to allow an easy and reliable DNA amplification and sequencing. (v) The target DNA region should be also short enough to allow amplification of degraded DNA (Taberlet et al. 2007).

Unfortunately, an ideal barcode does not exist. There is no universal DNA barcode gene, that is conserved in all domains of life and exhibits enough sequence divergence for species discrimination. However, all of the five mentioned criteria listed are not equally important e.g. for taxonomist the most important is a high level of variation with sufficient phylogenic information. In contrast, in forensics or for analyses of processed food, the crucial factors are the levels of standardization and robustness (Taberlet et al. 2007).

As already mentioned in chapter *2.2. Plant genome*: The rate of DNA mutation is related to the size of genome. DNA barcoding requires the mutation rate to be slow enough so that intraspecific variation is minimised, but sufficiently rapid to highlight interspecific variation. In the animal cell, the mutation in nuclear DNA is relatively slow compared to mitochondrial DNA (10 times slower). Hence, a much longer nucleotide sequence is required for nuclear DNA than for necessary with mtDNA in order to provide a barcode capable of differentiating species (Waugh 2007). Therefore, the ideal barcode for animal species is the mitochondrial gene encoding the cytochrome c oxidase. In the case of plant species, the plastid and nuclear genomes have a 3 to 10 time greater mutation rate than the mitochondrial genome (Sloane et al. 2012). Therefore, searching for universal DNA barcode in plants is focused on nuclear and plastid genomes, while the chloroplast genome is the most preferable because of its presence in each plant cell in a higher number of copies (Pečnikar & Buzan 2014).

2.3.3.1. Universal barcode for animals – COI

When Hebert et al. (2003) proposed biological identification through DNA barcodes, they also suggested a standard barcode, the cytochrome *c* oxidase I gene (COI). The sequence from mitochondrial protein-encoding gene is 648 bp long. The search for a universal animal barcode was focused on the mitochondrial DNA, given the high rate of evolution of mitochondrial genes in animals. The speed of evolution seems to be sufficient here to distinguish closely related species. Another advantage of mitochondrial DNA is the lack of introns, its limited exposure to recombination and its haploid mode of inheritance (Saccone et al. 1999). The COI does have two important advantages. First, the universal primers for this gene are very robust. Second, COI appears to possess a greater range of phylogenetic signals than any other mitochondrial gene (Hebert et al. 2003).

The COI DNA barcode fits the criteria for an ideal barcode well. It is a haploid, uniparentally-inherited, single locus that shows high levels of discriminatory power. It is

a protein-coding region present in high-copy numbers per cell, and it is not prone to drastic length variation or frequent mononucleotide repeats (Hollingsworth et al. 2011).

Since its 'discovery' the COI barcode has proven to be successful as a taxonomic tool and is highly effective in identifying many animal groups such as birds (Hebert et al. 2004), spiders, butterflies, fish, various insects and more (Barrett & Hebert 2005).

The COI sequence enables discrimination for more than 98% of animal species (Pečnikar & Buzan 2013).

2.3.3.2. Plant Barcode

Searching for suitable plant DNA barcodes is more problematic than in animals, due to lower heterogeneity in the mitochondrial COI gene of plants. Furthermore, the mitochondrial DNA in plants is limited in general and it evolves too slowly. Therefore, the universal barcode for plants had to be found outside of the mitochondrial genome (Hollingsworth et al. 2011). Thus, the search for universal DNA barcode for plants was focused on nuclear and plastid genomes, especially the chloroplast genome, which is an alternative to the animal mitochondrial genome. The chloroplast genome could contain suitable barcoding markers, because it is present in almost each plant cell in high number of copies and consists of conserved gene sequences. The chloroplast cells are present even in non-green parts of the plant species, such as branches, twigs and stem (Pečnikar & Buzan 2013).

An agreement on a common plant barcode is necessary to progress towards the creation of a shared community resource. To facilitate and formalize the selection of standard plant barcode, CBOL established a Plant Working group. The group was formed by representatives from different research groups, that had proposed or tested the leading candidate of barcoding markers (Table 1) (Hollingsworth et el. 2011). Among the tested barcodes was the *trnH-pbsA* intergenetic spacer (Kress et al. 2005) as well as phylogenic markers such as *rbcL* and *trnL-F* (Chase et al. 2005). Suggested barcodes also involved various combinations of seven plastid markers, these were *rpoC1+rpoB+matK* or *rpoC1+matK+trnH-psbA; rbcL+trnH-psbA* and *atpF-H+psbK-I+matK* (Kress & Erickson 2007). Each candidate marker had different strengths and weaknesses, therefore there is no unanimous decision, the conclusion is only based on what the majority preferred. In 2009, the recommendation of the CBOL Plant Working group was, to use

two plastid coding regions, *rbcL* and *matK*, as a 'core barcode' to be supplemented with additional markers as required (CBOL Plant Working group 2009). The reason behind the choice of the *rbcL+matK* pair as a core barcode is the fact, that they complement each other. Because of the straightforward recovery of the *rbcL* region, the region is also easy to amplify, sequence and align in most land plants. The strong side of the *matK* region is its discriminatory power, and the fact that it is the most rapidly evolving coding section of the plastid genome. These facts make *matK* to be the closes plant analogue to the COI animal barcode. On the other hand, *matK* can be difficult to amplify and *rbcL* does not have great discriminatory powers (Hollingsworth 2011). This combination will lead to a species-level identification in 72% of the cases (Pečnikar & Buzan 2013).

Table 1. Characteristics of different markers that have routinely been included in plant barcoding studies

Marker	Genomic source	Type	GenBank accessions	GenBank genera	GenBank species	Length - genomes	IQR length	Length - range	Number of samples	Frequency
nrITS	Nuclear	Transcribed spacer and 5.8S gene	102 684	13 307	52 450	705	683 - 724	407 - 1 630	5 020	0.013
nrITS2	Nuclear	Transcribed spacer	111 370	15 817	57 579	494	492 - 506	157 - 670	646	0.005
atpF-H	Plastid	Inter-genic spacer	1 180	274	664	669	578 - 707	390 - 918	134	0.440
matK	Plastid	Protein coding	34 647	7 454	22 701	889	880 - 889	862 - 910	132	0.235
psbK-I	Plastid	Inter-genic spacer	1 241	208	626	468	444 - 492	112 - 1 253	134	0.500
rbcL	Plastid	Protein coding	27 725	8 959	20 374	654	654 - 654	654 - 654	134	0.000
rpoB	Plastid	Protein coding	3 341	751	1 970	548	548 - 548	536 - 590	132	0.008
rpoC1	Plastid	Protein coding	5 314	1 110	3 075	616	616 - 616	610 - 622	132	0.000
trnH-psbA	Plastid	Inter-genic spacer	23 526	2 833	11 539	509	401 - 617	226 - 934	135	0.296
trnL-F	Plastid	Intron and inter-genic spacer	59 197	9 129	35 130	994	907 - 1 037	201 - 2 114	132	0.280
***trnL* (P6)**	Plastid	Intron	70 811	10 561	38 329	87	83 - 91	51 - 135	130	0.054

Source: Hollingsworth et al. 2011

* GenBank accessions/genera/species = Approximate number of GenBank accessions/genera/species
Length - genomes = Median amplicon length (bases) in completely sequenced plastid genomes
IQR length = IQR amplicon length (bases)
Length - range = Amplicon length range (bases)
Number of samples = Number of samples used to estimate amplicon length
Frequency = Frequency of amplicons with mononucleotide repeats ≥ 10 bases

2.3.4. DNA identification of tree species

Tropical plants present more challenges to DNA barcoding compared to plants from temperate zone. DNA extraction usually is more difficult in tropical plants, due to the greater abundance of proteins, lipids, polysaccharides and especially of secondary metabolites. Moreover, some lineages have a great number of species, this close relativeness is reducing levels of interspecific divergence. Finally, it has been shown that woody plant lineages show consistently lower rates of molecular evolution compared to herbaceous plant lineages, suggesting that the application of DNA barcoding will be more difficult for tree-flora than for non-woody floras (Gonzalez et al. 2009).

Barcoding trees is possible using a plant DNA barcode. However, in case of some trees, it is very difficult to obtain leaf material, because of their height or absence of leaves due to seasonality. These are one of the reasons explaining the need for DNA barcoding of wood material. But wood, in general, has different genetic characteristics from leaves that affects the selection of DNA barcodes (Jiao et al. 2018). Compared to the amount of DNA available in soft plant tissue, such as leaves, buds and fruits the amount of DNA preserved in wood is small. However, the degradation of wood DNA post-mortem is much more limited. Wood DNA is therefore a good candidate for forensic or archeological applications (Tang et al. 2011).

The formation of wood consists of five major steps: cell division, cell expansion, cell wall thickening, programmed cell death and hardwood formation. Because of this, wood is a poor source of DNA, as it is basically composed of dead cells. Generally, DNA in wood gradually reduces in quality and quantity and degrades into small fragments even before trees are felled. DNA extraction from wood is not as simple or direct as from other plant parts (Plomion et al. 2001).

In study by Deguilloux et al. (2002), it has been shown that chloroplast fragments of varying length could be recovered from different parts of wood. Only the cambium is composed exclusively of living cells, sapwood is composed of both living and dead cells, and all cells in the heartwood are dead (Figure 1). This is all related to the gradual transformation of the cells during aging. Therefore, long chloroplast fragments were obtained from fresh sapwood, shorter from sapwood of dried trees and from heartwood fragments longer 380 bp could not be amplified (Deguilloux et al. 2002). Genomes with

high copy number, like chloroplast and mitochondria, are potentially superior for DNA analysis, compared to single copied nuclear genes. A high copy number should ensure high amplification success rate, particularly for the highly degraded template (Jiao et al. 2018).

Therefore, the ideal DNA barcode for barcoding of wood should be short, making it easy for recovery, and has sufficient number of information to be able to discriminate species. The shorter the amplicon, the higher the chance of successful amplification. Therefore, mini-barcodes should be used instead of full-length barcodes. Mini-barcodes are suitable for species identification, when high-quality DNA is not available and highly degraded DNA is retrieved (Jiao et al. 2018).

Gonzalez et al. (2009) tested eight DNA plant markers in a large biodiversity survey of the Amazonian forest. They were able to extract DNA from cambium tissue, which can be useful in future routine tropical forestry monitoring programs. The markers *rpoC1*, *rbcLa* and *trnL* were all sequenced easily from both leaf and cambium tissue. However, from the previously mentioned options, the *rpoC1* consistently had the worst performance as a DNA barcode. The *trnL* intron ranked second in 'best close match' test and fifth in the monophyly and clustering test. It was also twice as variable as *rbcLa.* Its variability is comparable to one of the universal barcodes – *matK*, but it is much easier to sequence. Therefore, the conclusion of the study was that the *trnL* intron might be an interesting option for barcoding projects. This barcode has been used in various ecological projects already and it has proven to be very useful for the identification of degraded DNAs. Unfortunately, none of the rates of correct identification exceeded 70% (Gonzalez et al. 2009).

Gonzalez et al. (2009) examined if DNA barcoding could be used for identification of juvenile individuals, which is usually very hard to do using traditional taxonomy. They sampled 152 tree saplings, 96% of which could be identified to the species or at least to the genus. Thus, DNA barcoding does show much potential in the identification of species in different life stages. Even though, the coding plastid markers were not variable enough to identify samples to the species level, they were able to assign them to higher taxa. Finally, these results show that DNA barcoding could have important implications for ecological purposes, such as tropical plant diversity surveys (Gonzalez et al. 2009).

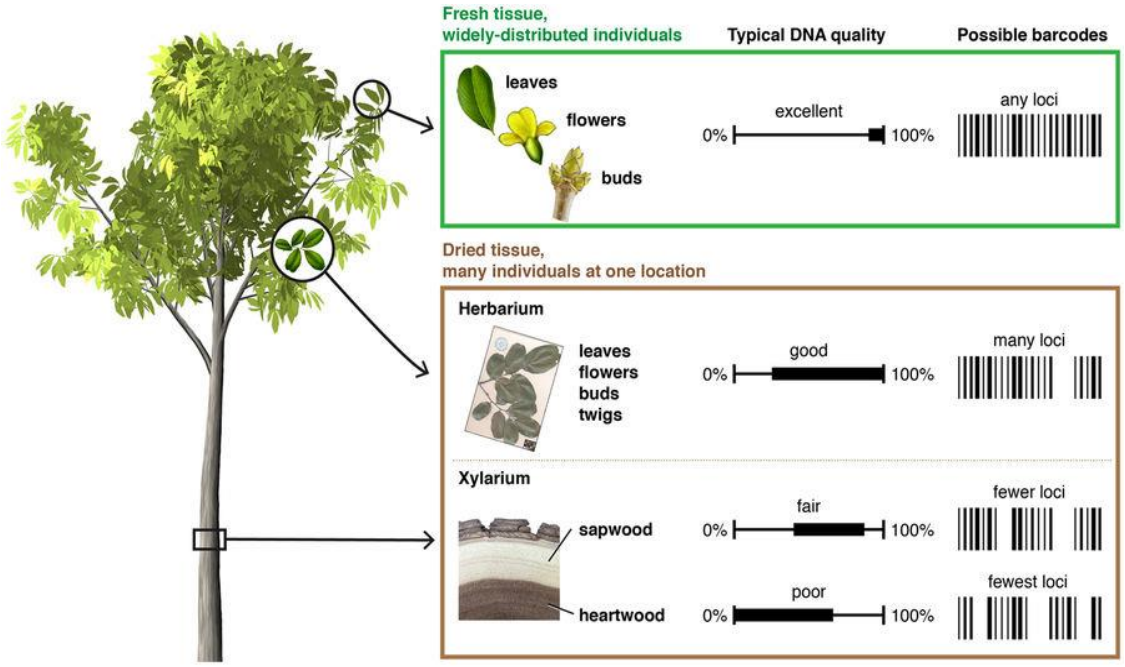

Figure 1. A schematic representation of the potential strengths and weaknesses of source tissue, source Jiao et al. (2002)

2.3.4.1. DNA mini-barcode

A DNA mini-barcode is a short DNA fragment, generally 100-250 bp, suitable for species identification. DNA mini-barcoding has extensively developed over the past 10 years. Many studies have highlighted its importance as an extension of DNA barcoding. A small portion of the barcode region may be used in place of full-length barcode to overcome DNA degradation for samples with poor DNA preservation (Gao et al. 2019). Conventional DNA barcoding uses the length of an approximately 650 bp and more. In most cases, it is easy to achieve a successful DNA isolation and PCR amplification, when using freshly collected and well-preserved specimens. It is more difficult to obtain a full-length barcode in older museum specimens and samples which have been preserved in formalin or similar DNA-unfriendly preservatives. When the DNA molecules degrade into fragments shorter than the spanning length of the primers (650 bp), it is nearly impossible to amplify the DNA barcodes (Hajibabaei & McKenna 2012). Experiments have repeatedly shown (Chen et al. 2005, Särkinen et al. 2012, Bergerová et al. 2011) that PCR amplification success greatly increases with the decrease in amplicon size. Therefore, the need to find shorter DNA barcodes is in place, the so-called mini-barcodes. DNA mini-barcodes can be used for species identification of digested material, old herbarium/museum specimens, ancient DNA, processed medicinal herbs and other

materials with poor source of DNA. Due to their reduced size, mini-barcodes are presumably PCR-amplified at higher rate than full-length barcodes. However, identification success and the taxonomic discriminatory power is impaired relative to that of a full-length barcode, because of the reduced number of nucleotides (Dong et al. 2015). Nevertheless, shorter DNA sequences – mini-barcodes – have been robustly recovered and shown to be effective in identifying majority of specimens to the species level (Hajibabaei & McKenna 2012).

Thus far, a few tries have been made to design a DNA mini-barcode. The selection of an optimal mini-barcode follows similar criteria as the criteria employed in the selection of full-length barcode: (i) PCR universality, (ii) sequence quality, (iii) taxonomic discrimination (Little 2014).

In a paper published by Taberlet et al. (2007) the first example of mini-barcode was demonstrated: the short region of tRNA-Leu (*trnL*) in chloroplasts, termed P6 loop (10-143 bp), which could be amplified from processed food and permafrost samples (Taberlet et al. 2007). Meusier et al. (2008) proposed mini-barcodes from the COI gene region. They analyzed 100 bp and 250 bp of DNA in COI region and accomplished successful identification rates of 90% and 95%, respectively. They also developed universal primer pairs for mini-barcodes of 120-150 bp, and achieved higher success rates for these amplicons than full-length barcoding (Meusnier et al. 2008). Little (2014) studied the *rbcL* barcode, 12 DNA mini-barcodes were selected from this region and PCR amplification for all mini-barcodes was successful for 90.2 – 99.8% (Little 2014).

Mini-barcodes offer a solution for samples with degraded DNA. The shorter the amplicon, the more likely it is for the PCR to be successful, but as well as the shorter the amplicon the lower possibility of correct taxonomic discrimination (Little 2014).

2.3.4.2. trnL intron

The *trnL* intron and the intergenic spacer between *trnL* and *trnF* have been widely used in plant systematics since the early 1990s. This frequent use is due to the early publication of a set of primers by Taberlet et al. (1991). The main advantage of the use of primers for a long time is that, since their publications, they have been extensively used and the *trnL* intron has been thoroughly analyzed and is well understood (Quandt & Stech 2005). These robust sets of primers allow a routine recovery, they are well conserved and

generally simple to sequence, although the mononucleotide repeats (Table 1) can have an impact on sequencing reads in some taxa. For the use of species identification, the biggest strength of *trnL* intron is the presence of a small-stem loop structure within the intron, the P6 loop. P6 has conserved priming sites flanking a variable loop of 10-143 base pair (Figure 2). This very short barcode, so-called 'mini-barcode', has proved to be very useful for ecological studies, particularly studying highly degraded DNAs found in processed food, in fossil remains or other materials poor in DNA (Hollingsworth et al. 2011).

The main, and maybe the only drawback is the low resolution of the *trnL* intron compared to several other non-coding chloroplast regions. This weakness is even more dramatic with the use of the short P6 loop. Therefore, the *trnL* intron does not represent the best choice for characterizing plant species and for phylogenetic studies among closely related species. However, this drawback can be compensated by the possibilities of standardization, the selection of one or more reference genes. Another advantage of this approach is the extremely well conserved primers, both the entire region (*c* and *d*) and the P6 loop (*g* and *h*). This advantage is crucial when amplifying multiple species within the same PCR. Another advantage is that the number of *trnL* intron sequences available in databases is already very high. Finally, the robustness of both systems represents an important advantage, coming mainly from the primer universality (Taberlet et al. 2007).

This '*trnL* approach' of ecological barcoding has been developed in parallel to the major international barcoding consortia of the International Barcode of Life Project (iBOL). The *trnl* approach has been widely used for reconstructing phylogenies between closely related species or for identification of plant species (Hollingsworth et al. 2011).

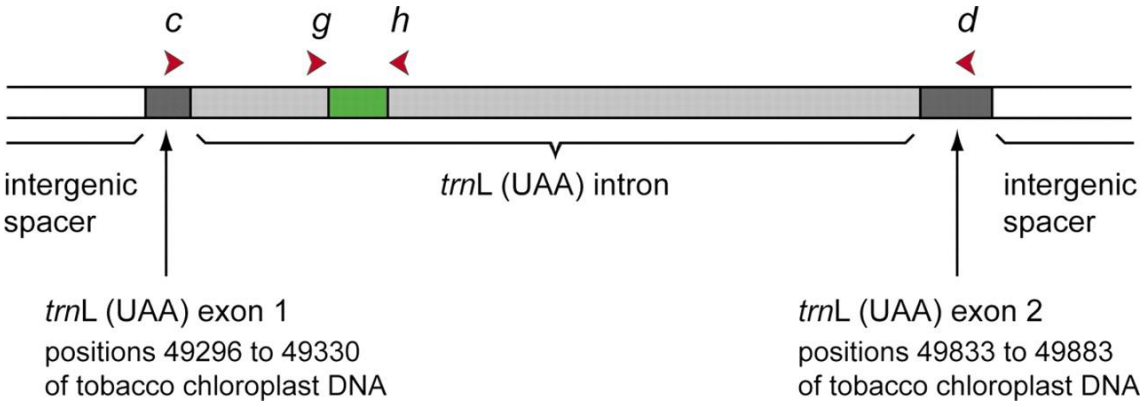

Figure 2. Position of the primers c, d, g and h on the chloroplast trnL (UAA) gene. The P6 loop amplified with primer g and h is indicated in green. Source: Taberlet et al. (2007)

2.4. Practical uses of DNA barcoding

The area of application of DNA barcoding ranges through different fields of study. One of the fields that benefits the most is taxonomy. The possible uses range from fast species discrimination, recognition of cryptic species, detection of alien species to the identification of endangered/ threatened species or the chance of linking egg or larval life stage to an adult species. One of the most known case of unrevealing cryptic species is from Hebert et al. (2004): *Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator.* The study revealed ten different species of butterflies, which were previously considered as a single species due to their morphological similarity (Hebert et al. 2004).

Another field that may use DNA barcoding is biosecurity and public health, e.g. detection of infections and various illnesses spread by parasites or insects. The identification of the parasite itself is extremely important, because it goes through different life stages or include multiple hosts. In Africa, DNA barcoding was used to identify mosquitos that spread lymphatic filariasis (Becker et al. 2010). DNA barcoding can be also used for determination of pests, because the pest in larvae form are very difficult to distinguish from other larval organisms. Floyd et al. (2010) discovered new pest species in California, *Epipyas postvittana*, an invasive species originated in Australia.

Assessing biodiversity using DNA barcodes provides advantages, especially in species-rich ecosystems. Sampling can be done by traditional method, by sampling living organisms or by analyzing samples from soil, water and air. DNA barcoding can also reconstruct ecological conditions on Earth in the past (Pečnikar & Buzan 2014).

The food industry also uses DNA barcodes, especially for the areas of food safety and food quality. It has been used in tracing the origin of seafood, whereas it can also prevent mislabeling and species substitution. Barcodes can trace the origin of meat in sausages and pork pates or the overall content of industrial food products (Teletchea et al. 2008).

DNA barcoding of wood material can be also used in the case of illegal logging. Thanks to the ability of tracking down timber resources from marketplace, DNA barcoding of wood material can be used to check the legality of logging or whether endangered species are not cut down (Nithaniyal et al. 2014).

However, to reach the full potential of barcoding, the method has to work at 100% accuracy for all species (Pečnikar & Buzan 2014). This is currently considered as the biggest drawback of the method.

2.5. Ecological and geographical characteristics of Cameroon

Cameroon is often referred to as 'Africa in miniature' since it mirrors the continent's diversity, both geological and cultural, it exhibits all ecological zones and vegetation types such as deserts, mountains, rainforests and savannas. It is the fourth most biodiverse country in Africa. The country is part of the Congo Basin which is the second largest tropical rainforest hot spot in the world and harbours a wide range of biological resources, Cameroon's forests represent about 10% of the total Congo basin area (Eyebe et al. 2012).

The south of the country is formed mainly by wetlands composed of tropical rain forests and mangroves. The northern parts of the country are essentially dominated by tropical grasslands, savannas mixed with some arid trees (Epule et al. 2014). Cameroon is administratively divided into 10 regions, these regions are grouped into three different agro-ecological zones: humid savannah (Northwest and Western regions), dry savannah (Adamawa, North and Extreme north regions) and forest zones (Southwest, South, Centre and East regions) (Foundjem-Tita et al. 2014).

The total surface area of Cameroon is estimated at 475.4 million km^2. About half of the country is forested (around 22.5 million hectares) designated forest area takes about 18 million hectares a half being under forestry concession and the other half serves as national parks or hunting zones. The export of sawn wood is one of the country's most important trade income as it accounts for more than 10% of the country's GDP (gross domestic product). Cameroon is the sixth largest tropical wood exporter in the world (Alemagi & Kozak, 2010). The 4-5 million hectares of forest which lie outside of the designated forestland is occupied by local farmers involved in agroforestry practices (de Wasseige et al. 2012).

Arable land covers 13% of the total surface area. Agriculture is the main pillar of Cameroon's economy accounting for nearly a half of GDP (43%), employing roughly 70% of the population. Also, more than one third of total export earnings is generated from agriculture. The key crops of the southern parts of the country consist generally of

tubers, yam (*Dioscorea* spp.), cassava (*Manihot esculenta*), potatoes (*Solanum tuberosum*) other starchy crops such as plantains (*Musa* × *paradisiaca*) and vegetables. Typical crops for the north are cereals and legumes: groundnuts (*Arachis hypogea*), beans (*Phaseolus* spp., *Vigna* spp.), maize (*Zea mays*), rice (*Oryza* spp.), sorghum (*Sorghum bicolor*). However, cereals and tubers such as cassava and maize are usually grown all over the country (Shackleton et al. 2007). Cameroon's primary export crops are cocoa (*Theobroma cacao*), cotton (*Gossypium* spp.), coffee (*Coffea arabica, canephora*), palm oil (*Elaeis guineensis*), rubber (*Hevea brasiliensis*) and bananas (*Musa* spp.) (FAO 2006).

Cash crops (coca, coffee, rubber) are usually planted in intensive nearly monoculture system, with scattered upper-storey shade trees. Arable crops are managed in the under-storey. These lands are reformed from primary or secondary forests into mixed agroforests and homegardes (Akinnifesi et al. 2007).

2.4.1. Characterization of study area

The study site was located in the South and Central region (Figure 3). The land cover of these regions consist mainly of humid lowland forest, predominantly of secondary origin with scattered paces of primary woods. In these dense rainforests hardwood evergreen trees (including: *Khaya ivorensis, Diospyros crassiflora, Triplochiton scleroxylon, Lovoa trichilioides, Entandrophragma cylindricum*) can grow more than 60 metres tall (Robiglio et al. 2010). The trees grow naturally in the agro-ecological forest zone and are quite abundant both in the wild and on the farm, therefore, the farmers implemented trees into their agricultural practices (Foundjem-Tita et al. 2014).

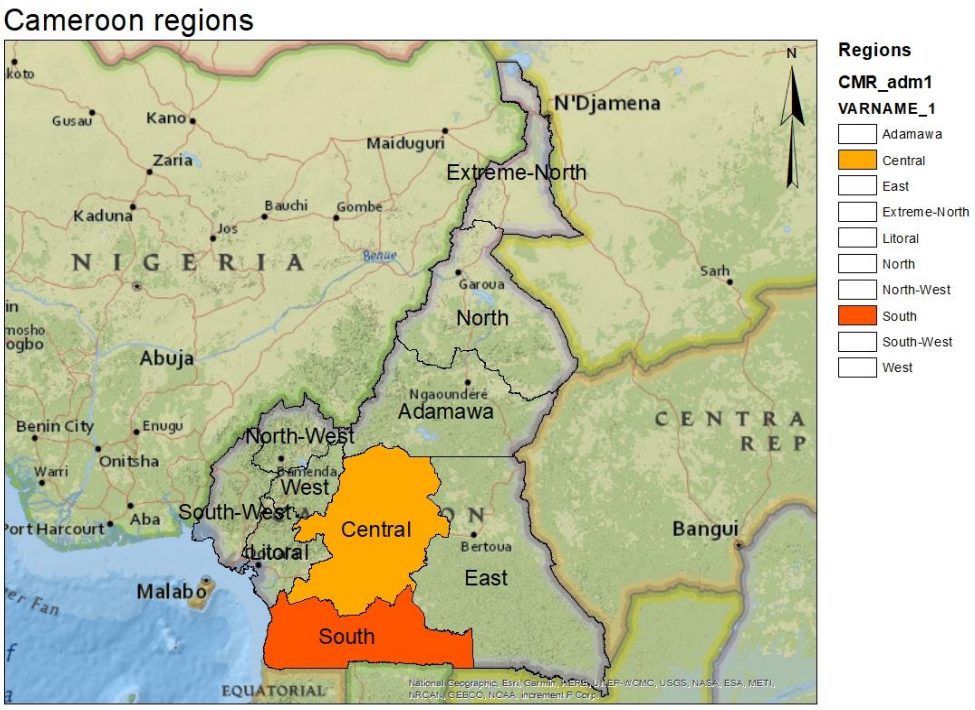

Figure 3. Map of Cameroon regions with marked South and Central regions

The climate of the regions is a tropical climate (type A, according to Köppen-Geiger climate classification) with an average temperature of 20°C and higher and significant levels of precipitation (1867mm). In the specific areas, where the research was conducted, the climate is classified as tropical monsoon climate (Am type), characterized by regular changing of dry and rainy seasons. Most months of the year are marked by significant rainfall. The short dry season has little impact (Figure 4) (Climate-Data, 2020).

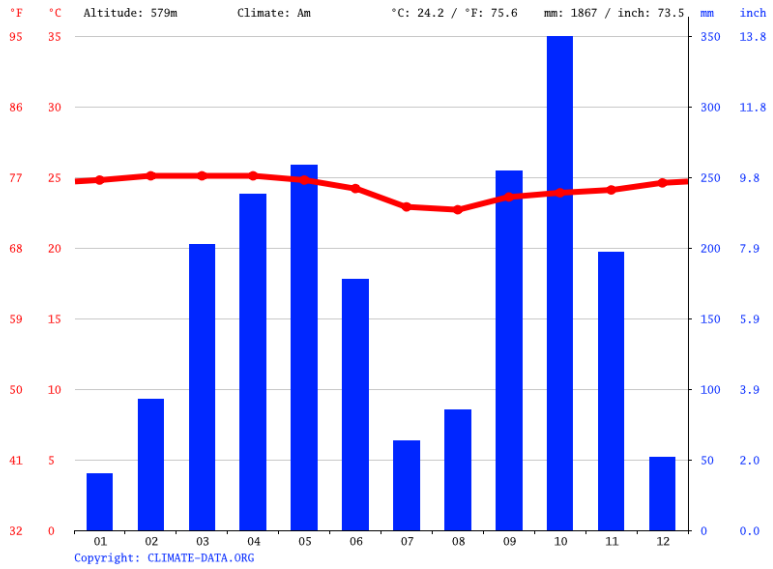

Figure 4. Climate diagram of South region, source: en.climate-data.org

2.5. Characterization of selected trees

The chapter presents a brief characterization of the trees, selected for the purposes of the thesis completion. All the species are common in the forests and farms of West/Central African region. Almost each of the trees is multipurpose, however, they are divided according to their major utilization.

2.5.1. Fruit trees

Fruiting trees are the most popular and widely cultivated tree group in Cameroon. They usually have many more functions such as medicinal, timber, fuel, fodder and more. The fruits are typically very rich in micronutrients thus contributing to balanced diet of local communities. The medicinal properties of fruit are also not negligible because they commonly serve as a primary source of healthcare. Moreover, the fruits and seeds provide a great additional income opportunity to the local farmers and might play a role as export commodity (Leakey & Van Damme 2014).

Cola acuminata x Cola nitida

Cola nitida (Vent.) Schott &Endl. (Figure 5, 7) and *Cola acuminata* (P.Beauv.) Schott & Endl. (Figure 6, 8) are species of the genus *Cola* belonging to the family Malvaceae. They are native to tropical Africa and generally referred to as the kola nut. They figure largely in historical and modern trade and play important cultural role throughout Central Africa. The evergreen small or moderate-sized trees are cultivated for their bitter-tasting edible nuts (cotyledons) which are high in caffeine and theobromine content and are chewed as a stimulant (Niemenak et al. 2008). These powerful stimulants counteract fatigue, suppress thirst and hunger, and are believed to enhance intellectual activity. Other parts of the trees are traditionally used, including the leaves, twigs, flowers, fruit follicles and bark. These parts were used to prepare a tonic as a remedy for coughs, diarrhea, vomiting and chest pain. The twigs are used as chewing sticks to clean the teeth (Tachie-Obeg & Brown 2004).

Cola genus comprises of about 125 - 140 species. In Cameroon, 39 *Cola* species and the most commonly used are *C. acuminate, C. nitida, C. verticillate, C. anomala* and *C.pachycarpa.* They are cultivated in association with cacao and/or coffee as a shade tree (Niemenak et al. 2008). *C. acuminata* and *C. nitida* are two of the most economically

important members of the genus. The fruits consist of 1 to 10 follicles and the seeds comprise of two cotyledons (*C. nitida*) or 3-6 cotyledons (*C. acuminate* and *C. anomala*). The number of cotyledons is one of the characteristic for distinguishing these very similar species (Table 2.) (Tachie-Obeng & Brown 2004).

Table 2. Morphological differences between *Cola nitida* and *Cola acuminata*

	Cola nitida	*Cola acuminata*
Tree height	8 – 12 m (sometimes up to 24 m)	18 m (6 – 9 m in cultivation)
Branching	Bole unbranched for few metres	Bole branching low
Leaves	Abruptly acuminate, flat, nerves promient	Broadly acuminate, curved, twisted
Fruits	Curved, prominent keel extended to form a beak, rugose or tuberculate, green, smooth to touch	Straight or slightly curved, not rugose or tuberculate, russet, rough to the touch
Seeds	2 cotyledons	3 – 6 cotyledons

Source: Tachie-Obeng & Brown 2004

Figure 5. *Cola nitida*, Nkongmeneck 1985

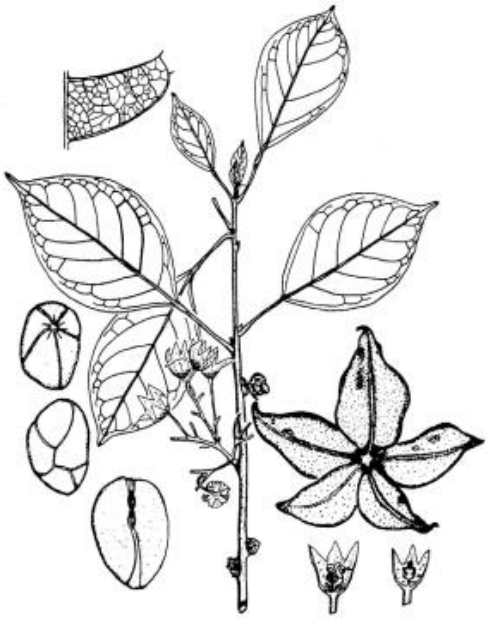

Figure 6. *Cola acuminata*, Nkongmeneck 1985

Figure 7. *Cola nitida*, author's archive

Figure 8. *Cola acuminata*, author's archive

Dacryodes edulis

Dacryodes edulis (G.Don) H.J.Lam (Figure 9, 10) also known as safou or African plum is dioecious shade-loving tree belonging to the Burseaceae family. It is an oliferous fruit tree found in equatorial and humid tropic climates, naturally occuring in the countries bordering the Gulf of Guinea (Rodrigues et al. 2018). It is a medium-sized, evergreen tree reaching a height of 18-40 metres in the forest, however not more than 12 metres in under the conscious cultivation. It is usually low branching with a deep dense crown (Ajibesin 2011). The tree is easily planted and found in many parts of Cameroon, usually in homegardens and cocoa farms. It represents 21-57% of all fruiting trees in farmer's fields and plays an important part in the economy of rural communities (Leakey 2012).

D.edulis is mainly cultivated for the fruit and its nutritional and market values. The fruits are ellipsoid drupe and vary in size and shades of colour by variety. The exocarp is thin and pink, ripening from greenish-pink to dark blue, purple or almost black (Ajibesin 2011). The edible fruit bears high oil content and is an important source of nutrients, especially lipids, proteins and minerals. The mesocarp is usually boiled in water, roasted or, sometimes, also consumed raw. The combination of the high content of fatty acids and amino-acids makes safou an alternative source of vegetable oils for the food, pharmaceutical and cosmetic industries (Tchoundjeu et al. 2002). All these characteristics makes this

agroforestry tree very important to the poor and malnourished people of West and Central Africa (Rodrigues et al. 2018).

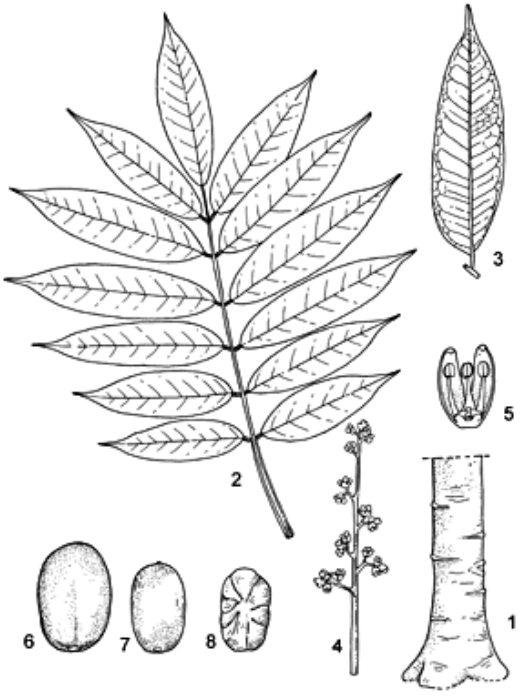

Figure 9. *Dacryodes edulis*, Maňourová

Figure 10. *Dacryodes edulis:* 1 - base of bole; 2 - leaf; 3 - leaflet; 4 - inflorescence; 5 - male flower in longitudinal section; 6 - fruit; 7 - endocarp; 8 – seed. Redrawn and adapted by W. Wessel-Brand, PROTA

Garcinia kola

Garcinia kola Heckel (Figure 11, 12) also referred to as bitter kola is part of the Clusiaceae family and naturally occurs from Congo to Sierra Leone. However, Cameroon and Nigeria are considered to be the major presence hotspots. It is a shade-tolerant medium-large tree growing up to 30m tall, usually it is found around 12 - 15m tall. It is highly preferred as a shading tree in coca plantations (Fondou & Manga 2000).

G. kola has been referred to as a "wonder plant" thanks to the frequent usage of all parts of the tree for medicinal purposes. Both bark and seeds are used as a cure for treatment of gastric and liver disorder. Bark of the species is also used for tanning leathers, twigs are used as a chewing sticks, gum treats gonorrhoea and latex is applied to wounds. The bark also serves in palm wine production, to enhance the flavour (Leakey 2012). However, the most valuable part of the tree is the seed. The seeds are chewed to suppress headaches, laryngitis, bronchitis, malaria and gonorrhoea and have supposed aphrodisiac

effect on men (Adegboye et al. 2008). The seed extract is used as a cure for various inflammations or liver cirrhosis. Vast evidence has been found that bioactive components of the seeds can serve as an alternative medicine to treat/prevent severe illnesses such as malaria, hepatitis and immune-destructive diseases (Maňourová et al. 2019).

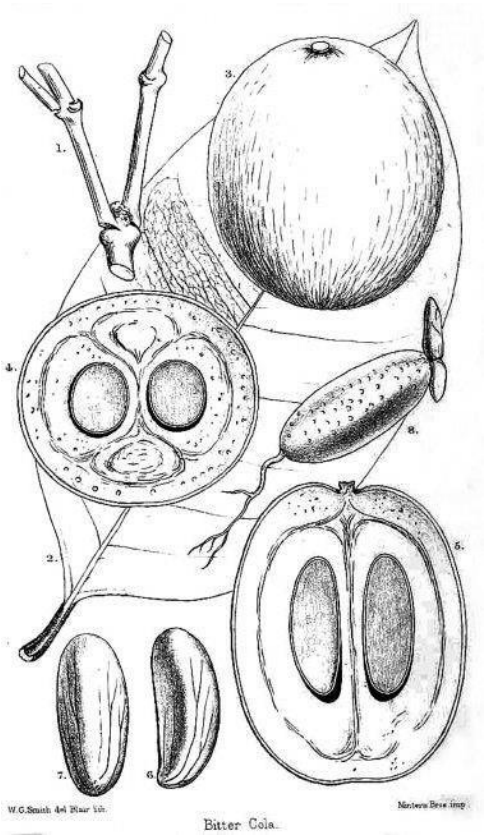

Figure 11. *Garcinia kola*, Maňourová

Figure 12. *Garcinia kola*: 1 - twig; 2 - leaf; 3 - fruit; 4, 5 - seeds disposition in the fruit; 6, 7 - seeds; 8 - germinated seedling, source: World Agroforestry Centre

Irvingia gabonensis

Irvingia gabonensis (Aubry-Lecomte ex O'Rorke) Baill. (Figure 13, 14), usually known by its common name bush mango, belong to Irvingiaceae family. This family has in total nine species and all of them occur in West and Central Africa. This large evergreen tree reaching up to 35 m, can be both found in the wild stands and on farm, where it provides shade for cocoa and coffee (Leakey 2012). There are two main varieties of *Irvingia:* the sweet one (*I.gabonensis*), which is more common in Cameroon and bitter one (*I.wombolu*) favoured in Nigeria (Okafor 1975).

I. gabonensis is known for its edible fleshy fruits. However, it is a multipurpose tree and it also used as a source of timber, utensils, foods and medicine. Bush mango shows beneficial effect for diabetes and obesity reduction, and is popular for its analgesic, antimicrobial, antioxidant and gastro-intentestinal effects as well. Traditionally, bark, kernels, leaves and roots (Okoronkwo et al. 2014). The fruit of this tree is a drupe, green in colour ripening into yellow, the flesh is stringy yellow, carrying large and woody seed. The fruits are predominantly consumed fresh but can be used in preparation of juice, jelly, jam and wine. However, the most valued part is the kernel, processed by grinding and crushing, and then used to thicken soups and stews. An edible oil is extracted from the seeds used in cooking. Kernels are traded extensively; Cameroon is probably the main exporter (Anegbeh et al. 2003).

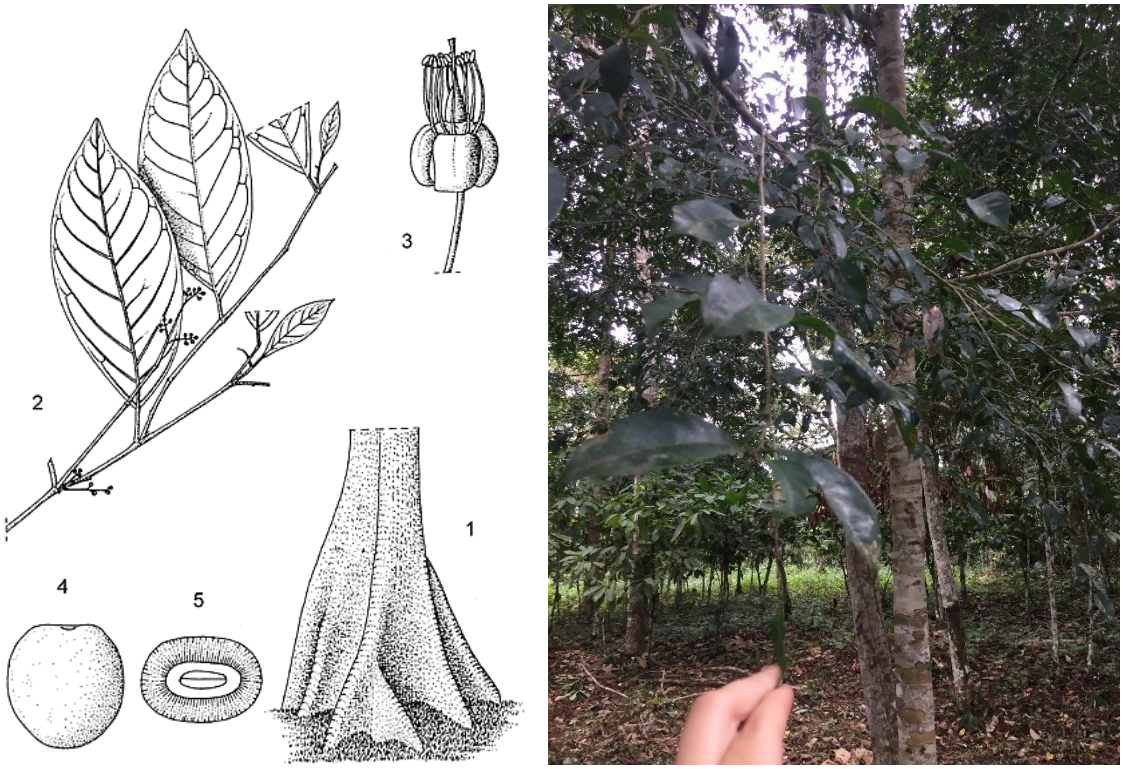

Figure 13. *Irvingia gabonensis*, author's archive

Figure 14. *Irvingia gabonensis:* - base of bole; 2 - flowering twig; 3 - flower; 4 - fruit; 5 - fruit in cross section. Redrawn and adapted by Achmad Satiri Nurhaman, PROTA

Ricinodendron heudelotii

Ricinodenron heudelotii (Baill.) Heckel (Figure 15, 16) belongs to the family *Euphorbiaceae* and locally is called Njangsang. This fast-growing, light-demanding tree

is native to Africa's tropics and subtropics areas and can reach up to 50 m in height, but usually the height varies between 20 – 30 metres. Njangsang can be easily recognized by its thick buttresses often extending into heavy superficial roots. In Cameroon, the tree is widely distributed in the Centre, South and West regions (Oyono et al. 2014).

Many parts of the tree are used as medicine. The root and the bark of the stem extract is used against cough, as a poison antidote and for the treatment of intestinal diseases. Bark is also used to treat yellow fever, malaria, stomach pains, painful menstruation, to prevent miscarriage, cure infertility in women and much more (Oyono et al. 2014). However, the most important part are the kernels, which have a great nutritional value due to the high protein content. The seeds are widely used in cooking in the whole West and Central African region. An edible oil is extracted from the seeds and a paste is made by crushing of the dried kernels. Then the paste is further used as a flavouring and thickening agent for soups and stews. Njangsang kernels are traded locally as well as in national and regional market, the kernels often ensure an additional income to farmers, when the cocoa beans price is fluctuating (Leakey 2012).

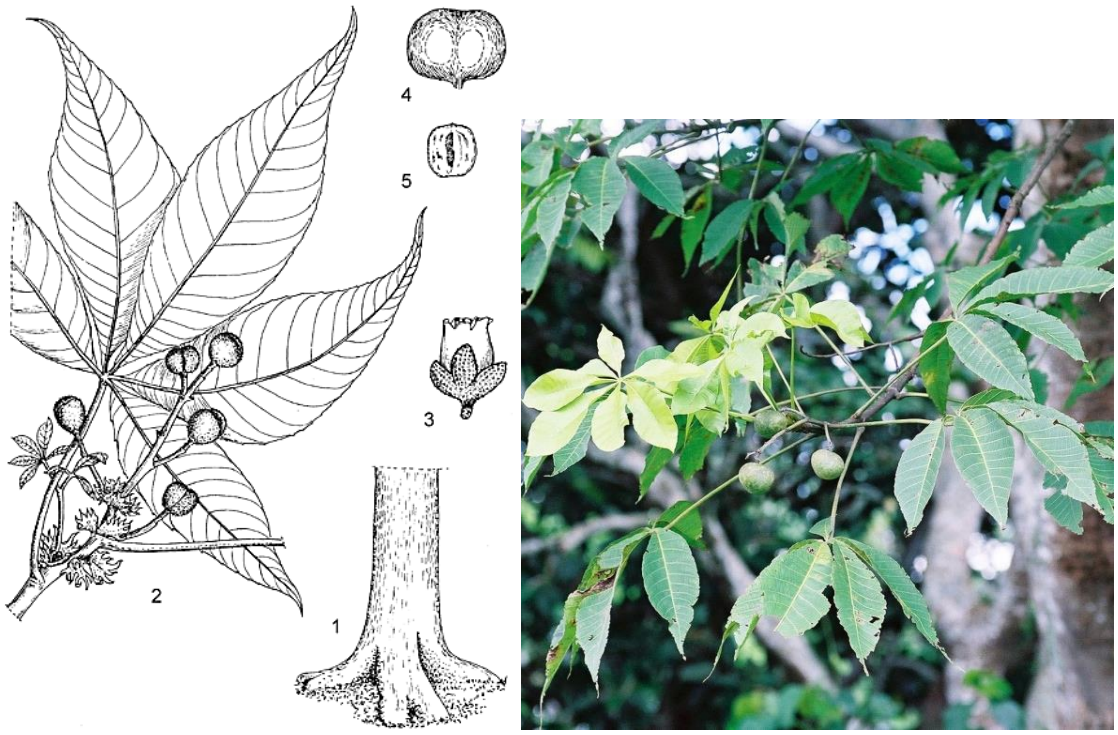

Figure 15. *Ricinodendron heudelotii*, tropical.theferns.info

Figure 16. *Ricinodendron heudelotii*: 1 - base of bole; 2 - part of branch with young fruits; 3 - male flower; 4 - fruit; 5 - seed. Redrawn and adapted by Iskak Syamsudin, PROTA

Spondias dulcis

Spondias dulcis Parkinson (Figure 17, 18) in Cameroon called Cas mango is a fruiting tree from the Anacardiaceae family. This fast-growing tree can reach up to 20 m in height, but usually is between 10 – 12 m tall. It has been introduced into tropical areas across the world. The tree is mainly cultivated for its edible fruits containing a fibrous pit, which has a considerable amount of sugars, vitamins and polyphenols. The fruit is green, turning into yellow when ripe. The species has medicinal properties as well, the roots, leaves, flowers, fruits and especially the bark are used in traditional medicine. It can also be used in flavouring other food products such as juices or tea (Zofou et al. 2019).

The fruit has an antioxidant, antimicrobial, cytotoxic and thrombolytic potential. It is used to cure itchiness, sore throat and skin inflammations. The fruit is used to treat anemia, regulate blood glucose levels, and to treat digestive problems as it contains high amount of dietary fibre (Jayarathna et al. 2020). Cas mango is also used in eyesight enhancement and to treat eye infections. Bark is used as a remedy for diarrhea (Islam et al. 2013). The leaves are used as an antiseptic, antidiarrheal and for the treatment of bronchitis, stomatitis, stomach pain, eye infection and diabetes (Fernandes et al. 2018).

Figure 17. *Spondias dulcis*, author's archive

Figure 18. *Spondias dulcis*, Little & Wadsworth – plantillustrations.org

2.5.2. Forest species

Forest plays a crucial role in Cameroon's economy and everyday life of local communities, which use the trees for building houses, fences, boats and fuel wood. Cameroon has the second largest forest estate from all African countries. The tropical forest composes of various types of trees, providing a high-quality timber that are traded even internationally. Unfortunately, in the last decades there has been a significant increase in illegal logging (Alemagi & Kozak, 2010).

Cola pachycarpa

Cola pachycarpa K.Schum. (Figure19, 20) is underutilized indigenous tropical fruit species found in humid West and Central African forests. It is a part of the *Cola* genus, family Malvaceae and usually occurs in understorey of taller forest species. The tree usually grows up to 10 m, it has very few branches with leaves crowded at the top of the stem. The tree is known by its common name monkey cola and it is a wild relative of the previously mentioned kola nut. Native people of southern Nigeria and Cameroon relish this edible tasty fruit as well as some wild primate animals, especially monkeys (Ogbu & Umeokechukwu 2014).

The fruits are composed of up to 6 large boat-shaped, shiny, pink or red glabrous carpels containing few large seeds, which are enclosed in white aril. The aril is the consumed part (Nwiisuator et al. 2012).

Not much is known about this species, mainly because local people do not see the need in planting such species, they prefer to collect the fruits in the wild. Therefore, the trees are disappearing at a faster rate than the nature can replenish them. Especially when local people are harvesting the trees by cutting down the whole stem (Ogbu & Umeokechukwu 2014).

Figure 19. *Cola pachycarpa*, author's archive

Figure 20. *C. pachycarpa:* fruit, author's archive

Lovoa trichilioides

Lovoa trichilioides Harms (Figure 21, 22) from Meliaceae family is an important commercial hardwood timber tree found throughout the humid forest zone of Cameroon. Its natural distribution extends from Sierra Leone to Uganda (from west to east) and from Tanzania to Angola (from south to north western). The tree is also called African walnut, Dibetou and Bibolo. It is a large forest tree, reaching a height of 45 metres or more and diameter at the breast height often exceeds 1m (Tchoundjeu & Leakey 2001).

The wood is highly valued for furniture, flooring, interior trim, panelling and decorative veneer and plywood. It is also suitable for ship building, toys, carving and crates. Moreover, it is used as firewood and for charcoal production. Bibolo is included in the IUCN Red list of threatened species as vulnerable, because of the high rates of its exploitation (Nyunaï 2008) and other factors hampering its regeneration such as difficult collection of seeds and large-scale destruction of seedlings by shoot borers (Tchoundjeu & Leakey 2001).

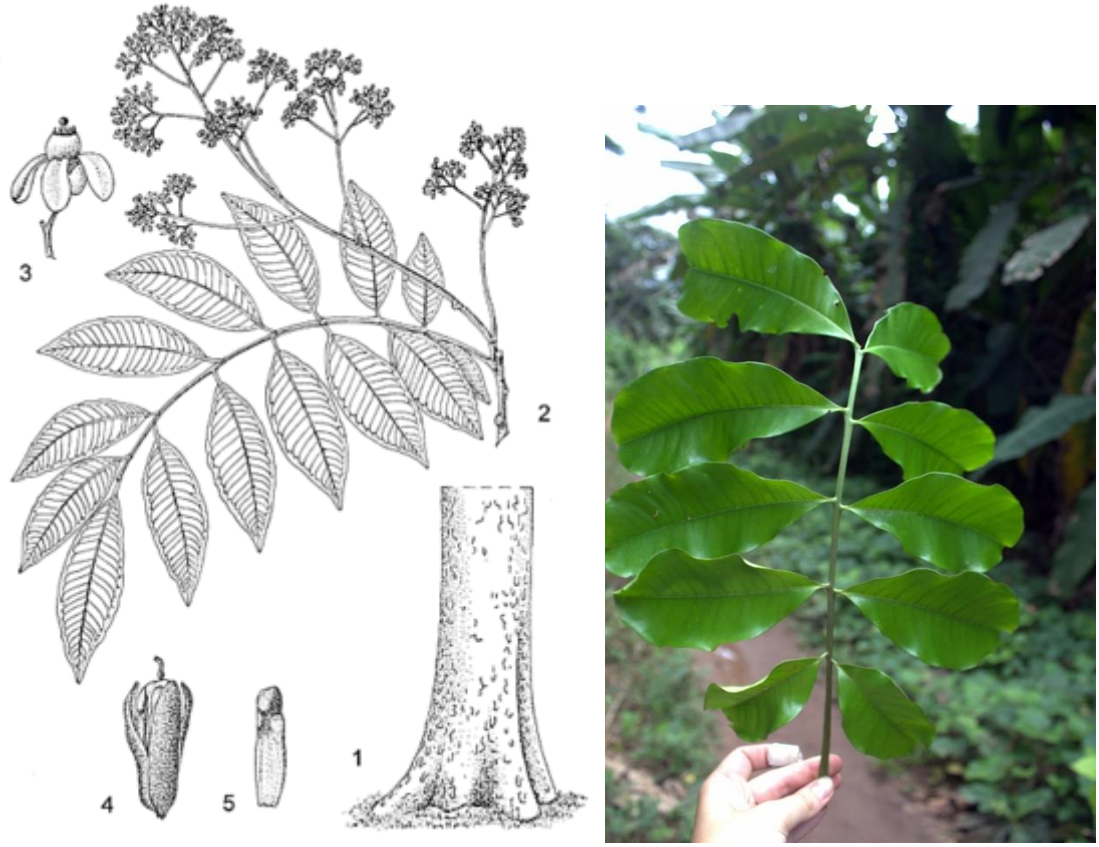

Figure 21. *Lovoa trichilioides*, author's archive

Figure 22. *Lovoa trichilioides*: 1 - base of bole; 2 - flowering twig; 3 - flower; 4 - fruit; 5 - seed. Redrawn and adapted by Achmad Satiri Nurhaman, PROTA

Triplochiton scleroxylon

Triplochiton scleroxylon K.Schum. (Figure 23, 24) also known as African whitewood or Ayous is a tree from the Malvaceae family. It is widely distributed in the West and Central African forest zone. This large, deciduous, forest tree reaches a height of 50 m and more, the trunk diameter is usually around 2 m. The species accounts as the highest timber volume extracted annually from the West African forests, mainly because of its fast growth, especially under the plantation conditions (Hall & Bada 1979).

The high-quality wood has multiple uses: interion joinery, panelling, furniture, boxes, sculptures, pencils, fibre and particle boards. It is often very important for house building and for roof shingles, the bole is used for dugout canoes and the wood pulp can be used for paper production. The trees also grow in cocoa plantations as a shade trees (Bosu & Krampah 2005).

Figure 23. *Triplochiton scleroxylon*, author's archive

Figure 24. *Triplochiton scleroxylon:* 1 - base of bole; 2 - flowering twig; 3 - fruit. Redrawn and adapted by Iskak Syamsudin, PROTA

2.5.3. Ornamental tree

Ornamental trees serve as an aesthetic component of a landscape. They are usually planted in gardens, parks or along roads. The ornament trees of tropical regions usually possess additional functions such as a timber or nitrogen fixing species, that enrich the soil fertility (Harmand et al. 2004).

Delonix regia

Delonix regia K.Schum. (Figure 25, 26) is a species of flowering plant from the family Fabaceae. In many tropical parts of the world it is grown as an ornamental tree and its English name flame tree is reflecting its beautiful large flowers with orange to red petals. This fast-growing tree usually reaches the height of 15 m and has wide spreading, umbrella shaped crown that shapes a diameter that is usuallywider than its height (Jensen 1995). Even though the species is one of the most widely cultivated ornamental plants in the world it has a lot more functions. This multipurpose tree can be harvested for

medicinal purpose, food, timber and fuel wood. It is cultivated as a shade tree in plantations and is used to stabilize and enrich the soil (Rahman et al. 2011).

The leaves, flowers, seed and bark contain a range of medically active compounds such as flavonoids, alkaloids, sterols, tannins, carotenoids and phenolic acids. The plant is reported to have antibacterial, antidiabetic, antidiarrheal, anti-inflammatory, antioxidant and other healing activities, explaining its frequent use in the traditional medicine (Azab et al. 2013).

Figure 25. *Delonix regia*, plantillustrations.org

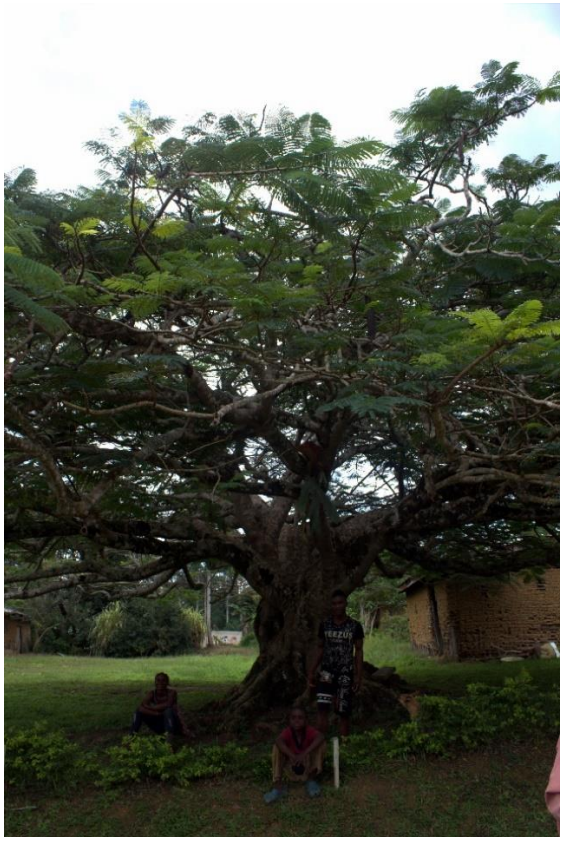

Figure 26. *Delonix regia*, author's archive

3. OBJECTIVES

The main objective of the thesis was to evaluate the use of the DNA barcoding method for the identification of tropical trees in humid tropical forests, particularly in Cameroon.

The specific objectives were following:

- To identify the trees both by traditional taxonomic methods and by molecular methods;
- To compare the results from both species identifications methods, and based on these results assess the reliability of the DNA barcoding method in practise;
- To compare the DNA extraction, PCR amplification and sequencing success of three different tree tissues types - leaf, young wood and old wood;

The main hypothesis of this study was, that all collected samples will be successfully identified both by taxonomic and molecular methods. It was assumed that DNA barcoding could successfully recognize collected samples in all cases at least to its family or genus level. The lower the quality of DNA, the lower the recognition. Therefore, the presumption was that the leaf material will yield more quality DNA and could be recognized to its species, whereas the wood material has lower quality of DNA and could be recognized only to its genus or family.

4. MATERIALS AND METHODS

4.1. Study area

The study was carried out in Cameroon, in the South and Central regions. The data collection was conducted nearby four bigger cities: Ebolowa, Sangmélima and Kye-Ossi (South region) and Mbalmayo (Central region) (Figure 27).

The main sampling was conducted in the South region, in the wild stands (humid forests) and in agroforestry systems (mostly homegardens) of local farmers. The additional sampling was done in the Central region in the World Agroforestry Centre (ICRAF) forest reserve in Mbalmayo.

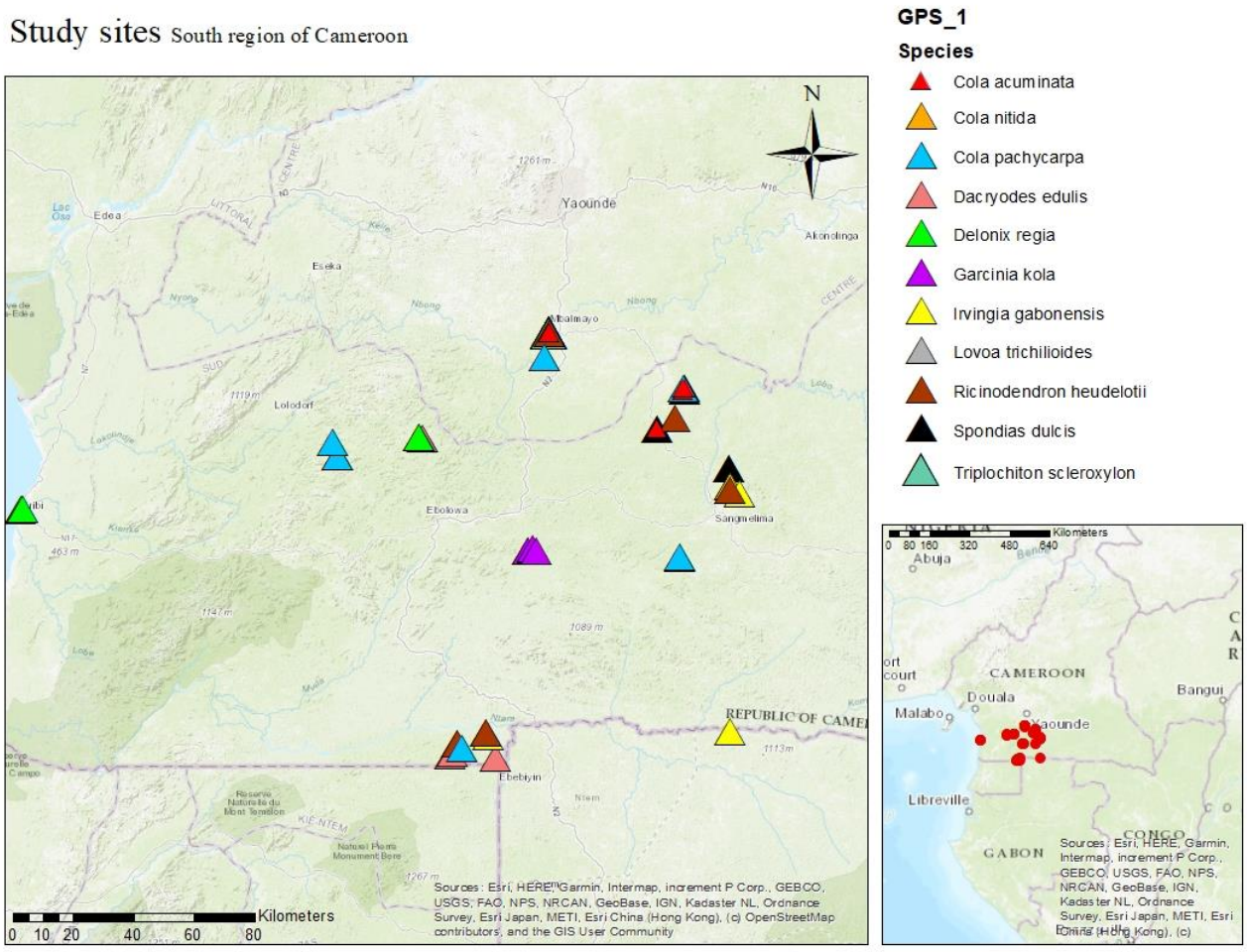

Figure 27. Map of Cameroon with marked study sites and collected samples

4.2. Sampling and data collection

The samples were collected during August and September 2019. With the help of experts from the ICRAF, 11 species of commonly grown trees in the area of South and Central region, have been selected. Out of the 11 species, seven are considered as fruit trees, three belong to forest species (two timber trees and one wild fruit tree) and one is an ornamental tree (Table 3).

From the selected species two sets of samples were collected, the first for the voucher specimen preparation and the second for the upcoming genetic analysis. Each species was marked by GPS and collected in six repetitions to test the reliability of the method. For the means of genetic analysis, three different types of tissue were sampled: leaves - six repetitions, young wood (twigs/cambium) – six repetitions and one sample of old wood for each species. In total, 150 samples have been collected. The fresh sampled material was immediately dried in silica gel. All the samples were brought back to the Czech Republic for further analyses.

Table 3. The list of collected species and number of collected samples

Characterization	**Scientific name**	**Leaf**	**Young wood**	**Old wood**	**Voucher specimen**
Fruit species	*Dacryodes edulis*	6	6	1	1
	Ricinodendron heudelotii	6	6	1	1
	Garcinia kola	6	6	1	1
	Cola acuminata	6	6	1	1
	Cola nitida	4	4	1	1
	Irvingia gabonensis	6	6	1	1
	Spondias dulcis	6	6	1	1
Forest species	*Cola pachycarpa*	6	6	1	1
	Triplochiton scleroxylon	6	6	1	1
	Lovoa trichilioides	6	6	1	1
Ornamental species	*Delonix regia*	6	6	1	1
	Samples sum	64	64	11	11
	Total				**150**

4.2.1. Tree selection and identification

Correct taxonomic identification was ensured in three sets of steps and controls. Noteworthy to say, the selected tree species are among the most commonly grown/cultivated in Cameroon and easy to be recognised even under the field circumstances. Before the field work, morphological characteristics of selected species were thoroughly studied in Yaoundé with the assistance of ICRAF botanists. In the field, the species were recognized with the help of local field-guides as well as by the owners of the trees. Additionally, the trees were checked by the author and the co-supervisor of this thesis. After the field identification, voucher specimens were collected for each species. The samples were preserved in newspapers, rinsed with 96% ethanol, flattened and dried in the plant press. When the specimens were dried, they were complemented by the pictures taken in the field and checked by Alain Tsobeng (ICRAF Assistant Scientist specialized in tree genetics) After the field work, voucher specimens were properly labelled, completed and stored in the herbarium of Czech University of Life Sciences in Prague. All the voucher specimens are displayed in Appendix A.

4.3. Genetic analyses

The DNA extraction, PCR amplification and the preparation for sequencing were performed in the Laboratory of Molecular Genetics at Faculty of Tropical AgriSciences, CZU. Afterward the samples were sent for Sanger sequencing provided by commercial company - SEQme s.r.o. and Microsynth AG.

4.3.1. Homogenization

The approach for sample homogenization differed according to the sample type (different tree tissue). For the homogenization of the leaf material, the tissue was ground manually with purified sand using a mortal and a pestle. The wood samples were ground using an oscillation mill. All of the components used (grinding jars, beads) were sterilized in following steps: cleaned in bleach, water and sterilized by ethanol and fire. Wood samples were cut into small shavings using sterilized scalpels, sterilization was done by 96% ethanol and fire. The shavings were transferred into 10 mL stainless-steel grinding jars with two stainless-steel grinding balls (Figure 28). Closed jars were put into liquid

nitrogen for 10 minutes, then the frozen jars were moved to the Mixer Mill (Retsch) apparatus and the grinding was set at maximum speed for 2 minutes (in the case of twigs/young wood), and for 5-10 minutes (in the case of old/hard wood). If needed, the last two steps were repeated until the material became fully homogenized in the form of fine wood powder.

Figure 28. Stainless-steel grinding jars and grinding ball, source: fishersci.co.uk

4.3.2. DNA extraction

In the previous study from 2018, two methods of DNA extraction were compared. The modified CTAB protocol (Doyle and Doyle, 1987; Faleiro, 2002) and DNA extraction using isolation kit (Qiagen DNeasy Mini Plant Kit. The CTAB (Cetyl trimethylammonium bromide) method was preferred, since it resulted higher nucleic acid concentration and better DNA purity (Legezová 2018).

50 mg of dry grounded material was mixed with 800 ul extraction buffer (CTAB 2.8%, NaCl 1.3 M, EDTA 20 mM, TRIS-HCl 100 mM, PVP 1%, mercaptoethanol 0.2%) and with 100 ng of Proteinase K for lysis of wood cells, heating the samples at 65°C for one hour while mixing them every 10 minutes. After cooling down of the sample to room temperature, to denature the contaminants in the samples, 700 ul of chloroform:isoamyalcohol (24:1) was added and mixed for 10 minutes, the phases were

separated by centrifugation for 10 minutes at14000 RPM and 4 °C. The supernatant, the upper phase, was transferred to a new microtube and 55ul of CTAB 7% was added. After that the chloroform:IAA extraction was repeated once more, to remove all the contaminants. The resulting supernatant was mixed with 900 ul of isopropanol in new tubes, which were placed in freezer (at -20°C) for one hour to allow the DNA to precipitate. After precipitation, the tubes were centrifuged for 10 minutes at 14000 RPM and 4 °C, the supernatant was discarded. To remove all the remaining salts, the pellet was washed twice in ethanol (96% and 70%). Subsequently, the pellet was dried at room temperature and dissolved in 100 ul of ddH_2O with addition of 30 ng of RNase. DNA was quantified using the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA), the concentration and quality of the extracted DNA was measured. DNA was diluted to 20ng/ul for the concertation to be equal in all samples. Until further processing the extracted DNA was stored in microtubes at -20 °C.

For the extraction of DNA from wood material, minor changes in the protocol were done, primarily to ensure higher purity of extracted DNA. Smaller amount of wood powder was added (about 30 mg). After lysis the phases were separated, the supernatant transferred to new microtube and only then the chloroform:IAA extraction was done. The chloroform:IAA extraction was done only once, with no repetition. The rest of the protocol was the same for all tissue material.

For the successful amplification and further processing of the samples, a threshold for the DNA purity was set. When using Nanodrop 2000 spectrophotometer the primary purity is assessed by the ratio of absorbance at 260nm and 280nm, the unit ~1.8 is generally accepted as “pure” DNA. Therefore, our threshold for acceptable DNA purity was set to 1.6. Thus, as a good isolate was considered a sample with the 260/280 value with 1.55 and higher. The process of DNA extraction was repeated and modified until the DNA purity reached near this unit. The ratio 260/230 is a secondary measure of nucleic acid purity and expected values are commonly in the range of 2.0-2.2.

4.3.3. PCR amplification

Currently the most suitable barcode loci for amplification of degraded wood samples are from plastid region, the *trnL* intron. The *trnL* intron gives more options, regarding four possible combinations of primer pairs. Figure 2. presents the location of the primers in the chloroplast *trnl* gene and Table 4. shows their sequences.

Table 4. Sequences of the two universal primer pairs amplifying the *trn*L intron

Name	Code	Sequence 5'-3'	Reference
c	A49325	CGAAATCGGTAGACGCTACG	Taberlet et al. (1991)
d	B49863	GGGGATAGAGGGACTTGAAC	Taberlet et al. (1991)
g	A49425	GGGCAATCCTGAGCCAA	Taberlet et al. (2007)
h	B49466	CCATTGAGTATCTGCACCTATC	Taberlet et al. (2007)

Source: Taberlet et al. 2007

Previously optimized regions from the *trnL* intron gene were amplified, in total there were three regions of different lengths, with working names short/medium/long region. The 'long' region is the total length of the intron with the use of *c* and *d* primers and it is 500bp long, the 'short' region is the P6 loop, it is only 50bp long and amplified using the *g* and *h* primer and the 'medium' region results from amplification using the combination of primer *c* and *h* and should be around 150bp long. DNA amplifications were carried out in a final volume of 25ul (Table 5).

For primer pair *c-d* and *c-h* the cycle was identical. The amplification started with initial denaturation at 95°C for 15minutes, followed by 35 cycles at 95°C for 30 seconds, annealing at 50°C for 30 seconds, extension at 72°C for 2 minutes was followed by the final extension step at 72°C for 10 minutes. For the very short *g-h* region the amplification was done according to Valentini et al. (2010), where they have removed the extension step. Therefore, the amplification was as follows: denaturation at 95 °C for 15 minutes, followed by 35 cycles of 30 seconds at 95 °C, and 30 seconds at 55 °C.

The amplified products were displayed on 1% agarose gel, stained with ethidium bromide and visualized under UV light. Electrophoresis ran for 45 minutes at 120 V. DNA ladder used on this gel was graduated by 50 bp.

Table 5. PCR mixture

PCR MasterMix - Qiagen	12.5 µL
primer *c/g* forward	1 µL
primer *d/h* reverse	1 µL
H_2O	7.5 µL
DNA	3 µL
	25 µL

Amplified samples were cleaned using Qiagen's kit QIAquick PCR Purification Kit following manufacturer's instructions. Purified PCR products were once more quantified using NanoDrop 2000 spectrophotometer, afterwards they were prepared for Sanger Sequencing according to specification of the sequencing company.

4.3.4. Evaluation of results

Sequences were displayed in Geneious Prime software. The sequences were searched in the NCBI GenBank database using a BLAST (Basic Local Alignment Search Tool) search, the database is searched to find close matches. More specifically a BLAST search against the Nucleotide collection (nr/nt).

Few criteria were considered in order to find the best match. The most important measurement was the Bit-score. Bit-score indicates how good is the alignment, the higher the score, the better the alignment. This score is calculated by formula that accounts for the alignment of similar or identical residues, as well as any gaps introduced to align the sequences (Geneious 2020).

Search hits can be also sorted by other values such as E value, pairwise identity, identical sites, query coverage and grade score. Pairwise identity is measured in percent and indicates how similar are the sequences found in database to the one sequence used as a query. Grade score is a percentage calculated by Geneious by combining query coverage, e-value and identity values. This allows for sorting the hits, the longest and highest identity hits are at the top (Geneious 2020).

5. Results

5.1. DNA extraction

Total of 139 samples was intended for DNA analyses. DNA was successfully extracted from all of these samples. The average of DNA concentration and quality per tissue type as well as their standard deviation are displayed in Table 6, the full version is available in Appendix B. As the quantity measurement was considered the nucleic acid concentration and as the quality measurement the purity indicator, the 260/280 ratio.

The DNA concentration of leaf material ranged from 130 ng/µl to 1500 ng/µl, therefore the average is 731.6 ng/µl, with standard deviation of 495.5 ng/µl, which represents the wide variation of values. In the case of young wood samples, the variation was also large, but the DNA concentration ranged from lower values of 20 ng/µl to 1000 ng/µl. Hence, the average is 304.8 ng/µl with a wide disperse of values represented by a standard deviation of 308.5 ng/µl. The DNA concentration values of isolates from old wood were also diverse, from 57 ng/µl to 1200 ng/µl, thus the average is 583.8 ng/µl and the standard deviation is 422.4 ng/µl.

The extraction from young wood material yielded only one third of good isolates, with the 260/280 purity ration near the 1.6 value, in some cases the extraction was repeated up to 4 times and still the purity was out of the defined range. Moreover, in some samples it was not possible to get to better purity than 1.3. However, the extraction of DNA from the leave samples was repeated only once and the repetition was necessary only in one third of the samples. The secondary purity measurement was lower than the required value in all cases. Both the primary and secondary purity as well as the DNA concentration was lower in the wood samples. The leave samples yielded the best quality as well as the best quantity.

Table 6. DNA quantification using NanoDrop 200 spectrophotometer

Sample type	Nucleic Acid Conc. (average)	Nucleic Acid Conc. (standard deviation)	260/280 (average)	260/280 (standard deviation)	260/230 (average)	260/230 (standard deviation)
Leaves	731.6 ng/µl	495.5	1.81	0.11	1.73	3.7
Young wood	304.8 ng/µl	308.5	1.67	0.15	0.82	0.4
Old wood	583.8 ng/µl	422.4	1.74	0.2	0.81	0.22

5.2. PCR amplification

All samples of three different tissue types were amplified using all three regions: the long one (*c*-*d*), the short one (*g*-*h*) and the medium one (*c*-*h*). Which resulted in 417 amplicons.

When visualised by gel electrophoresis, DNA appeared degraded in the 'long' region. The fragmentation is more visible in the case of wood samples, the fragment size ranged from approximately 100 bp to 700 bp (Figure 29). In the case of leave samples, the fragmentation is less severe, and the bands are more visible around the length of 500 – 600 base pairs. The short region's length is around 100bp and the medium region's length is around 200 bp.

Therefore, based on the visualization of the amplified samples, the long region (*c*-*d*) was excluded from further processing, due to the high fragmentation.

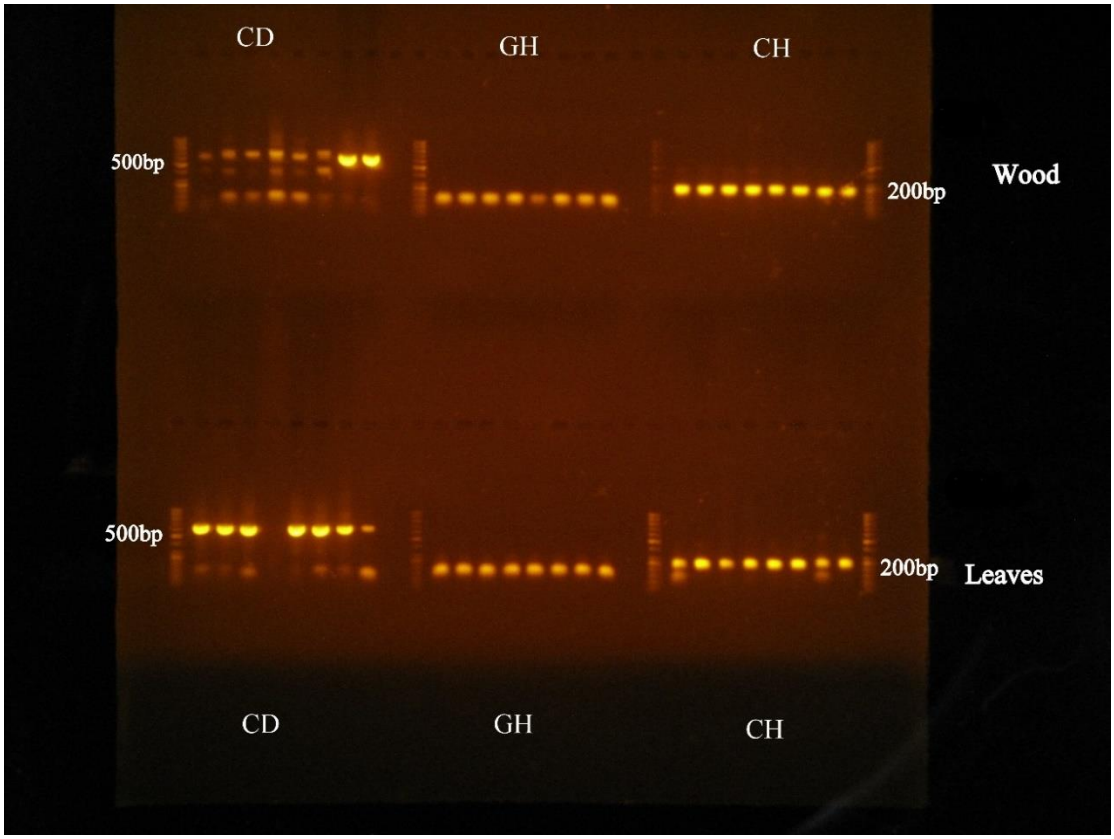

Figure 29. Electrophoresis of PCR products

5.3. Sequencing

24 representative samples were chosen, amplified in two different regions (the medium (*c-h*) and the short (*g-h*) region) and sent for Sanger sequencing. Therefore, the first batch sent for sequencing consisted of PCR products. The medium region showed considerably better results: the sequencing success rate was better by 30%. Therefore, all of the samples were subsequently amplified only using the medium region by the *trnL_G* and *trnL_H* primers and sent for sequencing.

In total, of the 139 samples sequenced, 128 resulted in applicable sequences. Therefore, the sequencing success rate was 92%. From all the successfully sequenced samples, 100% of them were correctly recognized to their family, 48% to the correct genus and 23% of them to their species (Table 7).

Because the species were precisely taxonomically identified, we believe that the sequences are also correct. Therefore, for all 11 studied species, the highest quality consensus sequences were selected for GenBank database submission. The sequences with annotation were successfully accepted and included in the GenBank database. Where

they can be found under their accession number and used for future barcode identifications (Table 8).

The sequence quality and successful identification did not differ that much in various tissue material (Table 7). However, even though the difference was small, it can be said that the higher the DNA quality of the material, the better the resolution to higher taxa e.g. genus, species. Table 9 represents a summarized version of the comparison of morphological identification results and molecular identification results; the full version is displayed in Appendix C.

Table 7. Sequencing success of resolution to higher taxa

	Genus	Species
Leaves	52%	24%
Young wood	44%	24%
Old wood	45%	18%
Overall	**48%**	**23%**

Table 8. Species and *trnL* sequences accession numbers submitted to GenBank

Species	**Accession number**
Garcinia kola	MT876415
Dacryodes edulis	MT876416
Irvingia gabonensis	MT876421
Ricinodendron heudelotii	MT876422
Cola pachycarpa	MT876417
Triplochiton scleroxylon	MT876418
Lovoa trichilioides	MT876419
Delonix regia	MT876420
Cola acuminata	MT876414
Spondias dulcis	MT876423
Cola nitida	MT876424

Table 9. Summarized sequencing results

	Morphological recognition		*trnL* barcode identification		
	Scientific name - Species	**Family**	**Scientific name - Species**	**Family**	**Bit-score**
Leaves	*Garcinia kola*	Clusiaceae	*Garcinia gummi-gutta*	Clusiaceae	281.811
Young wood	*Garcinia kola*	Clusiaceae	*Garcinia gumni-guta*	Clusiaceae	268.884
Old wood	*Garcinia kola*	Clusiaceae	*Garcinia mangostana*	Clusiaceae	185.785
Leaves	*Dacryodes edulis*	Burseraceae	*Dacryodes rostrata*	Burseraceae	265.191
Young wood	*Dacryodes edulis*	Burseraceae	*Canarium album*	Burseraceae	270.731
Old wood	*Dacryodes edulis*	Burseraceae	*Dacryodes rostrata*	Burseraceae	263.345
Leaves	*Irvingia gabonensis*	Irvingiaceae	*Klainedoxa gabonensis*	Irvingiaceae	268.884
Young wood	*Irvingia gabonensis*	Irvingiaceae	*Klainedoxa gabonensis*	Irvingiaceae	276.271
Old wood	*Irvingia gabonensis*	Irvingiaceae	*Klainedoxa gabonensis*	Irvingiaceae	191.325
Leaves	*Ricinodendron heudelotii*	Euphorbiaceae	*Ricinodendron heudelotii*	Euphorbiaceae	278.118
Young wood	*Ricinodendron heudelotii*	Euphorbiaceae	*Ricinodendron heudelotii*	Euphorbiaceae	279.964
Old wood	*Ricinodendron heudelotii*	Euphorbiaceae	*Klainedoxa gabonensis*	Euphorbiaceae	191.325
Leaves	*Cola pachycarpa*	Malvaceae	*Eriotheca discolor*	Malvaceae	289.198
Young wood	*Cola pachycarpa*	Malvaceae	*Eriotheca discolor*	Malvaceae	287.351
Old wood	*Cola pachycarpa*	Malvaceae	*Eriotheca discolor*	Malvaceae	278.118
Leaves	*Triplochiton scleroxylon*	Malvaceae	*Bombax ceiba*	Malvaceae	243.031
Young wood	*Triplochiton scleroxylon*	Malvaceae	*Bombax ceiba*	Malvaceae	252.265
Old wood	*Triplochiton scleroxylon*	Malvaceae	*Bombax ceiba*	Malvaceae	219.025
Leaves	*Lovoa trichilioides*	Meliaceae	*Lovoa trichilioides*	Meliaceae	252.265
Young wood	*Lovoa trichilioides*	Meliaceae	*Lovoa trichilioides*	Meliaceae	270.731
Old wood	*Lovoa trichilioides*	Meliaceae	*Lovoa trichilioides*	Meliaceae	239.338
Leaves	*Delonix regia*	Leguminosae	*Delonix regia*	Leguminosae	259.651
Young wood	*Delonix regia*	Leguminosae	*Delonix regia*	Leguminosae	274.424
Old wood	*Delonix regia*	Leguminosae	*Delonix regia*	Leguminosae	185.785
Leaves	*Cola acuminata*	Malvaceae	*Sterculia foetida*	Malvaceae	279.964
Young wood	*Cola acuminata*	Malvaceae	*Sterculia foetida*	Malvaceae	274.424
Old wood	*Cola acuminata*	Malvaceae	*Sterculia foetida*	Malvaceae	298.431
Leaves	*Spondias dulcis*	Anacardiaceae	*Spondias tuberosa*	Anacardiaceae	261.498
Young wood	*Spondias dulcis*	Anacardiaceae	*Spondias tuberosa*	Anacardiaceae	265.191
Old wood	*Spondias dulcis*	Anacardiaceae	*Spondias tuberosa*	Anacardiaceae	252.265
Leaves	*Cola nitida*	Malvaceae	*Sterculia foetida*	Malvaceae	272.578
Young wood	*Cola nitida*	Malvaceae	*Sterculia foetida*	Malvaceae	279.964
Old wood	*Cola nitida*	Malvaceae	*Sterculia foetida*	Malvaceae	239.338

6. Discussion

6.1. DNA extraction

Preparation of high-quality DNA is a crucial prerequisite for many subsequent downstream applications, including DNA barcoding. Many studies stated that tropical plants are rich in lipids, proteins and secondary metabolites, which complicates the nucleic acid separation, purification and therefore downstream applications (Colpaert et al. 2005, Gonzalez et al. 2009, Huang et al. 2013). Even though DNA was successfully isolated from all the samples with relatively high nucleic acid concentration (the average varies from 300 to 700 ng/μl), the purity was quite low. This supports the above-mentioned findings.

Many manufacturers are producing a number of kits addressing such problematic samples. Commercial DNA isolation kits have many advantages, such as shorter isolation steps, faster achievement of results and smaller amounts of chemicals used. On the other hand, costs of these kits are often high and usually the DNA yields were found to be lower than those obtained with conventional methods, such as CTAB (Akkurt 2012, Sousa et al. 2014, Stefanova et al. 2013). Pipan et al. (2018) as a result of their comparison of six chosen commercial kits, used for extraction of DNA from apple tree leaves, recommends the use of DNeasy Plant Pro Kit (Qiagen), which produced the purest product.

However, prior this study we have tried various commercial kits: DNeasy Plant Mini Kit (Qiagen), DNeasy Plant Pro Kit (Qiagen), NucleoSpin Plant (Macherey-Nagel) and others. These kits worked quite well in the case of leaf tissue but not when extracting DNA from wood tissue, both DNA concentration and purity were not sufficient enough. Even though CTAB method is very time consuming and requires the use of toxic substances, it yielded the best results (Legezová 2018). Different modifications of this method are also recommended by previously published studies (Colpaert et al. 2005, Huang et al. 2013).

The 260/280 purity ratio was lower that the ideal value, especially in the case of wood samples. However, this lower value was not that significant, indicating that a sample is probably contaminated by residual reagent used in the extraction protocol, or some other contaminant, which absorbs at 280 nm. Of course, it can also simply mean that the nucleic

acid concentration is very low. The secondary purity measurement, the 260/230 ration was very low, especially in the wood samples again. This low value can indicate the presence of organic contaminants, absorbing at 230 nm, such as TRIzol, chaotropic salts and other aromatic compounds. It may also be a result of carbohydrate carryover, which is a common problem in plant species (Wilfinger et al. 1997).

The young wood material is the most problematic tissue for DNA extraction. The isolation had to be repeated up to 4 times in some samples and around 10% of the samples did not reach the required purity threshold of 1.6. In comparison, the leaf material was the easiest to extract, yielding good nucleic acid concentration and the primary purity ratio was almost ideal in every case. Even though, the secondary purity measurement was lower, the difference was not that significant. Finally, it can be said, that when comparing different tissue types, DNA extraction from leaf material led to the best results. This outcome was expected, because the leaves are considered as the highest quality source of DNA from the different tissue types stated above.

In wood tissues, the plastids are differentiated in living parenchyma cells. However, the number of plastids in living parenchyma cells are not as high as the number of chloroplasts in photosynthetic tissues (Deguilloux et al. 2002). Therefore, the DNA concentration is lower, and the extraction is more complicated.

The lower nucleic concentration values should not be a problem for the success of subsequent PCR amplification. At least not in our case where the lowest values of nucleic acid concentration were around 20 ng/ul. In addition, it is usually better to have low DNA concentration but better purity than the other way around. Because, the low DNA purity is usually an indicator of some kind of contamination, which could inhibit the PCR amplification.

Even though the purity values were not ideal, the extraction was successful in all three tissue types.

6.2. PCR amplification

The recommended DNA barcode for plant species according to the CBOL Plant Working Group is the combination of two plastid coding regions *rbcL* and *matK*. However, this recommended barcode is not ideal for our purposes. The *rbcL* region is 599 bp long and limitation of this region is that it has low discriminatory power. The *matK* is even longer, 841 bp long and it is difficult to be PCR amplified (Hollingswort et al. 2011). Therefore, we were searching among the mini-barcodes, the *trnL* region compliments our needs well. The *trnL* intron is highly recommended in several studies on the identification of plant species, especially with the use of degraded DNAs (Taberlet et al. 2007, Mallot et al. 2018, Valentini et al. 2010).The chosen *trnL* approach was as well suggested by the results of our previous study (Legezová 2018). Despite the tree itself is a living organism, the wood is considered as a degraded material because it is composed of dead cells, thus consisting of degraded DNA. We also believe, that thanks to the *trnL* introns short length it could have wider use and it has the potential to become the universal barcode for plant species.

The isolated DNA was used for PCR amplification of three different overlapping *trnL* regions. The resulting sequence lengths corresponded with the length given by Taberlet et al. (1991, 2007) who designed these primers. However, different primer pairs yielded different amplification results quality. The most problematic amplification was using the *c* and *d* primers for the 'long' region, which appeared fragmented when visualized by electrophoresis. The fragmentation was even more severe in the case of wood samples. This can be explained by the fact, that the wood material is a poor source of plastid DNA and moreover the DNA is degraded and sequences longer than 500 bp are rarely possible to be amplified. The fragmentation is also visible in the case of leave samples. Even though the leaf material is not considered as a poor source of DNA or as degraded material. Nevertheless, the leaf tissue is often heavily defended against herbivors by high concentrations of anti-predation compounds, which may inhibit downstream applications, particularly PCR (Colpaert et al. 2005). The two shorter sequences were amplified uniformly with no significant fragmentation. However, when comparing these sequences by different tissue type, there can hardly be seen any differences.

Nevertheless, PCR amplification was successful, and the resulting product could be used for further processing. Hence, the two regions the ‘short’ and ‘medium’ were sent for sequencing.

6.3. Sequencing

The two sequenced regions differ in output quality. The sequencing success of the short (*g*-*h*) region was considerably lower than the success of the medium (*c*-*h*) region. Based on this result plus the fact mentioned by Little (2014) that the shorter the amplicon, the lower possibility of correct taxonomic discrimination, it has been decided to continue the sequencing only with the medium region. Noteworthy to say, this result was almost anticipated from our previous study, where it is mentioned that the long sequence may be too long for wood DNA samples and the short sequence may be too short for the interspecies variability (Legezová 2018). It was also suggested that a possible solution could be to optimize a medium region, which is shorth enough to avoid fragmentation and, at the same time, long enough to be interspecifically variable. Therefore, we are pleased that the current results are confirming our previous assumption.

The overall sequencing success was 92%, which can be considered as a very good result. The remaining 8% can be attributed to many factors, but it does not necessary mean that the method has an 8% error rate. For example, it is possible that some contamination by different DNA occurred in one of the steps of the analysis – DNA extraction, PCR amplification, purification of PCR products or the preparation for Sanger sequencing itself. Or the concentration of the PCR product was too low/too high. Moreover, since 8% of the not successful samples were in the same strip, we can also mark these unsuccessful results as a sign of human error.

The recognition to higher taxonomic ranks was rather low. From all the samples, 48% were assign to their correct genus and only 23% to its species level. Nevertheless, 100% were recognized to the correct family. This can be explained by the fact that the unidentified species do not have a listed *trnL* sequence in the GenBank database. Although, the *trnL* barcode has been used for more than 20 years and has about 18,200 sequences in the GenBank (Taberlet et al. 2007). Some species (*Garcinia kola, Irvingia gabonensis, Dacryodes edulis*) have only around 20-30 nucleotide sequences entries in

the database. Moreover, it consists mainly of microsatellites, or other barcodes such as *matK* or *rbcL*, as they are more often targeted as barcodes of plants. In the case of *Triplochiton scleroxylon, Cola nitida and Spondias dulcis,* there are only two or three listings and unfortunately none of them is the *trnL* intron sequence. *Cola pachycarpa* was not even listed as taxonomical record in the GenBank database, which can be explained by the fact that it is a wild forest species. The search of keyword *"Cola pachycarpa"* retrieved only two relevant results on Web of Science and only one on PubMed. Therefore, we can assume that the collected samples had lower recognition rate to higher taxa, because they are indigenous to West and Central Africa and are not extensively studied using DNA barcodes.

The previously stated assumption that the recognition ability decreases with DNA quality was not confirmed by these results. The sequencing success and the recognition to species and genus level were more or less equal for all tissue type. This discovery disproves our previously stated hypothesis, though it can be considered it as a positive result and an important finding. It implies, that there it is not necessary to sample the leaf material, which can be difficult to obtain, proving that the wood material is also reliable source of DNA. However, this result can be affected by the low number of successful identification to higher taxa.

7. Conclusion

This study aimed on the summary of the DNA barcoding knowledge and usage with the focus on DNA barcoding of tropical trees. 11 tree species were sampled and identified both by morphological and molecular methods. In total, 150 samples were collected, 11 intended for the creation of voucher specimen to serve as a control of the morphological identification and the remaining 139 samples were intended for genetic analyses. Three different tissue types were collected (leaf, young wood and old wood). For DNA barcoding *trnL* approach was used, as it was the most suitable DNA barcode for our purposes based both on the literature review and on previous experiences.

The DNA quality differed according to tissue material. Overall, the DNA extraction was easiest from the leaf material and also yielded the best results. The extraction from the wood material was quite difficult, requested many repeats and yielded lower quality. The PCR amplification was done in three different regions. The 'long' region (*c*-*d*) and 'short' region *(g*-*h*) and region that was suggested and optimized for this study the 'medium' region (*c*-*h*). The most successful was the medium region.

The sequencing resulted with 92% of applicable sequences and 100% taxonomic discrimination to the family level. However, the identification to higher taxa was quite low and resulted only in 48% of correct assignment to the genus and 23% to the species level. This low level of recognition is probably caused by the low number of listed nucleotide sequences of selected species in the GenBank database. The sequencing success and the recognition to higher taxa were more or less equal for all tissue type.

The main output of the thesis is the finding that DNA barcoding can be used for identification of tropical tree species, using different tissue materials. Even though the DNA extraction from wood tissue is more difficult the sequencing success was not significantly influenced by this fact. Which implies, that wood is also a good source of DNA for the means of species recognition by molecular methods.

Furthermore, 11 sequences of the studied tree species were submitted to GenBank, from which one of them did not have any previous record. For DNA barcoding to fulfil its goal of assigning individuals to species with 100% reliability, more sequencing has to be done and more sequences have to be submitted to the database.

8. References

Adegboye MF, Akinpelu DA, Okoh AI. 2008. The bioactive and phytochemical properties of Garcinia kola (Heckel) seed extract on some pathogens. African Journal of Biotechnology **7:** 3934–3938.

Ajibesin KK. 2011. *Dacryodes edulis* (G. Don) H.J. Lam: A review on its medicinal, phytochemical and economical properties. Research Journal of Medicinal Plant **5**: 32-41.

Akkurt M. 2012. Comparison between modified DNA extraction protocols and commercial isolation kits in grapevine (*Vitis vinifera* L.). Genetics and Molecular Research: GMR **11**:2343–2351.

Akinnifesi FK, Ajayi OC, Sileshi G, Kadzere I, Akinnifesi AI. 2007. Domesticating and Commercializing Indigenous Fruit and Nut Tree Crops for Food Security and Income Generation in Sub-Saharan Africa. Smart J, Haq N, editors. 5th International Symposium on New Crops and Uses: their role in a rapidly changing world. Southampton, UK: University of Southampton, p300-325.

Anegbeh PO, Usoro C, Ukafor V, Tchoundjeu Z, Leakey RRB, Schreckenberg K. 2003. Domestication of *Irvingia gabonensis*: 3. Phenotypic variation of fruits and kernels in a Nigerian village. Agroforestry Systems **58**: 213–218.

Arnot DE, Roper C, Bayoumi RA. 1993. Digital codes from hypervariable tandemly repeated DNA sequences in the Plasmodium falciparum circumsporozoite gene can genetically barcode isolates. Molecular and Biochemical Parasitology 16: 299-314.

Azab SS, Abdel-Daim M, Eldahshan OA. 2013. Phytochemical, cytotoxic, hepatoprotective and antioxidant properties of *Delonix regia* leaves extract. Medicinal Chemistry Research **22.9**: 4269-4277.

Baker RJ, Bradley RD. 2006. Speciation in mammals and the genetic concept. Journal od Mammalogy **87**: 643-662.

Barret RD, Hebert PDN. 2005. Identifying spiders through DNA barcodes. Canadian Journal of Zoology **83**: 481-491.

Becker N, Petrić D, Zgomba M, Boase C, Madon M, Dahl C, Kaiser A. 2010. Mosquitoes and their control. Springer-Verlag, Berlin.

Bergerovová E, Godálová Z, Siekel P. 2011. Combined effects of temperature, pressure and low pH on the amplification of DNA of plant derived foods. Czech journal of food science **29**: 337–345.

Bosu PP, Krampah E. 2005. Triplochiton scleroxylon K.Schum. PROTA (Plant Resources of Tropical Africa). Available at https://uses.plantnet-project.org/en/Triplochiton_scleroxylon_(PROTA): Accessed 2020-03-04.

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

Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP, Haidar N, Savolainen V. 2005. Land plants and DNA barcodes: short-term and long-term goals. Philosophical Transactions of the Royal Society B: Biological Sciences **360**: 1889-1895.

Chen WY, Hung TH, Shiao SF, 2004. Molecular identification of forensically important blow fly species (Diptera: Calliphoridae) in Taiwan. Journal of Medical Entomology **41**: 47–57.

Chorev M, Carmel L. 2012. The Function of Introns. Frontiers in Genetics **3**: 55.

Climate-Data. 2020. Climate-data.org. Available at: http://en.climate-data.org/: Accessed 2020-06-21.

Colpaert N, Cavers S, Bandou E, Caron H, Gheysen G, Lowe AJ. 2005. Sampling tissue for DNA analysis of trees: trunk cambium as an alternative to canopy leaves. Silvae Genetica **54**: 265-269.

De Queiroz K. 2007. Species concepts and species delimitation. Systematic biology **56**: 879-886.

De Wasseige C, Flynn J, Louppe D, Hiol Hiol F, Mayaux P. 2012. The Forests of the Congo Basin: State of the Forest 2010. Weyrich, Luxembourg.

Deguilloux MF, Pemonge MH, Petit RJ. 2002. Novel perspectives in wood certification and forensics: dry wood as a source of DNA. Proceedings of the Royal Society: Biology **269**: 1039-1046.

Duellman WE, Venegas P. 2005. Marsupial frogs (Anura : Hylidae : Gastrotheca) from the Andes of northern Peru with descriptions of two new species. Herpetologica **61**: 295-307.

Epule TE, Peng C, Lepage L, Chen Z. 2014. Policy options towards deforestation reduction in Cameroon: An analysis based on a systematic approach. Land use policy **36**: 405-415.

Eyebe A, Simeon AE, Angu KA, Endamana D. 2012. Integrating biodiversity conservation into national development policy: A case study of Cameroon. IIED Poverty Conservation Learning Group: London, UK.

FAO. 2006. Key Statistics of Food and Agriculture External Trade. Food and Agricultural Organization of the United Nations (FAO), Statistics Division. Available at

www.fao.org/es/ess/toptrade/trade.asp?dir=exp&country=3&ryear=2004: Accessed 2020-07-14.

Fernandez FHA, Boylan F, Salgado HRN. 2018. Quality standardization of herbal medicines of *Spondias dulcis* Parkinson using analytical and microbiological analysis. Journal of Thermal Analysis and Calorimetry **134**: 1923-1928.

Floyd R, Lima J, deWaard J, Humble L, Hanner R. 2010. Common goals: policy implications of DNA barcoding as a protocol for identification of arthropod pests. Biological Invasions **12**: 2947-2954.

Fondoun JM, Manga TT. 2000. Farmers indigenous practices for conserving Garcinia kola and Gnetum africanum in southern Cameroon. Agroforestry systems **48**: 289-302.

Foudjem-Tita D, D'Haese M, Speelman S, Degrande A, Gyau A, Van Damme P, Tchoundjeu Z, Van Huylenbroeck G. 2014. Would strictly enforced forestry regulations affect farmers' stated intentions to plant indigenous fruits trees? Insights from Cameroon. Food Policy **49**: 95-106.

Friedheim S. 2016. Comparison of Species Identification Methods: DNA Barcoding versus Morphological taxonomy. Manoa Horizons 1: 74-86.

Geneious Prime 2020.2 User Manual. 2020. Biomatters Ltd. Available at: https://assets.geneious.com/documentation/geneious/GeneiousPrimeManual.pdf Accessed 2020-07-05.

Gonzalez MA, Baraloto C, Engel J, Mori SA, Pétronelli P, Riéra B, Roger A, Thébaud C, Chave J. 2009. Identification of Amazonian Trees with DNA Barcodes. PLoS ONE 4 (e7483) DOI: 10.1371/journal.pone.0007483.

Hall JB, Bada SO. 1979. The Distribution and Ecology of Obeche (*Triplochiton scleroxylon*). Journal of Ecology **67**: 543-564.

Hebert PDN, Cywinska A, Ball SL, deWaard JR. 2003. Biological identification through DNA barcodes. Proceedings of the Royal Society of London B: Biological Sciences **270**: 313-321.

Hebert PDN, Gregory TR. 2005. The promise of DNA barcoding for taxonomy. Systematic Biology **54**: 852-859.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proceedings of the National Academy of Sciences **101**: 14812-14817.

Hillis DM. 1987. Molecular versus morphological approaches to systematics. Annual Review of Ecology and Systematics **18**: 24-42.

Hollingsworth PM, Graham SW, Little DP. 2011. Choosing and Using a Plant DNA Barcode. PLoS ONE 6 (e19254) DOI: 10.1371/journal.pone.0019254.

Huang QX, Wang XC, Kong H, Guo YL, Guo AP. 2013. An efficient DNA isolation method for tropical plants. African Journal of Biotechnology **12**:19.

Hubby JL, Lewontin RC. 1966. A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in Drosophila pseudoobscura. Genetics 54: 577-594.

Islam SMA, Ahmed KT, Manik MK, Wahid MA, Kamal CSI. 2013. A comparative study of the antioxidant, antimicrobial, cytotoxic and thrombolytic potential of the fruits and leaves of *Spondias dulcis*. Asian Pacific Journal of Tropical Biomedicine **3**: 682-691.

Jayarathna PLI, Jayawardena JAEC, Vanniarachchy MPG. 2020. Identification of Physical, Chemical Properties and Flavor Profile of *Spondias dulcis* in Three Maturity Stages. International Research Journal of Advanced Engineering and Science **5**: 208-211.

Jensen M. 1995. Trees commonly cultivated in Southeast Asia: An illustrated field guide. FAO Regional Office for Asia and the Pacific, Bangkok.

Jiao L, Yu M, Wiedenhoeft AC, He T, Li J, Liu B, Jiang X, Yin Y. 2018. DNA Barcode Authentication and Library Development for the Wood of Six Commercial *Pterocarpus* Species: the Critical Role of Xylarium Specimens. Scietific Reports **8**: 1945

Kress WJ, Erickson DL. 2007. A two-locus global DNA barcode for land plants: The coding rbcL gene complements the non-coding trnH-psbA spacer region. PLoS One 2 (e508) DOI: 10.1371/journal.pone.0000508.

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. 2005. Use of DNA barcodes to identify flowering plants. Proceedings of the National Academy of Sciences **102**: 8369-8374.

Leakey RRB, Van Damme P. 2014. The role of tree domestication in green market product value chain development. Forests, Trees and Livelihoods **23**: 116-126.

Leakey RRB. 2012. Living with the trees of life. Oxfordshire, CABI.

Legezová B. 2018. DNA barcoding of wood [BSc. Thesis]. Czech University of Life Sciences Prague, Prague.

Little DP. 2014. A DNA mini-barcode for land plants. Molecular Ecology Resources **14**: 437-446.

Lyal CHC. 2007. The effects of tropical forest management on biodiversity and ecosystem functioning [PhD Thesis]. University of Oxford, Oxford.

Mallet J, Willmott K. 2003. Taxonomy: renaissance or Towel of Babel? Trends in Ecology and Evolution **18**:57-59.

Mallet J. 2006. Species Concepts. Evolutionary Genetics: Concepts and Case Studies. Oxford University Press, Oxford.

Mallott EK, Garber PA, Malhi RS. 2018. trnL outperforms rbcL as a DNA metabarcoding marker when compared with the observed plant component of the diet of wild white-faced capuchins (Cebus capucinus, Primates). PloS one 13.6 (e0199556) DOI: 10.1371/journal.pone.0199556.

Maňourová A, Leuner O, Tchoundjeu Z, Van Damme P, Verner V, Přibyl O, Lojka B. 2019. Medicinal Potential, Utilization and Domestication Status of Bitter Kola (*Garcinia kola* Heckel) in West and Central Africa. Forests **10**: 124.

Mayr E. 1942. Systematics and the origin of species, from the viewpoint of a zoologist. Harvard University Press, Cambridge.

Mayr E. 1982. The growth of biological thought: Diversity, evolution, and inheritance. Harvard Univeristy Press, Cambridge.

McAndrew BJ, Majumdar KC. 1983. Tilapia stock identification using electrophoretic markers. Aquaculture 30: 249-261.

Meusnier I, Singer G, LandryJF, Hickey D, Hebert P, Hajibabaei M. 2008. A universal DNA mini-barcode for biodiversity analysis. BMC Genomics **9**:214.

Meyer CP, Paulay G. 2005. DNA barcoding: Error rates based on comprehensive sampling. PLoS Biology **3**: 2229-2238.

Niemenak N, Onomo PE, Fotso, Lieberei R, Ndoumou DO. 2008. Purine alkaloids and phenolic compounds in three *Cola* species and *Garcinia kola* grown in Cameroon. South African Journal of Botany **74**: 629–638.

Nixon KC, Wheeler QD. 1990. An Amplification of the Phylogenetic Species Concept. Cladistics **6**: 211-223.

Nwiisuator D, Oddo E, Emerhi EA, Owuno F, Sangha P. 2012. Mineral composition of *Cola parchycarpa* (K. Schum) Arils and Seeds. American Journal of Food and Nutrition **2**: 37-41.

Nyunaï N. 2008. Lovoa trichilioides Harms. PROTA (Plant Resources of Tropical Africa). Available at https://uses.plantnet-project.org/en/Lovoa_trichilioides_(PROTA): Accessed 2020-03-05.

Ogbu JU, Umeokechukwu CE. 2014. Aspects of Fruit Biology of Three Wild Edible Monkey Kola Species Fruits (*Cola* spp: Malvaceae). Annual Research & Review in Biology: **4**: 2007-2014.

Okafor JC. 1975. Varietal Delimitation in *Irvingia gabonensis* (Irvingiaceae). Bulletin du Jardin botanique national de Belgique **45**: 211-221.

Okoronkwo CU, Agoha EEC, Ogodo AC, Nwachukwu NO. 2014. Physical and Chemical Characteristics of the African Bush Mango (*Irvingia Gabonensis Var Garbonensis*) Seed Oil. International Journal Of Advances in Engineering and Management **1**: 28-31.

Oyono VA, Fokunang C, Assam Assam JP, Voundi S, Tsafack P, Mouafo ET, Ngandjui BT, Beng VP. 2014. Acute toxicity studies, antioxidant and in vitro antibacterial activities of extract from the barks of *Ricinodendron heudoletti* (Euphorbiaceae). Journal of Pharmacognosy and Phytotherapy **6**: 47-53.

Pace NR. 1997. A Molecular View of Microbial Diversity of the Biosphere. Science 276: 734-740.

Packer L, Gibbs J, Sheffield C, Hanner R. 2009. DNA barcoding and the mediocrity of morphology. Molecular Ecology Resources **9:**42-50.

Pečnikar ŽF, Buzan EV. 2014. 20 years since the introduction of DNA barcoding: from theory to application. Journal of applied genetics **55**: 43-52.

Pipan B, Zupančič M, Blatnik E, Dolničar P, Meglič V. 2018. Comparison of six genomic DNA extraction methods for molecular downstream applications of apple tree (Malus X domestica). Cogent Food & Agriculture (e1540094) DOI: 10.1080/23311932.2018.1540094

Plomion C, Leprovost G, Stokes A. 2001. Wood formation in trees. Plant physiology **127**: 1513-1523.

Ratnasingham S, Hebert PDN. 2007. BOLD: The Barcode of Life Data System (www.barcodinglife.org). Molecular Ecology Notes **7**: 355-364.

Řepková J. 2013. Genetika rostlin. Přírodověděcké fakulta, Masarykova univerzita, Brno. Available from https://is.muni.cz/do/rect/el/estud/prif/js13/genetika/web/index.html (accessed March 2020).

Robiglio V, Ngendakumana S, Gockowski J, Yemefack M, Tchienkoua M, Mbile P, Tchawa P, Tchoundjeu Z, Bolognesi M. 2010. REDUCING EMISSIONS FROM ALL LAND USES IN CAMEROON: FINAL NATIONAL REPORT. ASB Partnership for the Tropical Forest Margins. Nairobi, Kenya.

Robinson DR, Wu YM, Lonigro RJ, Vats P, Cobain E, Everett J, Cao X, Rabban E, Kumar-Sinha C, Raymond V, Schuetze S, Alva A, Siddiqui J, Chugh R, Worden F. 2017. Integrative clinical genomics of metastatic cancer. Nature 548: 297-303.

Rodrigues S, de Brito ES, de Oliveira Silva E. 2018. Exotic Fruits: Safou – *Dacryodes edulis*. Academic Press, Elsevier.

Saccone C, Degiorgi C, Gissi C, Pesole G, Reyes A. 1999. Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. Gene **238**:195–209.

Särkinen T, Staats M, Richardson JE, Cowan RS, Bakker FT. 2012. How to open the treasure chest? optimising DNA extraction from herbarium specimens. PLoS One 7 (e43808) DOI: 10.1371/journal.pone.0043808.

Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R. 2005. Towards the writing the encyclopaedia of life: an introduction to DNA barcoding. Philosophical Transactions of the Royal Society B: Biological Sciences **360**: 1805-1811.

Sloane DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR. 2012. Rapid Evolution of Enormous, Multichromosomal Genomes in Flowering Pant Mitochondria with Exceptioonally High Mutation Rates. PLoS Biology 10 (e1001241) DOI: 10.1371/journal.pbio.1001241.

Sonet G, Jordaens K, Braet Y, Bourguignon L, Dupont E, Backeljau T, De Meyer M, Desmyter S. 2013. Utility of GenBank and the Barcode of Life Data Systems (BOLD) for identification of forensically important Diptera from Belgium and France. ZooKeys **365**: 307-328.

Sousa, C. C., Gomes, S. O., Lopes, A. C. A., et al. (2014). Short communication comparison of methods to isolate DNA from Caesalpinia ferrea. Genetics and Molecular Research **13**:4486–4493.

Stefanova P, Taseva M, Georgieva T, Gotcheva V, Angelov A. 2013. A modified CTAB method for DNA extraction from soybean and meat products. Biotechnology & Biotechnological Equipment **27**: 3803–3810

Stoeckle M, Janzen DH, Hallwachs W, Hanken J, Baker J. 2003. Taxonomy, DNA, and the barcode of life. Draft Conference Report, Meeting held at Banbury Center, Cold Spring Harbor Laboratory, New York, NY, September 10–12 2003.

Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G, Brochmann C, Willerslev E. 2007. Power and limitations of the chloroplast *trn*L (UAA) intron for plant DNA barcoding. Nucleic Acids Research 35 (e14) DOI: 10.1093/nar/gkl938.

Taberlet P, Gielly L, Pautou G, Bouvet J. 1991. Universal primers for amplifications of three non-coding regions of chloroplast DNA. Plant Molecular Biology **17**: 1105-1109.

Tachie-Obeng E, Brown N. 2004. KOLA NUTS (*Cola acuminate* and *Cola nitida*). Clark LE, Sunderland TCH editors. The Key Non-Timber Forest Products of Central Africa: State of the Knowledge. Washington DC, USA: The Mitchell Group, p87-120.

Tang X, Zhao G, Ping L. 2011. Wood identification with PCR targeting noncoding chloroplast DNA. Plant Molecular Biology **77**: 609–617.

Tchoundjeu Z, Kengue J, Leakey RRB. 2002. Domestication of *Dacryodes edulis*: State of the art. Forests, trees and livelihoods 12: 3-13.

Tchoundjeu Z, Leakey RRB. 2001. VEGETATIVE PROPAGATION OF LOVOA TRICHILIOIDES: EFFECTS OF PROVENANCE, SUBSTRATE, AUXINS AND LEAF AREA. Journal of Tropical Forest Science **13**:116–129.

Teletchea F, Bernillon J, Duffraise M, Ladet V, Hänni C. 2008. Molecular identification of vertebrate species by oligonucleotide microarray in food and forensic samples. Journal of Applied Ecology **45**: 967-975.

Tropical Plants Database, Ken Fern. tropical.theferns.info. Available at http://tropical.theferns.info/viewtropical.php?id=Delonix+regia: Accessed 2020-08-02.

Valentini A, Miquel C, Taberlet P. 2010. DNA barcoding for Honey Biodiversity. Diversity **2**: 610-617.

Waugh J. 2007. DNA barcoding in animal species: progress, potential and pitfalls. Bioessays **29**: 188-197.

Wiemers M, Fiedler K. 2007. Does the DNA barcoding gap exist? – a case study in blue butterflies (Lepidoptera: Lycaenidae). Frontiers in Zoology 4: 8.

Wiley EO. 1978. The evolutionary species concept reconsidered. Systematic Zoology **27**: 17.

Wilfinger WW, Mackey K, Chomczynski P.1997. Effect of pH and Ionic Strength on the Spectrophotometric Assessment of Nucleic Acid Purity. BioTechniques **22**:474-481.

Will KW, Rubinoff D. 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics 20: 47-55.

Wilson EO. 2003. The encyclopaedia of life. Trends in Ecology & Evolution 18: 77-80.

World Agroforestry Centre. 2020. Available at http://worldagroforestry.org: Accessed 2020-08-02.

Zofou D, Shu GL, Foba-Tendo J, Tabouguia MO, Assob JCN. 2019. In vitro and in vivo anti-Salmonella evaluation of pectin extracts and hydrolysates from "Cas Mango" (Spondias dulcis). Evidence-Based Complementary and Alternative Medicine (e3578402) DOI:10.1155/2019/3578402

Appendices

List of the Appendices:

Appendix A: Voucher specimens

Irvingia gabonensis

Dacryodes edulis

Lovoa trichilioides

Triplochiton scleroxylon

Cola acuminata

Spondias dulcis

Ricinodendron heudelotii

Cola pachycarpa

Delonix regia


CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE
Faculty of Tropical AgriSciences
Department of Crop Sciences and Agroforestry
HERBARIUM
Collection: Cameroon 2019
Scientific name: Cola nitida (Vent.) Schott & Endl.
Synonym(s): Sterculia nitida Vent.
Family: Malvaceae
Vernacular name(s): Kola nut
Habitat (Locality): Mbalmayo, Cameroon
GPS: 3°28'00.0"N, 11°29'036"; elevation: 665 m
Description: Tree
Collected by: Barbora Legezová (Anna Maňourová)
Date collected: 25.9.2019
Collection No.: BL004
Signature:

Cola nitida

Garcinia kola

Appendix B: DNA isolation

	Leaves			Young wood		
	Nucleic Acid Concentration	260/280	260/230	Nucleic Acid Concentration	260/280	260/230
Garcinia kola	895.9	1.9	1.68	869.4	1.69	1.03
	1230.7	1.85	1.55	1097.4	1.75	1.25
	2079.8	1.75	1.07	533.6	1.81	0.81
	1689.4	1.63	0.87	66.2	1.79	1.71
	696.3	1.85	1.5	36.8	1.85	0.15
	736.5	1.8	1.53	63.9	1.82	1.85
Dacryodes edulis	541.5	1.73	1.1	69.1	1.67	1.06
	399.1	1.97	1.78	414.3	1.56	0.78
	272.4	1.67	0.98	159.5	1.3	0.35
	525.8	1.74	1.12	211	1.64	0.76
	889.5	1.81	1.4	34.8	1.79	1.6
	824	1.81	1.42	30.4	1.76	1.2
Irvingia gabonensis	830.7	1.87	1.66	2.7	1.31	0.18
	790.3	1.77	1.44	592.4	1.32	0.49
	1705.5	1.76	1.39	586	1.69	1
	1151	1.81	1.42	248.1	1.34	0.43
	1012	1.8	1.49	604.1	1.59	0.77
	1501.9	1.64	2.12	375.7	1.53	0.67
Ricinodendron heudelotii	1279.8	1.91	1.79	250.9	1.82	1.13
	132.2	1.89	1.62	1598.6	1.67	1.04
	421.2	1.82	1.32	192.1	1.54	0.87
	491.5	1.85	1.32	216	1.51	0.73
	1547.2	1.86	1.61	115.9	1.28	0.61
	1255.9	1.86	1.6	136.1	1.66	0.93
Cola pachycarpa	220.3	1.9	1.46	211.7	1.77	1.09
	794.8	1.74	1.09	64.8	1.58	0.5
	1066.2	1.85	1.28	138.3	1.58	0.44
	898.4	1.76	1.08	144.4	1.8	0.63
	305	1.78	1.01	99.6	1.7	0.48
	203.5	1.83	1.29	89.1	1.68	0.69
Triplochiton scleroxylon	316.7	1.93	1.77	29.1	1.7	0.41
	1079.3	1.79	1.08	24.2	1.8	0.43
	175.1	1.84	0.84	66.4	1.97	1.04
	981.8	1.66	0.9	953	1.82	1.42
	567.7	1.78	1	175.9	1.73	0.85
	412.9	1.77	1.05	181.5	1.78	1.04

	Leaves			Young wood		
	Nucleic Acid Concentration	**260/280**	**260/230**	Nucleic Acid Concentration	**260/280**	**260/230**
Lovoa trichilioides	568.1	1.77	1.08	393.8	1.79	0.99
	931.9	1.91	1.5	33.6	1.52	0.39
	546.3	1.92	1.41	95.9	1.89	0.95
	332.8	1.85	1.13	50.3	1.6	0.54
	881.4	1.9	1.35	106.3	1.76	0.8
	457.8	1.88	1.21	2.4	1.57	0.17
Delonix regia	388.4	1.83	1.05	770.3	1.78	1.2
	267.7	1.58	30.75	745	1.87	1.48
	1512.6	1.92	1.56	921.5	1.87	1.02
	722.4	1.89	1.42	318.4	1.8	1.09
	1628.3	1.99	1.91	168.8	1.74	0.78
	2192.3	2.08	2.15	214	1.6	0.63
Cola acuminata	208.3	1.81	0.81	33.1	1.61	0.56
	790.3	1.96	1.3	325.9	1.59	0.39
	338.2	1.96	1.51	238.4	1.74	0.73
	521.8	1.98	1.43	538.1	1.36	0.27
	679.3	1.69	0.73	24.7	1.52	0.33
	400.7	1.81	1.13	250.1	1.63	0.98
Spondias dulcis	500.4	1.57	0.69	590.7	1.64	0.84
	460.3	1.47	0.67	626	1.66	1.02
	1099.8	1.75	1.1	161	1.8	0.84
	100.8	1.66	0.66	511.2	1.65	0.87
	298.6	1.87	1.26	401.3	1.66	0.82
	143	1.81	0.91	109.6	1.88	1.15
Cola nitida	183.8	1.77	0.81	337.9	1.6	0.53
	221.6	1.84	0.76	408.8	1.76	0.9
	161.2	1.93	1.17	365.4	1.74	0.85
	362.7	1.84	0.94	194.7	1.8	0.83

	Old wood		
	Nucleic Acid Concentration	**260/280**	**260/230**
Garcinia kola	684.3	1.71	0.95
Dacryodes edulis	830.6	1.66	1.01
Irvingia gabonensis	1205.4	1.15	0.78
Ricinodendron heudelotii	112.6	1.74	0.71
Cola pachycarpa	439.4	1.67	0.68
Triplochiton scleroxylon	1036.7	1.88	1.25
Lovoa trichilioides	649.7	1.74	0.78
Delonix regia	71.2	1.71	0.67
Cola acuminata	57.6	1.46	0.43
Spondias dulcis	167.9	1.45	0.6
Cola nitida	116.7	1.85	1.05

Appendix C: Sequencing results

Morphological recognition		***trnL* barcode identification - Leaves**				***trnL* barcode identification - Young wood**		
Scientific name - Species	**Family**	**Scientific name - Species**	**Family**	**Bit-score**	**Grade**	**Scientific name - Species**	**Bit-score**	**Grade**
Garcinia kola	Clusiaceae	*Garcinia mangostana*	Clusiaceae	217.178	97.2%	*Garcinia gummi-gutta*	265.191	97.6%
Garcinia kola	Clusiaceae	*Garcinia gummi-gutta*	Clusiaceae	231.952	98.2%	*Garcinia gumni-guta*	261.498	97.6%
Garcinia kola	Clusiaceae	*Garcinia gummi-gutta*	Clusiaceae	281.811	98%	*Garcinia mangostana*	215.332	99.6%
Garcinia kola	Clusiaceae	*Garcinia gummi-gutta*	Clusiaceae	265.191	97%	*Garcinia gumni-guta*	268.884	98.2%
Garcinia kola	Clusiaceae	*Garcinia gummi-gutta*	Clusiaceae	276.271	97.4%	*Garcinia gumni-guta*	248.571	98.3%
Garcinia kola	Clusiaceae	*Garcinia gummi-gutta*	Clusiaceae	281.811	98%	*Garcinia gumni-guta*	235.645	97.7%
Dacryodes edulis	Burseraceae	*Dacryodes rostrata*	Burseraceae	252.265	100%	*Dacryodes rostrata*	274.424	99.7%
Dacryodes edulis	Burseraceae	*Dacryodes rostrata*	Burseraceae	265.191	100%	*Canarium album*	270.731	100%
Dacryodes edulis	Burseraceae	*Dacryodes rostrata*	Burseraceae	220.872	100%	*Canarium album*	270.731	100%
Dacryodes edulis	Burseraceae	*Dacryodes rostrata*	Burseraceae	265.191	100%	*Canarium album*	270.731	100%
Dacryodes edulis	Burseraceae	*Canarium album*	Burseraceae	268.884	100%	*Canarium album*	268.884	100%
Dacryodes edulis	Burseraceae	*Canarium album*	Burseraceae	268.884	100%	*Canarium album*	272.578	100%
Irvingia gabonensis	Irvingiaceae	*Klainedoxa gabonensis*	Irvingiaceae	261.498	99.6%	*Klainedoxa gabonensis*	267.038	100%
Irvingia gabonensis	Irvingiaceae	*Klainedoxa gabonensis*	Irvingiaceae	268.884	100%	*Klainedoxa gabonensis*	244.878	100%
Irvingia gabonensis	Irvingiaceae	*Klainedoxa gabonensis*	Irvingiaceae	268.884	100%	*Klainedoxa gabonensis*	276.271	100%
Irvingia gabonensis	Irvingiaceae	*Klainedoxa gabonensis*	Irvingiaceae	268.884	100%	*Klainedoxa gabonensis*	268.884	100%
Irvingia gabonensis	Irvingiaceae	*Klainedoxa gabonensis*	Irvingiaceae	263.345	100%	*Klainedoxa gabonensis*	274.424	100%
Irvingia gabonensis	Irvingiaceae	*Klainedoxa gabonensis*	Irvingiaceae	263.345	100%	*Klainedoxa gabonensis*	220.872	100%
Ricinodendron heudelotii	Euphorbiaceae	*Ricinodendron heudelotii*	Euphorbiaceae	248.571	100%	No results		
Ricinodendron heudelotii	Euphorbiaceae	*Ricinodendron heudelotii*	Euphorbiaceae	274.424	100%	*Ricinodendron heudelotii*	274.424	100%
Ricinodendron heudelotii	Euphorbiaceae	*Ricinodendron heudelotii*	Euphorbiaceae	278.118	100%	*Ricinodendron heudelotii*	274.424	100%

Morphological recognition		***trnL* barcode identification - Leaves**				***trnL* barcode identification - Young wood**		
Scientific name - Species	**Family**	**Scientific name - Species**	**Family**	**Bit-score**	**Grade**	**Scientific name - Species**	**Bit-score**	**Grade**
Ricinodendron heudelotii	Euphorbiaceae	*Ricinodendron heudelotii*	Euphorbiaceae	274.424	100%	*Ricinodendron heudelotii*	279.964	100%
Ricinodendron heudelotii	Euphorbiaceae	*Ricinodendron heudelotii*	Euphorbiaceae	278.118	100%	*Ricinodendron heudelotii*	195.019	100%
Ricinodendron heudelotii	Euphorbiaceae	*Ricinodendron heudelotii*	Euphorbiaceae	274.424	100%	*Ricinodendron heudelotii*	274.424	100%
Cola pachycarpa	Malvaceae	*Firmiana simplex*	Malvaceae	424.003	98.1%	*Pachira macrocarpa*	333.517	98.7%
Cola pachycarpa	Malvaceae	*Eriotheca discolor*	Malvaceae	285.504	98.9%	*Eriotheca discolor*	285.504	99.7%
Cola pachycarpa	Malvaceae	*Eriotheca discolor*	Malvaceae	283.658	99%	*Eriotheca discolor*	283.658	99.7%
Cola pachycarpa	Malvaceae	*Eriotheca discolor*	Malvaceae	289.198	99.7%	*Eriotheca discolor*	281.811	99.7%
Cola pachycarpa	Malvaceae	*Eriotheca discolor*	Malvaceae	287.351	99.7%	*Eriotheca discolor*	287.351	99.1%
Cola pachycarpa	Malvaceae	*Eriotheca discolor*	Malvaceae	289.198	99.1%	*Eriotheca discolor*	250.418	97.7%
Triplochiton scleroxylon	Malvaceae	*Bombax ceiba*	Malvaceae	243.031	97.1%	*Bombax ceiba*	241.185	97.5%
Triplochiton scleroxylon	Malvaceae					*Bombax ceiba*	248.571	97.4%
Triplochiton scleroxylon	Malvaceae					*Reevesia thyrsoidea*	213.485	100%
Triplochiton scleroxylon	Malvaceae	No results				*Bombax ceiba*	252.265	97.5%
Triplochiton scleroxylon	Malvaceae					*Bombax ceiba*	246.725	97.2%
Triplochiton scleroxylon	Malvaceae					*Bombax ceiba*	250.418	96.9%
Lovoa trichilioides	Meliaceae					*Lovoa trichilioides*	250.418	100%
Lovoa trichilioides	Meliaceae	No results				*Lovoa trichilioides*	265.191	100%
Lovoa trichilioides	Meliaceae					*Lovoa trichilioides*	270.731	100%
Lovoa trichilioides	Meliaceae					*Lovoa trichilioides*	265.191	100%
Lovoa trichilioides	Meliaceae	*Lovoa trichilioides*	Meliaceae	235.645	100%	*Lovoa trichilioides*	191.325	100%
Lovoa trichilioides	Meliaceae	*Lovoa trichilioides*	Meliaceae	252.265	100%	*Lovoa trichilioides*	220.872	100%
Delonix regia	Leguminosae	*Vachellia nilotica*	Leguminosae	250.418	99.6%	*Vachellia nilotica*	235.645	100%

Morphological recognition		***trnL* barcode identification - Leaves**				***trnL* barcode identification - Young wood**		
Scientific name - Species	**Family**	**Scientific name - Species**	**Family**	**Bit-score**	**Grade**	**Scientific name - Species**	**Bit-score**	**Grade**
Delonix regia	Leguminosae	*Delonix regia*	Leguminosae	257.805	100%	*Delonix regia*	268.884	100%
Delonix regia	Leguminosae	*Delonix regia*	Leguminosae	239.338	100%	*Delonix regia*	274.424	100%
Delonix regia	Leguminosae	*Delonix regia*	Leguminosae	248.571	100%	*Delonix regia*	274.424	100%
Delonix regia	Leguminosae	*Delonix regia*	Leguminosae	259.651	100%	*Senegalia catechu*	198.712	100%
Delonix regia	Leguminosae	*Delonix regia*	Leguminosae	255.958	100%	*Delonix regia*	279.964	100%
Cola acuminata	Malvaceae	*Sterculia foetida*	Malvaceae	274.424	98.1%	*Craigia yunnanensis*	217.178	99.6%
Cola acuminata	Malvaceae	*Sterculia foetida*	Malvaceae	274.424	98.1%	*Sterculia foetida*	274.424	98.1%
Cola acuminata	Malvaceae	*Sterculia foetida*	Malvaceae	274.424	98.1%	*Sterculia foetida*	274.424	98.1%
Cola acuminata	Malvaceae	*Sterculia foetida*	Malvaceae	274.424	98.1%	*Heritiera fomes*	274.424	98.3%
Cola acuminata	Malvaceae	*Sterculia foetida*	Malvaceae	270.731	99.7%	*Sterculia foetida*	274.424	98.1%
Cola acuminata	Malvaceae	*Sterculia foetida*	Malvaceae	279.964	99.7%	*Sterculia foetida*	274.424	98.1%
Spondias dulcis	Anacardiaceae	*Spondias tuberosa*	Anacardiaceae	261.498	100%	*Spondias tuberosa*	259.651	100%
Spondias dulcis	Anacardiaceae	*Spondias tuberosa*	Anacardiaceae	261.499	100%	*Spondias tuberosa*	219.025	99.6%
Spondias dulcis	Anacardiaceae	*Spondias tuberosa*	Anacardiaceae	257.805	100%	*Spondias tuberosa*	259.651	100%
Spondias dulcis	Anacardiaceae	*Spondias tuberosa*	Anacardiaceae	246.725	100%	*Spondias tuberosa*	265.191	100%
Spondias dulcis	Anacardiaceae	*Spondias tuberosa*	Anacardiaceae	233.798	100%	*Spondias tuberosa*	265.191	100%
Spondias dulcis	Anacardiaceae	No results	Anacardiaceae			*Spondias tuberosa*	265.191	100%
Cola nitida	Malvaceae	*Sterculia foetida*	Malvaceae	265.191	99.7%	*Sterculia foetida*	279.964	98.1%
Cola nitida	Malvaceae	*Heritiera fomes*	Malvaceae	283.658	99.1%	*Sterculia foetida*	279.964	98.1%
Cola nitida	Malvaceae	*Sterculia foetida*	Malvaceae	255.958	99.6%	*Sterculia foetida*	274.424	98.1%
Cola nitida	Malvaceae	*Sterculia foetida*	Malvaceae	272.578	99.7%	*Sterculia foetida*	274.424	98.1%

Morphological recognition		***trnL* barcode identification – Old wood**			
Scientific name - Species	**Family**	**Scientific name - Species**	**Family**	**Bit-score**	**Grade**
Garcinia kola	Clusiaceae	*Garcinia mangostana*	Clusiaceae	185.785	99.5%
Dacryodes edulis	Burseraceae	*Dacryodes rostrata*	Burseraceae	263.345	99.7%
Irvingia gabonensis	Irvingiaceae	*Klainedoxa gabonensis*	Irvingiaceae	191.325	100%
Ricinodendron heudelotii	Euphorbiaceae	*Jatropha curcas*	Euphorbiaceae	339.057	99.5%
Cola pachycarpa	Malvaceae	*Eriotheca discolor*	Malvaceae	278.118	99.4%
Triplochiton scleroxylon	Malvaceae	*Bombax ceiba*	Malvaceae	219.025	97.2%
Lovoa trichilioides	Meliaceae	*Lovoa trichilioides*	Meliaceae	239.338	100%
Delonix regia	Leguminosae	*Delonix regia*	Leguminosae	185.785	98.2%
Cola acuminata	Malvaceae	*Sterculia foetida*	Malvaceae	298.431	94.4%
Spondias dulcis	Anacardiaceae	*Spondias tuberosa*	Anacardiaceae	252.265	100%
Cola nitida	Malvaceae	*Sterculia foetida*	Malvaceae	239.338	99.6%