Czech University of Life Sciences Prague

Faculty of Tropical AgriSciences

Czech University of Life Sciences Prague
Faculty of Tropical AgriSciences

Barcoding of Tropical Woods

Bachelor thesis

Prague 2018

Supervised by:

doc. Ing. Bohdan Lojka, Ph.D.

Elaborated by:

Barbora Legezová

Declaration

I hereby declare that I have done this thesis entitled Barcoding of tropical woods 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, April 19th, 2018

…...............................

Barbora Legezová

Acknowledgements

First of all, I would like to sincerely thank my supervisor doc. Ing. Bohdan Lojka, Ph.D. for his overall help, comments, suggestions and useful information during his guidance of my thesis. Special thanks must go to Ing. Marie Kalousová for all her comments, corrections, also for the guidance during the laboratory analysis and for always being her nice, supportive and helpful self.

My thanks also belong to my co-supervisor Mgr. Petr Stiblík for introducing me to the laboratory work, teaching me all the right procedures and always being patient during this process, and also for introducing me to the barcoding problematic altogether. Next, I want to thank Mgr. Kateřina Votýpková Ph.D. for her quick and informative answers, for her help and all the advices and explanation concerning my study. And I have to thank to the whole Termite Research Team for enabling me to work with them and for everybody's help.

Furthermore, I have to thank my family for being patient and supportive while I was writing my thesis and always withstanding my complaining. The biggest thanks go to my best friend Valentýna Foytlová for being such an amazing language support and always helping me to finding the right words for what I wanted to say and more. Finally, to Jakub Foytl for being the vice-president of the language support family team.

Czech University of Life Sciences Prague
Faculty of Tropical AgriSciences

Termite
Research
Team

Abstract

The method of DNA barcoding was initially introduced in 2003 in the effort to facilitate the species identification. It is a molecular method of species identification based on a short region of DNA sequences. These DNA barcodes should be unique for each species. This method has since proven very effective mainly among animal species – however, further development is required for plant species. This thesis aims to describe the current knowledge of the barcoding technique, in the field of plant identification, with a special focus on tropical trees. Moreover, it aims to test the information gained by isolating DNA from dead wood material, select useful barcode loci for amplification, amplify the isolates by polymerase chain reaction (PCR) and verify the success of the reaction by agarose gel electrophoresis. This study evaluates the positive and negative aspects of the DNA barcoding method. It shows different plant barcodes proposed and provides examples of its successes in the use for wood barcoding. Moreover, this thesis explores the practical uses of DNA barcoding ranging from ecological studies, biosecurity, biomonitoring, biodiversity sciences, diet analysis, illegal trade and forensics, and databases in which barcode sequences are stored. The successes of DNA extractions were quantified using NanoDrop spectrophotometer and PCR amplifications were visualized using electrophoresis. Results show that out of the two DNA extraction methods (CTAB and DNeasy Mini Kit-Qiagen), the CTAB method isolates greater quantity of DNA, however both methods produced DNA of not ideal quality. Even though the poor quality of DNA, PCR amplification worked perfectly, therefore CTAB extraction was successful. The *trn*L approach chosen for this study has proven effective, PCR amplification worked well with chosen sequences. To improve this method, further optimization of DNA isolation may help with better results, mainly the optimization of 'medium sequence' (approximate length 400 base pair).

Key words: dead wood, DNA isolation, PCR, plant barcode, species identification, taxonomy, *trn*L approach

Table of contents

1. Introduction 1
2. Literature review 3
   2.1. Taxonomy 3
      2.1.1. Species definition and identification 3
      2.1.2. Species identification methods 4
   2.2. Plant genome 6
   2.3. DNA barcoding 8
   2.4. History of DNA barcoding 9
   2.5. Plant DNA Barcode 10
   2.6. DNA barcode for wood 13
   2.7. Practical uses of DNA barcoding 14
   2.8. Databases and other components of DNA barcoding 16
      2.8.1. Barcode of Life 16
      2.8.2. GenBank 17
3. Objectives 19
4. Methodology 20
   4.1. Plant materials for laboratory analysis 20
   4.2. Laboratory analysis 22
      4.2.1. DNA extraction 22
      4.3.1. PCR amplification 23
4. Results 26
   4.3. DNA extraction results 26
   4.4. PCR amplification results 28
5. Discussion 30
6. Conclusions 32
7. References 33

List of figures

Figure 1: Plant cell structure (p. 13)

Figure 2. A schematic representation of the potential strengths and weaknesses of source tissue (fresh, herbarium, xylarium) (p. 20)

Figure 3. Location of sampled trees (p. 26)

Figure 4. Position of the primers *c, d, g* and *h* on the chloroplast *trn*L (UAA) gene (p. 30)

Figure 5. Electrophoresis of leaf PCR products (p. 33)

Figure 6. Electrophoresis of wood PCR products (p.34)

List of tables

Table 1. Characteristics of different markers hat have routinely been included in plant barcoding studies (p. 18)

Table 2. Sequences of the two universal primer pairs amplifying the *trn*L (UAA) intron (p. 30)

Table 3. PCR mixture – wood (p.30)

Table 4. PCR mixture – leaves (p.30)

Table 5. fresh leaf material DNA quantification using NanoDrop 2000 spectrophotometer of DNeasy Mini Kit extraction method (p. 32)

Table 6. wood DNA quantification, comparation of CTAB and DNeasy Mini Kit (Qiagen) extraction method (p. 32)

List of abbreviations

COI……………………………………………….Cytochrome c oxidase I

*mat*K……………………………………………...Maturase K

*rbc*L……………………………………………….ribulose-bisphosphate carboxylase

CBOL………………………………………The Consortium for the Barcode of Life

BOLD……………………………………………Barcode of Life Data System

1. Introduction

Planet Earth is populated by millions of species; however, their discrimination is not an easy task. There are approximately 1.7 million species identified using morphological characters, but this figure may be a gross under-estimate of the true biological diversity on Earth, which is estimated to be 5 to 30 million species (Wilson 2003). It is common to discriminate species using morphological features such as colour, shape, pattern and size of the organism. Nevertheless, this approach has its limitations. Routine identifications are made using morphological keys, which are often only effective for a particular life stage or gender, thus many individuals cannot be identified. Furthermore, the use of the key requires a high level of expertise, and so misidentifications are common (Hebert et al. 2003). Identification of damaged or incomplete specimen, with only a small section of tissue available, renders morphological determination extremely difficult (Pečnikar & Buzan 2014).

These limitations inherent in specific morphology and decreasing number of skilled taxonomists showed that finding a new approach in taxa recognition was needed (Hebert et al. 2003). DNA barcoding is a relatively new method where each taxon can be uniquely branded by its genetic information. DNA barcoding is a revolutionary approach which is designed to provide rapid, accurate and automatable specimen identification using DNA sequences (Kress & Erickson 2008). These very short DNA sequences of a standardized genomic region can be viewed as genetic barcodes and can be identified in the same way that scanners in supermarkets distinguish products (Ratnasingham & Hebert 2007). DNA barcodes are sequences between 400 and 800 base pairs long that can be amplified using Polymerase Chain Reaction (PCR) and then sequenced. This approach can serve dual purpose: a new tool for taxonomists and a device for non-experts to objectively identify species. It can also help to discover and identify new species.

Although DNA barcoding is an important aid for taxonomic workflow, it cannot replace taxonomy altogether. There are cases where barcoding methods were proven to be more accurate and less expensive than the traditional morphological taxonomic survey (Thomson & Newmaster 2014), however comprehensive taxonomic analysis and molecular phylogenetics are still needed (Pečnikar & Buzan 2014). The methods of barcoding are already quite well developed for animals, however there is still quite a lot of limitations of using it for identification of plant species. Moreover, the DNA extraction

and amplification from the wood of trees, e.g. in a case where we do not have fresh leaf material, is still a big challenge.

The objectives of this thesis were i) to review the current state of knowledge and usage of DNA barcoding technique with focus on tropical trees ii) to select barcode loci for PCR amplification useful for identifying tropical trees and iii) to optimize isolation and PCR amplification of wood DNA for selected barcode loci.

2. Literature review

2.1. Taxonomy

Taxonomy is the science of naming, describing and classifying living things according to shared features. Using morphological, behavioural, biochemical and genetic observations, taxonomists can identify species and arrange them into categories (Lyal 2007). These categories are called taxa. A taxon, plural taxa, is defined as any unit used in the science of biological classification. Usually it is a group of organisms that have common characters, which differ from the other taxa. Taxa are usually named and arranged in a hierarchical ranking: kingdom, phylum, class, order, family, genus and species. The main aim of taxonomy is grouping organisms on the basis of mutual similarities into units, taxa. Species are the basic unit of classification and one of the fundamental units of biology (De Queiroz 2007).

2.1.1. Species definition and identification

The fundamental unit of biological diversity is usually considered to be species (Mayr 1982). However, it remains difficult to define. Several definitions and concepts were suggested, though only some have found 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 concept of the species was the typological concept (phenotypical) (e.g. Linnaeus 1758), according to which a species can be defined based on phenotypic characteristics of individual organism that do not occur in other species (Herbert & Gregory 2005). Such characteristics may include morphological, anatomical, physiological, biochemical or ethological features.

A concept that defines species with regard to the development in time is phylogenetic, which says that species is the smallest population of populations which has fixed heritable differences from other such populations (Nixon & Wheller 1990).

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

The most common - morphological species concept - is used for defining species by morphological characteristics. This concept originated from comparative anatomy, where it is possible to visually determine the difference amongst species (Hillis 1987).

Morphological features include colour, shape, size or 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.

The newest concept is the genetic species concept, that arose with the development of molecular biology, where DNA sequences started to be 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 in this frame is DNA barcoding.

2.1.2. Species identification methods

Several methods of identifying species are commonly used today. The most popular methods are morphological taxonomy and molecular systematics, or DNA barcoding. 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 comparative method is the foundation for all species identification. Among the advantages of morphological systematics is defining extinct species, which were identified based on fossil records, as it is difficult and time-consuming to extract DNA from fossilized organisms (Hillis 1987). Since this method has existed for over 250 years, some species are very well known and described, and for these species it is therefore the most reliable and developed method.

There are also cases where species are difficult, or impossible to determine through morphological analysis. For example, microbial species are difficult, or nearly impossible to see with the naked eye (Savolainen et al. 2005). Some organisms in comparison look morphologically indistinguishable, but are in fact distinct species, they are called cryptic species (Duellmand & Venegas 2005). Many species go through different life stages, most common example being caterpillars to butterflies. Sexual dimorphism is one of the issues as well, when female and male looks may vary in colour or size (Savolainen et al. 2005). Discrimination of damaged or incomplete specimen is also complicated using morphological methods.

Molecular systematics comes from molecular genetics, where DNA sequences of organisms are used to distinguish species. Each species is genetically unique, no genome is identical to another (Herbert & Gregory 2005). Intraspecific variation is one of the downfalls of this method. The assumption is, that intraspecific genetic variability should

be lower than interspecific variability. But firstly, this variation must be specified. If not, there is risk that some organisms could be mistaken for new species or subspecies. Some of the benefits include the fact that only a small sample is needed, and that it is a relatively quick and easy method (Savolainen et al. 2005). One of the issues with basing species barriers off of morphological features, is that many species go through different life stages, larvae look very different from an adult individual. Furthermore, many larvae look very similar to one another (Park 2008). The phenotypic differences between various developmental stages of organisms do not pose a problem for genetics since their genome remains the same. The same applies to sexual dimorphism, where the female looks very different from the male (Savolainen et al. 2005). Cryptic species identification has been overcome by molecular techniques, in cases where the organisms look morphologically indistinguishable, but are in fact very different. The most known case of unravelling cryptic species is from Hebert et al. (2004): *Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator.* Ten species of butterflies were considered to be one, the only difference was that some caterpillars ate different plants. It was considered that these caterpillars eat a lot of different plants, but in fact they were different species, morphologically alike.

Traditional taxonomists argue that the genomic species identification is not sufficient to replace morphological phenotypic characterization (Will & Rubinoff 2004). However, molecular scientists suggest morphological taxonomy is outdated and time-consuming (Hebert et al. 2003). Where one method succeeds, the other has a downfall, that is why a combination of the two methods should be used to ensure the highest possible accuracy (Friedheim 2016).

DNA barcoding involves all of the positives and negatives of molecular systematics, since it is a part of this method, and more.

In the case of DNA barcoding, where the whole genome is not needed, only a small segment, the DNA sequence is much shorter. This has its advantage mainly for damaged or incomplete specimens. It is relatively quick and available, useful for non-experts. One of the limitation of this method is that – there is not one universal barcode and there may never be one. A suitable barcode must be found for the specific group we want to identify, at least if it is of animal or plant origin. COI (Cytochrome c oxidase I) is proven to be working well, but only for some groups of organisms (Hebert et al. 2003).

2.2. Plant genome

DNA is the hereditary or genetic material, present in all cells, that carries information for the structure and function of living things. Genetic information is encoded in genes which are composed of exons and introns. An exon is a 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).

Genetic information of plant cells is carried in DNA molecules, chromosomes, most of the DNA is stored in the nucleus. The structure of plant cell is displayed in *Figure 1*. The nuclei of plant cells contain linear molecules of DNA, and the number and length of these molecules differ by species. The main function of nuclear DNA is to transfer structure, build, features and all other information about specific species throughout generations (Řepková 2013). Introns are found in all fully sequenced eukaryotic genomes, they are removed by DNA splicing of the final RNA product. Even though introns are the non-coding part of the genome, they have many other purposes. One of the intronic function is the increase in protein abundance of intron-bearing genes, they usually guarantee high expression of various genes (Chorev & Carmel 2012).

In addition to the chromosomes in the nucleus, the chloroplasts and mitochondria have their own DNA, but these molecules have a circular form. This fact is explained by the endosymbiotic theory: an endosymbiont is a cell which lives inside another cell with mutual benefits. Therefore, it is possible that these organelles may have once been independently living organisms that were incorporated into other cells to form the eukaryotic cell (Wernegreen 2012). Organelle DNA is inherited uniparentally, that means that offspring inherit their genotype from only one parent – there are usually many copies of DNA in these organelles. Mitochondria convert the energy of chemical bonds into the cell energy currency, that is adenosine triphosphate (ATP), and their genome contains DNA for this function as well as genes for 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).

Mutation is a change of genetic information, usually caused by some force such as physical, chemical or biological factors, and it is considered to be one of the mechanisms of evolution. The rate of DNA mutation is inversely related to the size of the genome. A

mutation rate includes all kinds of mutations in a mutation target: base pair substitutions, base addition and deletions. 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 nuclear DNA undergoes relatively slow mutation compared with mitochondrial DNA, sometimes the mtDNA mutation rate is 10 times faster than nuclear. For this reason, a much longer nucleotide sequence would be required for nuclear DNA than is necessary with mitochondrial DNA in order to provide a barcode capable of differentiating species (Waugh 2007). Hence, the ideal barcode for animal species is the mitochondrial gene encoding the cytochrome c oxidase (COI). For land plants, the plastid and nuclear genomes have a 3 to 10-fold greater mutation rate than the mitochondrial genome (Sloane et al. 2012), because mitochondrial genomes have a generally low rate of nucleotide substitutions. Therefore, looking for universal DNA barcode for plants was focused on nuclear and plastid genomes, chloroplast genome is more preferred because it is present in each plant cell in a high number of copies (Pečnikar & Buzan 2014).

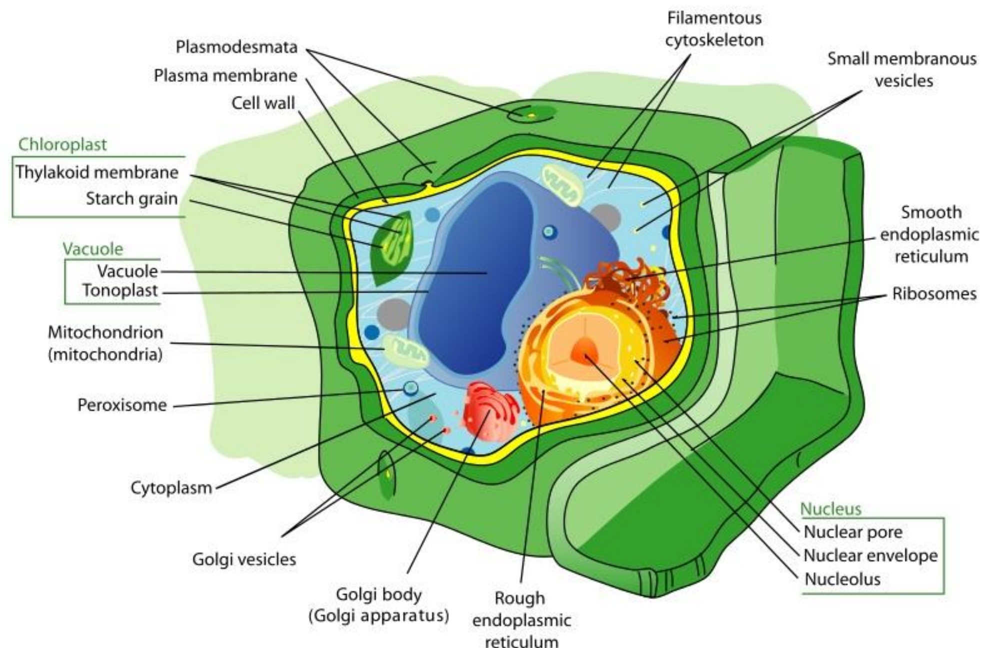


Figure 1. Plant cell structure,
source: https://biologydictionary.net/plant-cell/

2.3. DNA barcoding

The definition of bar coding is: using barcode symbols to identify an item. The DNA barcode goes as an analogy to the universal product code barcode, where by scanning black stripes of the barcode you can identify your purchases, and analogically by scanning the DNA barcode you could identify all living species. (Ratnasingham & Hebert 2007).

DNA barcoding is a method developed to identify species using a short DNA sequence that should be as diverse as possible between species and at the same time be similar within the same species. According to these differences individual species can be determined (Hebert at el. 2003).

DNA barcoding is a molecular and bioinformatics tool that aims to identify biological species. The ideology of DNA barcoding is to find easy-to-use, fast and accurate method that could identify all species on Earth. DNA barcoding in general is quite a new taxonomic method. What makes this method especially novel is that we only use a short segment of DNA instead of using whole genome. The molecular marker should be present in all species and should have enough discriminatory power to distinguish them. These short sequences usually have from 400 to 800 base pairs (Kress & Erickson 2008). Simply put, DNA is isolated from sample specimen, only a small amount of the sample is required. Then it is amplified using primers, to amplify specific region of chosen barcode gene, and after that, sequenced. This sequence produces a DNA barcode that is specific to the sample specimen. The newly sequenced DNA barcode is compared with known barcodes in large online databases such as GenBank or Barcode of Life Database (BOLD).

The barcode should fulfil these criteria: (i) it should be variable enough to be able to distinguish all species, but conserved, to be more variable between species than within. (ii) It should be standardized; different taxonomic groups could be discriminated with the same DNA region. (iii) The DNA region should contain high level of phylogenetic information to easily assign species to its higher taxa. (iv) It should be extremely robust, it must be easy to amplify and sequence. (v) To amplify degraded samples the target DNA region should be short (Taberlet et al. 2007). Unfortunately, this ideal marker does not exist. However, all of the five criteria listed are not equally important for different categories of users. For example, for taxonomists' needs, a high level of variation and sufficient phylogenetic information will be the most important. In contrast, for ecologists,

in forensics or when analysing processed food robustness and level of standardization will be important.

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 specimen, or to distinguish a matrix containing mixture of biological species (Valentini et al. 2008). DNA barcoding is suitable for two purposes: the molecular identification of already described species (Hebert et al. 2003), and the discovery of undescribed species (Hebert et al. 2004).

The key to identifying species is to determine the minimum amount of sequence differences, according to which it is possible to reliably distinguish all types of the selected group, the threshold value for differences between intraspecific variability and interspecific divergence. The assumption is that intraspecific genetic variability should be smaller than interspecies genetic divergence (Hebert et al. 2003).

If interspecies divergence exceeds intraspecific genetic variability, a gap called the barcode gap arises (Meyer & Paulay 2005). According to the size of the gap, or the extent of the differences between the intraspecific and the interspecies variability, it is possible to determine whether the samples represent one or more species. Some researchers state that the barcoding gap does not really exist at all and is only an artefact due to insufficient number of samples (Wiemers & Fiedler, 2007). This would be fatal for DNA barcoding and for 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.

2.4. History of DNA barcoding

The term "barcode" was first used in a paper published by Arnot et al. (1993). The authors mention DNA-based genotyping using PCR amplification, which increases the sensitivity of the technique. However, using molecular tools to determinate organism was mentioned even sooner (McAndrew & Majumdar 1983). The idea of using a small segment of the genome to discriminate organisms was first broadly accepted by those working with viruses and bacteria, which are the least morphologically traceable groups (Pace 1997). Determination of species differences using molecular tools has been in use since the mid-1960s (Hubby & Lewontin 1966). Hebert et al. (2003) were the first to propose biological identification through DNA barcodes; they were the first ones who put microgenomic identification onto a large scale. 648 base-pair regions in mitochondrial

cytochrome *c* oxidase I gene (COI) were selected as the standard barcode for almost all animal groups. Since then COI has proved successful as a taxonomic tool and is highly effective in identifying many animal groups such as birds (Hebert et al. 2004), spiders (Barrett & Hebert 2005), butterflies, fish and more. Unfortunately, COI cannot work as a universal barcode; it works only for animal species. Universal barcode for plants was introduced by The Consortium for the Barcode of Life (CBOL) Plant Working Group in 2009, two protein coding regions from plastid genome *rbc*L (ribulose-bisphosphate carboxylase) and *mat*K (Maturase K) were chosen (Hollingsworth et al. 2011).

2.5. Plant DNA Barcode

In plants COI is only useful in some algae (Saunders 2005), because variation in mitochondrial DNA is limited in general and it evolves too slowly. Since COI is not an effective barcode region in plants, and searching for suitable plant equivalent has proven difficult. Finding a universal plant barcode had to be done outside the mitochondrial genome. Among many suggestions, there was the *trn*H-*pbs*A intergenic spacer (Kress et al. 2005), some phylogenetics markers such as *rbc*L and *trn*L-F (Chase et al. 2005). There were also proposed barcodes which involved seven plastid markers in various combinations, e.g. *rbc*L+*trn*H-*pbs*A (Kress & Erickson 2007). *Table 1* presents all DNA markers proposed by various research groups, that have been included in their plant barcoding studies in different combinations.

The only nuclear marker proposed was the nuclear ribosomal internal transcribed spacer (nrITS). Unfortunately, it was found that this marker is very similar in plants and fungi. Fungal DNA is often amplified from plant samples, which leads to misleading sample identifications. Another limitation is that ITS is sometimes difficult to amplify and sequence. Therefore, ITS is not recommended to be used as single barcode locus but can be considered as a supplementary barcode (Hollingsworth 2011). It could be used in some parasitic plants with highly reduced plastid genomes. Hence, the focus shifted to the plastid genome.

An agreement on common plant barcode is of the highest essence, so the process of barcoding plants could progress towards the creation of a shared community resource. To formalise the selection of universal plant barcode, large consortium formed of the different research groups the Consortium for the Barcode of Life (CBOL) Plant Working Group. In 2009, CBOL approved two coding regions from plastid genome *rbc*L and *mat*K as a "core barcode" for plants (Hollingsworth et al. 2011). This decision was based on

straightforward recovery of the *rbc*L region, the fact that it is easy to amplify, sequence and align in most land plants and the discriminatory power and rapid evolving of the *mat*K region. *mat*K is perhaps the closest plant analogue to the COI animal barcode. Unfortunately, *mat*K can be difficult to amplify and *rbc*L discriminatory power is only modest. These two markers complement each other well. It was not a unanimous decision, only a majority preference, since each of the candidate markers had different strengths and weaknesses. It was also advised to use additional markers as required (CBOL Plant Working Group 2009).

Among the most widely used plastid barcoding marker is the intergenic spacer *trn*H-*psb*A, which is straightforward to amplify across land plants. It is the obvious choice of a supplementary barcode and shows high species discrimination power mainly in species such as *Ficus* and *Alnus*. One of its main limitations is the premature termination of sequencing reads by mononucleotide repeats, which leads to unidirectional reads in up to 30 % of sequences (Hollingsworth et al. 2011).

The *trn*L intron and the intergenetic spacer between *trn*L and *trn*F have been widely used since 1990s, firstly proposed by Taberlet et al. (1991). These robust sets of primers allow routine recovery, are well conserved and generally simple to sequence, although mononucleotide repeats (*Table 1)* can impact on sequencing reads. The major strength of the *trn*L intron for species identification is the presence of a small stem-loop structure within the intron, the P6 loop. This very short "minibarcode" has conserved priming sites flanking a variable loop of ca 10-143 base pair. P6 has proved very useful to ecologists studying highly degraded DNAs found in processed food or in fossil remains. The main drawback is the low resolution of the *trn*L intron compared with several other non-coding chloroplast regions (Taberlet et al. 2007).

Table 1: Characteristics of different markers hat 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 $\geq$ 10 bases

2.6. DNA barcode for wood

Barcoding trees could be done using plant DNA barcodes, as mentioned above, but when we do not possess leaf material, it can be challenging. However, the study by Deguilloux et al. (2002) demonstrated the potential of genetic analysis on dry wood, as plastid, mitochondrial and nuclear DNA sequences could all be recovered through PCR amplification.

Wood is composed of cells that die during the cell development. Mature wood cells are dead, they are devoid of protoplasm and nucleus, even in the living tree. Protoplasm and nucleus are present in the early stages of the short life of a wood cell (Plomion et al. 2001).

Wood is a botanically poor source of DNA. DNA extraction from wood is not necessarily as simple or direct as from other plant parts that can be collected and analysed in the living state *(Figure. 2)*. It has been shown that woody plant lineages show consistently lower rates of molecular evolution as compared with herbaceous plant lineages, suggesting the application of DNA barcoding concept should be more difficult for tree floras than for non-woody floras (Smith & Donoghue 2008).

An ideal DNA barcode should be short, making it easy for recovery, and have sufficient information to provide maximum species discrimination. Shorter amplicons showed a generally higher recovery rate than longer ones. Hence, short portions of the barcode region – mini-barcodes – may be used in place of full-length barcodes (Meusnier et al. 2008). DNA mini-barcodes are short DNA sequences of 100-250 base pair, they are suitable for species identification when high-quality DNA is not available and seriously degraded DNA is retrieved (Jiao et al. 2018).

According to Gonzales et al. (2009), ITS does not seem to suit as a universal DNA barcode for tropical forest inventories, given the limited sequencing success in said study. However, one of the highest ranked markers was *trn*L intron, which was highly variable and easy to sequence. Furthermore, by their conclusions it is an interesting option for future barcoding projects, especially considering its short sequence, that is suitable for degraded or fragmented DNA. The best performance showed the non-coding plastid DNA spacer *psb*A-*trn*H. On the contrary, Lee et al. (2016) showed in their study a combination barcode of *trnL-trnF*+ITS2 works best for species discrimination. Yu et al. (2017) propose the combination of ITS2 and *trn*H-*psb*A marker.

Since *trn*L approach has proven very useful for ecologists, which are studying degraded DNA and usually only small sections of highly fragmented DNA is recovered, it could work for this purpose as well.

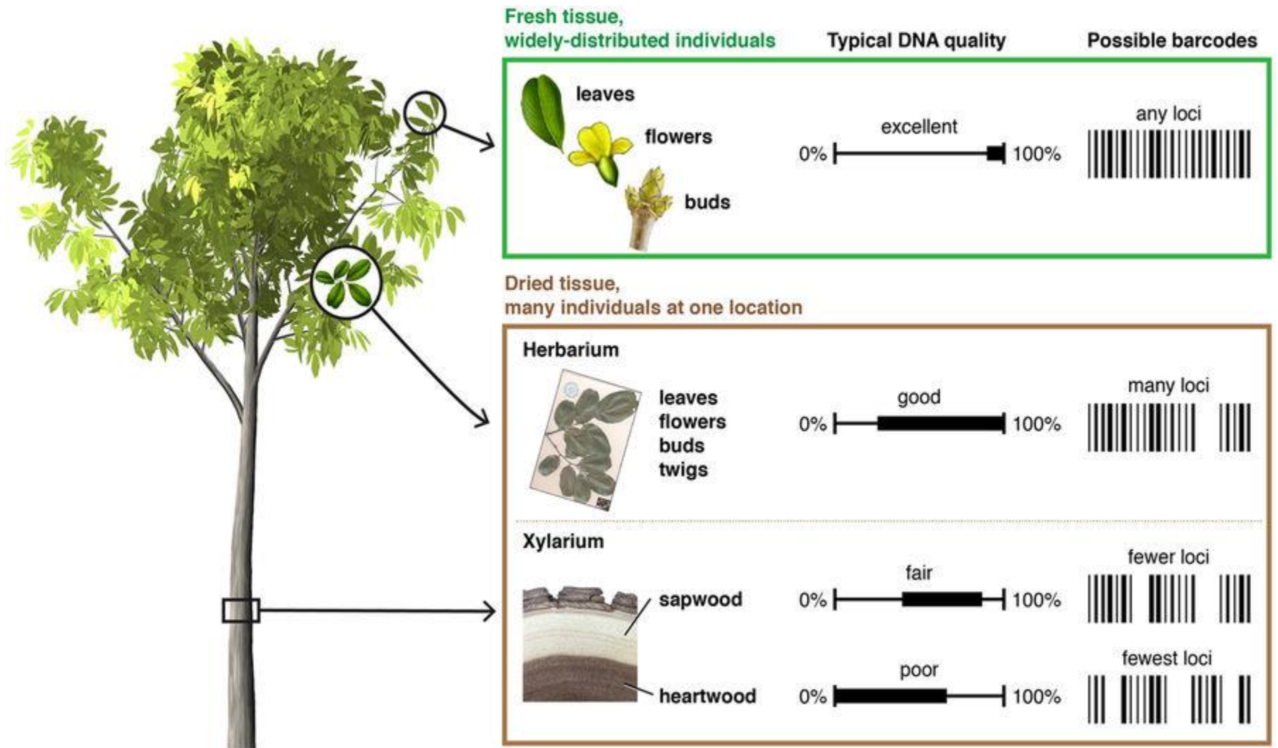


Figure 2. A schematic representation of the potential strengths and weaknesses of source tissue (fresh, herbarium, xylarium),

source: https://www.nature.com/articles/s41598-018-20381-6

2.7. Practical uses of DNA barcoding

The main field that can benefit most from DNA barcoding is taxonomy. Taxonomy can be used in other disciplines than just taxonomy. The development of new, faster and simpler molecular genetic methods has made DNA sequences more accessible and therefore useful to other branches of biology. The most promising use of DNA barcoding technique is in fields of conservation biology, ecological studies, medicine, pharmaceuticals and systems biology (Pečnikar & Buzan 2014).

Biosecurity and public health may use barcodes for identifications of parasites. Infections with parasites borne by widespread vectors result in high number of illnesses and mortality (Besansky et al. 2003). Identifying parasites means identifying the source of the disease and understanding the interactions between the host and the parasite. Correct taxonomic identification is therefore crucial as it enables differentiation between

morphologically similar species that cause different diseases. Beside cryptic species another challenge in right determination is that parasites go through different stages of development, their life circle may include multiple hosts, and the fact that they live deep in the host tissue (Pečnikar & Buzan 2014). The determination of pests poses very similar problems as the identification of parasites. This helped with tracking a new pests species in California, the light brown apple moth, *Epiphyas postvittana*, an invasive species originating from Australia (Floyd et al. 2010). This branch also implies to biomonitoring, which includes the pathogen spread and their associated vectors. With climate change, spread of tourism and trade, the movement of exotic species around the world is increasing. It is estimated that 1 % of the species introduced to novel environments will become invasive and have serious economic impact (Williamson 1996).

Food industry is one of the fields of biology that benefits the most from DNA barcoding, especially in food safety and quality. DNA barcoding is effective in certifying both origin and quality of raw materials, and to detect adulterations occurring in the industry food chain (Galimberti et al. 2013). This was proven to be particularly effective in the traceability of seafood. The increase in demand for seafood and the globalization of the market have made the control of the trade routes, the industrial processing systems and identification by the area of origin more difficult. DNA barcoding can be a tool to prevent mislabelling and fish species substitution (Barbuto et al. 2010). Between many uses of DNA barcoding in this branch of biology analysis of food composition is also included. Such as meat that is contained in pork pates, sausages content or origin of smoked salmon (Teletchea et al. 2008).

Assessing biodiversity by barcodes has its advantages in ecosystem that are species-rich, difficult to access and poorly catalogued. There is a risk that many species will become extinct before they are taxonomically recorded (Mora et al. 2011). Most of the unknown Earth's biodiversity is concentrated in developing countries. DNA barcodes could lower the cost and time requirements for such assessments (Gaston & O'Neill 2004). Samples can be done in the traditional way by sampling separate organism or by analysing samples from soil, water and air. The usefulness of DNA barcoding is not restricted to the recent biodiversity, it can also be used as a reconstruction method of ecological conditions on Earth in the past by analysing sediments or ice and through remnant biological remains (Kuch et al. 2002).

One of the other problems that DNA barcoding could prevent and provide evidence for, is international (illegal) trade with both living and dead biological material, especially rare species. It has proven useful in identifying species listed in the Convention on International Trade of Endangered Species (CITES) appendices (Muellner et al. 2011). Forest destruction and degradation continue to be major threats to global biodiversity and cause severe environmental damage. The ability to track timber resources from forest to marketplace is important for successful management and proper regulation of the timber trade as it could prevent illegal logging (Gonzalez et al. 2009).

The use of DNA barcoding also has perspective in forensic genetics. One of the interesting examples is the use of DNA sequences of dogs in the determination of murderers and other criminals, where in stolen cars or in the suspect's clothes hair was found which could belong to the dogs of the victims (Savolainen & Lundeberg 1999). In forensic entomology, the COI sequence was used to identify the *Calliphoridae* family, whose females place eggs on dead bodies, and according to the stage of larval development, an approximate time of death can be determined (Chen et al. 2004). Forensic botany can also support evidence during criminal investigations with its most common application limited to identifying specific as well as suspect illegal plants, it can present additional information in many forensic cases involving plant evidence that may be useful to link a suspect, a victim, or a vehicle to the crime scene (Ferri et al. 2009).

Uses for DNA barcoding of wood can be the ability to track timber resources, prevent illegal logging, or cutting down rare species. Or as it is in our case, identifying species which were eaten by termites. If is there preferred or favoured species of trees, eventually which species do they prefer and which not.

However, all of the advantages of DNA barcoding cannot reach its full potential until a barcoding method will be proven 100 % effective and reliable.

2.8. Databases and other components of DNA barcoding

One of the main aims that came with proposing DNA barcoding technique was to create a database that would link barcodes to organisms.

2.8.1. Barcode of Life

In May 2004, the CBOL consortium (The Consortium for the Barcode of Life) was established, to help develop DNA barcoding. Since then, many other organisations were added among the members, such as natural science museums, biological collections,

government agencies working with academic and commercial experts in the area of biotechnology, informatics, genomics, taxonomy and computer science. CBOL and its members are committed to creating public and freely available non-commercial libraries of data obtained by DNA barcoding. The goal of CBOL is to develop new methods, as well as primers, to obtain additional barcode records and apply these methods in public spheres.

Database for all barcode sequences has become BOLD (Barcode of Life Data System). It is a workspace where all working groups can insert results from their barcoding projects, barcode sequences from their analysis, along with information about the species, and at the same time to draw on it for further studies (Ratnasingham & Hebert 2007). For that purpose, a web database was created (http://www.boldsystems.org/). There are available public records of barcode sequences, lists of primer used for DNA barcode amplifications. Here we can also find all the published papers related to DNA barcoding.

Barcode of Life also has some subbranches as a FISH-BOL (The Fish Barcode of Life Initiative). Their effort is to assemble a standardised reference library for all fish species, create a valuable public resource in the form of electronic database containing DNA barcodes, images, and geospatial coordinates of examined specimens. Another project which focuses on specific field is Tree-BOL. Tree-BOL is a Barcode of Life (BOL) initiative to sample all the species of trees of the world, this project intends to DNA barcode all 100,000 species of trees of the world. Tree-BOL should significantly help to advance plant DNA barcoding in general. Trees were selected for their economic value as sources of fuel, fibre, food, flowers, as well as for their ecological value as producers of nearly half of the oxygen necessary for life on Earth (Aplin 2007).

2.8.2. GenBank

GenBank is a genetic sequence database, 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 Information (NCBI) (www.ncbi.nlm.nih.gov) sealed a partnership with CBOL which achieved that barcode standard DNA sequences and relevant supporting data can be now archived in GenBank.

NCBI provides a Barcode Submission Tool (BarSTool), which is a web-based tool that can be used to submit sets of barcode sequences to GenBank. In addition, it also collects other information to help identify the organism sequenced and to support the accuracy of the sequence. Each Barcode record is ensured to be complete with information including: specimen voucher, geographic location, collection date, primer data and trace files.

BOLD and GenBank contain the same public records, but they offer different options for optimizing their use as reference libraries. For barcode data, Sonet et al. (2013) recommended using the BOLD Identification System and searching the dataset including early-released sequences. This option gives access to a workbench with supplementary tools and it allows users to verify the quality of the data and optimises the number of best-matches.

3. Objectives

The main objective of the thesis was to summarize information known about DNA barcoding, the current knowledge about this novel taxonomic technique and its usage. Specifically, the focus was on DNA plant barcode, with emphasis on tropical trees. This thesis also describe downfalls and benefits of this method, maps the criteria by which barcode loci is selected and explores the current and possible use for DNA barcoding among scientific community or the general public. The outcome of this thesis could be considered as a summary report or a guide by which barcode sequences for wood barcoding could be used.

The second aim was also to optimize DNA isolation from dead wood samples – to select the best barcode loci suitable for wood identification and optimize PCR amplification of selected barcode loci. First verify this procedure on green leaves of trees and then apply it to wood samples.

The main idea behind creating this thesis was to be able to identify wood digested by termites. In future, we could use DNA barcoding of wood for this purposes of identifying termites diet.

4. Methodology

The methodology consists of a theoretical part (literature review) and practical part carried out in the Laboratory of Molecular Genetics at the Faculty of Tropical AgriSciences and with the help of Termite Research Team of Faculty of Forestry and Wood Sciences in their molecular genetics laboratory.

Most of the information was gained from scientific papers found through databases such as BOLD (Barcode of Life Data System) (http://www.boldsystems.org), Web of Knowledge (http://apps.webofknowledge.com) or Google Scholar (https://scholar.google.com). All of the information sources and papers used are listed in 'References'.

4.1. Plant materials for laboratory analysis

Leaves were sampled from trees it the campus of Czech University of Life Sciences. The samples were a variety of tree species such as Ash (*Flaxinus*), Chestnut (*Castanea*), Loquat (*Eriobotrya japonica*), Hazel (*Corylus*), Oak (Quercus) and Willow (*Salix*). These samples were collected during spring 2017.

Wood for identification was sampled in French Guiana, in pristine lowland tropical rainforest (4°05' N, 52°40' W) *(Figure. 3)* during November 2014. Four samples of different, unknown wood species were recovered. The four wood samples were marked as 'FGT: 1-10, 16-17, 24-7, 25-7'.

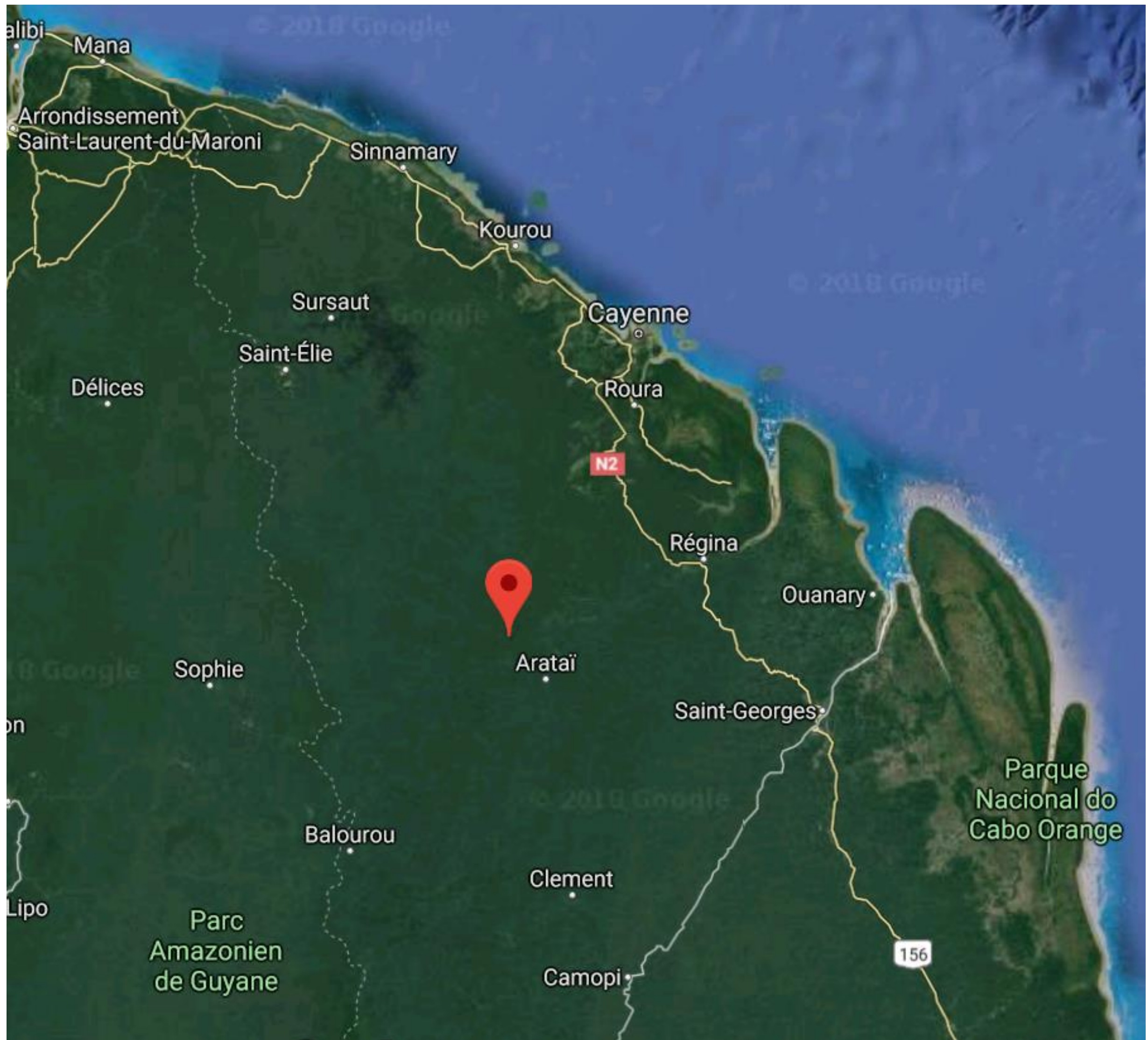


Figure. 3 Location of wood samples, adapted from maps.google.com.

4.2. Laboratory analysis

Methods used in this study were: DNA extraction using two different extraction methods – CTAB and DNeasy Mini Plant Kit – Qiagen, and PCR amplification.

4.2.1. DNA extraction

DNA extraction from leaf material

The leaf material was teared by sterilized pipets into 2 mL microcentrifuge tubes and 2 small stainless steeled beads were added. Tissue was homogenized using Mixer Mill (Retch) apparatus.

Isolation was done using DNeasy Mini Plant Kit (Qiagen) by strictly following the manufacturer's instructions.

DNA extraction from wood material

Wood preparation – powdering

Wood samples were cut using sterilized scalpels, sterilization was done by 96% ethanol and fire, into shavings. 2 mL microcentrifuge tubes were filled with the wood shavings and put into liquid nitrogen for 10 minutes to freeze the wood samples. Frozen wood shavings were transferred to chemically sterilized, by sodium hypochlorite, grinding jars with stainless steel bead, put into Mixer Mill (Retch) apparatus and grinded for 5 minutes. For 2 samples grinding step was repeated, until they became a fine wood powder.

Extraction with DNeasy Mini Plant Kit

Genomic DNA was extracted from wood powder by use of a DNeasy Mini Plant Kit (Qiagen) following the manufacturer's instruction with some minor modifications. The incubation with Buffer AP1 at 65 °C was prolonged from 10 minutes to 30 minutes and the incubation with Buffer P3 on ice was for 15 minutes instead of 5.

CTAB extraction

DNA was extracted from wood powder using a modified CTAB protocol (Doyle & Doyle 1987; Faleiro 2002). 800 ul of extraction buffer (CTAB 2.8%, NaCl 1.3 M, EDTA 20 mM, TRIS-HCl 100 mM, PVP 1%, mercaptoethanol 0.2%) and 100 ng of Proteinase K were added to the ground tissue for lysis of wood cells, samples were heated at 65°C for one hour while mixing them every 10 minutes. Then 700 ul of chloroform:isoamyalcohol (24:1) was added to denature the contaminations. The contents were mixed for 10 minutes and the phases were separated by centrifugation for 10 minutes

at 14000 RPM and 4 °C. The supernatant was transferred to a new microtube. After that 55 ul of CTAB 7% was added and the chloroform:IAA extraction repeated once more, so all contaminants were removed. This supernatant was mixed with 900 ul isopropanol in new tubes, to allow the DNA to precipitate, and the tubes were placed at 20 °C for one hour. After that, the tubes were centrifuged for 10 minutes at 14000 RPM and 4 °C. The supernatant was discarded and to remove the remaining salts, the resulting pellet was washed twice with ethanol (96% and 70%). The pellet was dried at room temperature and dissolved in 100 ul of TE buffer with addition of 30 ng of RNase.

DNA quantification

Genomic DNA was quantified, more specifically the concentration and quality of extracted DNA, by measurement of sample with a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA).

4.3.1. PCR amplification

For this study the *trn*L approach was used, which means using the chloroplast *trn*L (UAA) intron for PCR amplification. *Figure 4.* presents the location of the primers in the chloroplast *trn*L (UAA) gene, *Table 2.* Gives their sequences. The primers *c* and *d* cover the entire *trn*L (UAA) intron. The primers *g* and *h* were designed on two highly conserved regions inside the intron.

Each sample was amplified with two primer pairs, *c* and *d* for the 'long sequence', which reaches 456 base pair (bp) long in tobacco as found by Taberlet et al. (2007) and *g* and *h* for the 'short sequence', which in tobacco is 40 bp long.

DNA amplifications were carried out in a final volume of 25 μL.

PCR amplification of leaf isolates

For amplification of DNA were used the isolates by DNeasy Mini Kit (Qiagen). The components of amplification mixture and its exact volume are listed in the *Table 4.* For both sequences the same cycle was applied, started with initial denaturation at 94 °C, followed by 40 cycles of denaturation for 30 seconds at 94 °C, annealing for 30 seconds of 56 °C, and extension for 30 seconds of 72 °C, continued the final extension step for 7 minutes at 72 °C.

The amplified products were displayed on 1% agarose gel, ethidium bromide stained and visualized under UV light. Electrophoresis ran for 90 minutes at 90 V. In this PCR amplification we also used a positive and a negative control. A positive control is a sample

that is expected to work and provides the expected result. A negative control is a sample that is not supposed to work, if it does, it can mean some contaminations in the samples. As a positive control was used isolate – 'Javor' and negative control is marked as 'K-'. On the sides, DNA ladder was placed and the length between each band was 50 bp.

PCR amplification of wood isolates

The amplified DNA was extracted by CTAB extraction method. The components of amplification mixture and its exact volume are listed in the *Table 3*. For primers *c* and *d* the amplification started with initial denaturation for 15 minutes at 95 °C, followed by 35 cycles of denaturation for 30 seconds at 95 °C, annealing for 30 seconds at 50 °C, and extension for 2 minutes at 72 °C, then continued in a final extension step for 10 minutes at 72°C. For primers *g* and *h* in the process of PCR amplification the elongation was removed as suggested by Valentini et al. (2010), the mixture was denatured at 95 °C for 15 minutes, followed by 35 cycles of 30 seconds at 95 °C, and 30 seconds at 55 °C.

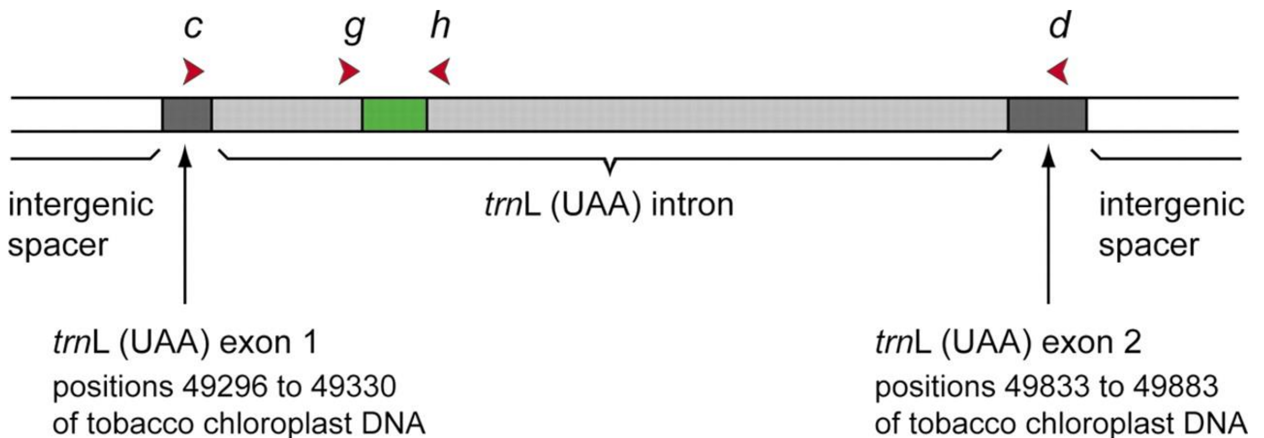
The amplified products were displayed on 1% agarose gel, ethidium bromide staining and visualized under UV light. Electrophoresis ran for 45 minutes at 120 V. DNA ladder used on this gel is graduated by 100 bp.

Table 2. Sequences of the two universal primer pairs amplifying the *trn*L (UAA) 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

Figure 4. Position of the primers *c, d, g* and *h* on the chloroplast *trn*L (UAA) gene. The P6 loop amplified with primer *g* and *h* is indicated in green.



Source: Taberlet et al. 2007

Table 3. PCR mixture – wood

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

Table 4. PCR mixture – leaves

PCR Nucleotide Mix	1 µL
AmpliTaq Polymerase	0.35 µL
Buffer	2.5 µL
BSA	1.5 µL
MgCl2	0.2 µL
primer *c/g* forward	1 µL
primer *d/h* reverse	1 µL
H2O	12.45 µL
DNA	5 µL
	25 µL

4. Results

4.3. DNA extraction results

The success rate of the DNA extraction from fresh wood material is displayed in *Table 5*. The secondary measure of nucleic acid purity, the 260/230 ratio was extremely low. The 260/280 was also lower than the model ratio, but not that significantly.

As we can see in tables below (*Table 6.*), the isolation using CTAB extraction method yielded in significantly higher amounts of DNA than DNeasy Mini Kit extraction. Averaged nucleic acid concentration in CTAB is 3.5 times higher than in Qiagen kit. Unfortunately, both methods lack in DNA quality. The ratios of absorbance are used to assess the purity of DNA. The ratio at 260nm and 280nm of ~1.8 is generally accepted as "pure" DNA. However, the average ratio in DNeasy Mini Kit extraction was 2.31, which is more than ideal by 0.51, nonetheless high 260/280 purity ratios are not necessarily indicative of a problem. The average ratio in CTAB extraction was 1.32, which is the opposite, it is lower than the ideal ratio by 0.48. The ratio at 260nm and 230nm is a secondary measure of nucleic acid purity, expected values are commonly in the range of 2.0-2.2. The average ratio in Qiagen kit extraction was 1.55 and in CTAB extraction was 1.7, which is not much of a difference. In both extraction methods, 260/230 ratio was lower than the model. Even though this slight deficiency in quality, DNA was successfully isolated from dead wood material, mainly by CTAB extraction method.

Table 5. fresh leaf material DNA quantification using NanoDrop 2000 spectrophotometer of DNeasy Mini Kit extraction method.

Sample ID	**Nucleic Acid Concetration**	**260/280**	**260/230**
Loquat	36 ng/µl	1.51	0.17
Chestnut	79 ng/µl	1.63	0.32
Ash	25 ng/µl	1.62	-1.58
Willow	176 ng/µl	1.49	0.91
Hazel	29 ng/µl	1.43	-1.34
Oak	43 ng/µl	1.54	-4.75

Table 6. wood DNA quantification, comparation of CTAB and DNeasy Mini Kit (Qiagen) extraction method.

Sample ID	**NA Qiagen**	**NA CTAB**	**260/280 Qiagen**	**260/280 CTAB**	**260/230 Qiagen**	**260/230 CTAB**
1-10	26 ng/µl	216.5 ng/µl	2.75	1.03	3.45	5.09
16-17	25 ng/µl	47.7 ng/µl	2.69	1.44	1.13	0.6
24-7	38 ng/µl	137.4 ng/µl	2.52	1.39	1.15	0.59
25-7	69 ng/µl	148.6 ng/µl	1.27	1.42	0.47	0.52

*NA = Nucleic Acid Concentration

260/280 = the 260nm and 280nm ratio, measure of nucleic acid purity

260/230 = the 260nm and 230nm ratio, secondary measure of nucleic acid purity

4.4. PCR amplification results

In the visualized leaf PCR products (*Figure 5.*). The *c-d* sequence appeared clearly every time, the sequence length shows above 500 base pair, it varies between 500 base pair and 600 base pair. The *g-h* sequence is less visible, and its length varies around 100 base pair.

Looking at the gel of wood PCR product (*Figure 6.*). The *c-d* sequence, DNA appeared degraded, and fragment sizes ranged from approximately 80 bp to 1,000 bp. The most visible segments were at approx. 600 base pair length, except for the sample 1-10, which was strongest at 300 bp. The *g-h* sequence base pair length varied around or under 100bp.

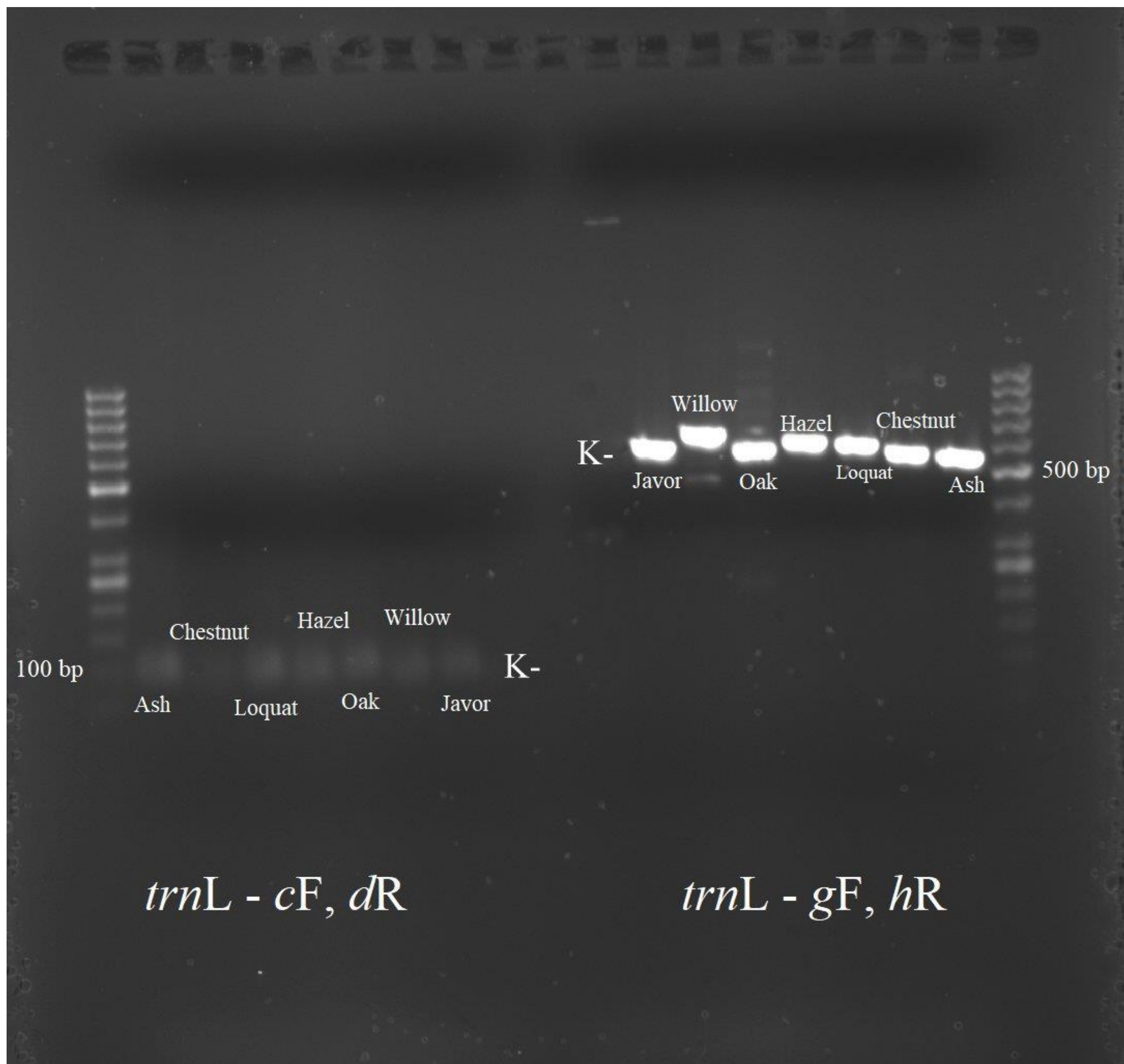


Figure 5. Electrophoresis of leaf PCR products.

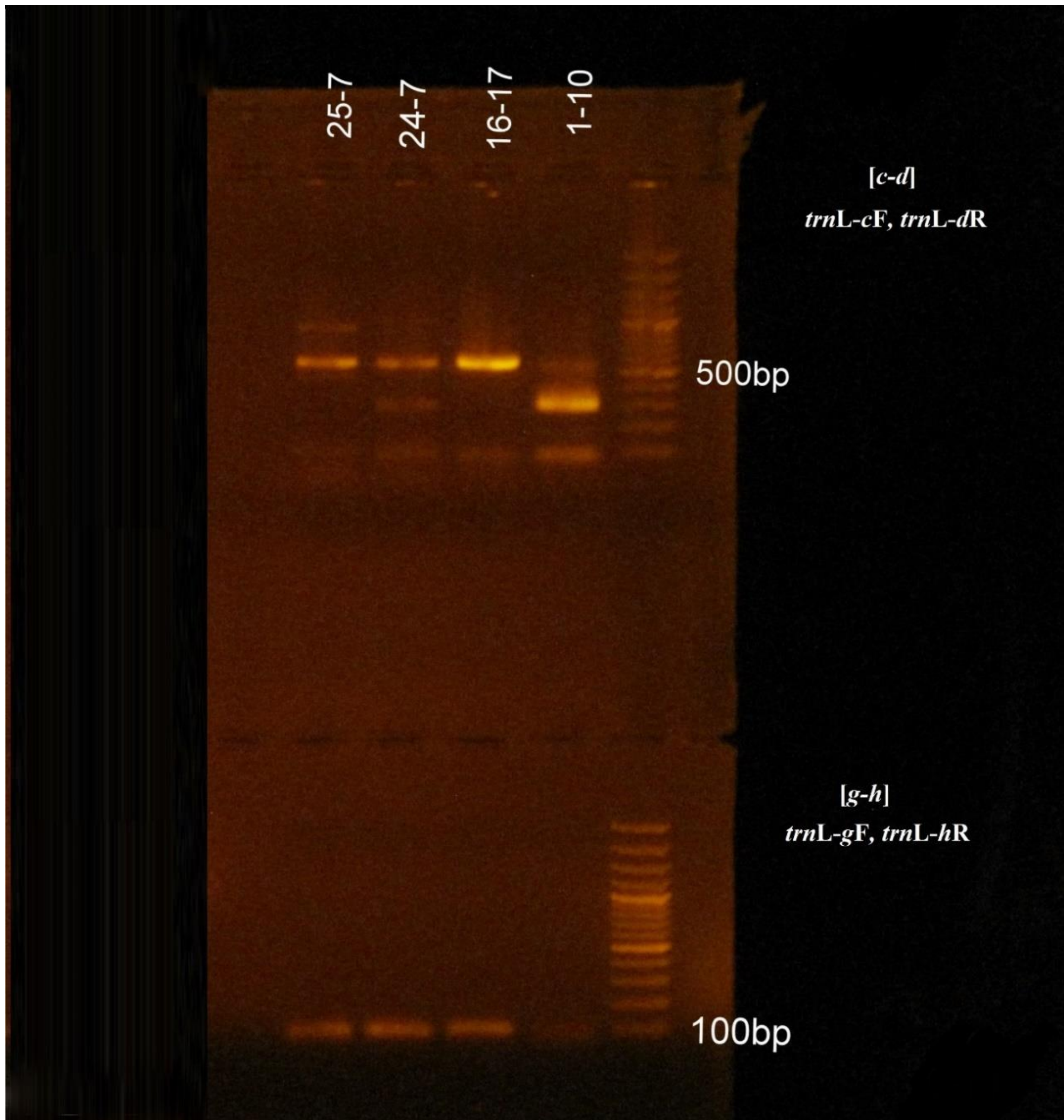


Figure 6. Electrophoresis of wood PCR products.

5. Discussion

DNA barcoding was proposed as a new taxonomical tool to facilitate species identification, it promises many advantages and practical uses that can be very useful to the species identification not only for taxonomic and scientific use, but it can serve for non-experts as well. DNA barcoding to hit its full potential it is necessary for this method to be 100 % affective and reliable. It is well established for animal species identification and plant species identification looks promising as well. Since DNA barcoding is mainly used to identify living species, it is not well developed for dead organism such as wood. Many papers published regarding this topic have been dealing with the same problem, that wood is a poor DNA source and basically even a living tree is formed by dead wood cells. Therefore, it is essential to choose the right barcode that fits these criteria

Universal barcodes for plant species identification are set and proven to be working. Moreover, it is suggested to use supplementary barcodes if needed. For wood identification were tried many barcodes designed for plant identification. Variety of combinations were successful by different studies and for different species. However, there is no universal barcode that appeared to be working for wood identification. Since in identifying of wood species we are working with small amount of low-quality, more importantly fragmented DNA. The rule in this kind of identification is that, the chance of successful amplification increases as the size of amplicon decreases. Therefore, looking for the shortest barcode, more so a 'minibarcode' is needed. By these standards, wood barcoding appears to be the closest to forensic and ecological studies. Thus, a barcode region widely used in ecological studies may be the best option. The shortest barcode suggested for plant species identification is the *trn*L loop, which length varies from 10 to 143 base pair. Its major strength is in the robust set of primers that allow routine recovery, which has been widely used by ecologists studying degraded DNAs.

Firstly, the method was tried on a green leaf material. DNA extraction was, except for some impurities in isolates, quite successful. The low 260/230 ratio, signalises possible presence of organic contaminants, which absorb at 230nm, such as TRIzol, chaotropic salts and other aromatic compounds. PCR amplification worked for both regions well, it shows that DNA from live material is not that fragmented as is in the case of dead wood material.

The lower 260/280 purity ration in CTAB extraction was lower than the ideal value, which usually indicates that a sample is contaminated by residual guanidine, or other

reagent used in the extraction protocol. The 230/260 secondary purity ratio was low in both cases and lower ratio may indicate the presence of organic contaminants.

As expected CTAB extraction was more successful. Unfortunately, its success was more visible in the quantity of extracted DNA. DNeasy Mini Kit (Qiagen) produced low-quality and low quantities of DNA. Whereas CTAB yielded higher quantity of DNA, its quality was similarly low to the Qiagen kit. The low-quality of extracted DNA is an unpleasant fact that may need more exploring such as the examination of the causes of impurities caused by residual reagents used in extraction protocol or by other contaminants, in order to extract purer DNA from dead wood. However, despite the fact of the low purity ratios PCR amplification worked perfectly. Therefore, dead wood was successfully isolated by CTAB extraction method. Between the three extraction methods (Qiagen kit, CTAB, PTB), Asif & Cannon (2005) evaluated the *N*-phenacylthiazolium bromide (PTB, Prime Organics) as the best extraction method. This method has primarily been used for DNA extraction from ancient bone in palaeontological studies. Thus, it may be useful in dead, degraded wood samples as well.

We confirmed that the *trn*L approach is a good choice for this study. The sequence lengths corresponds with the findings of Taberlet et al. (2007). However, the 'long sequence' shows fragmented DNA, it may mean the sequence is too long for wood DNA samples. Thus, not that quality of DNA is extracted from dead wood to use the *c-d* gene region. The 'short sequence' works well in degraded DNA, but the shorter the sequence the harder is to recognize the interspecies variability. Therefore, the ability to amplify both long and short regions of DNA will allow better species recognition. Unfortunately, this may not be possible. Another solution for this problem may be to optimize 'medium sequence', for example trying the combination of primer *c* and *h*. This region, theoretically, could be short enough to avoid fragmentation and at the same time long enough to be interspecifically variable.

Despite the fact, that we amplified low-quality DNA, the amplification was successful in two regions of different lengths. Thus, we have proven that barcoding of wood is possible.

6. Conclusions

DNA barcoding is a novel method used for species identification. This thesis demonstrated that the barcoding technique, although very well developed for animal identification, lacks its reliability for plant species. It works with a 96% success rate for animal species and around 70% for plant species. It is usually used for living organisms, and barcoding of degraded and dead material is not that common. Since identification of plant species using barcodes relies on the use of green parts of the plants, barcoding of wood is even more problematic and not that well explored. No universal DNA isolation method nor a barcode for wood exists. The successful extraction of useful DNA from various wood samples is a necessary first step for PCR amplification and further DNA sequencing. Extraction methods designed for isolation of plant species may not work perfectly, looking outside of this direction may be a possible solution. Universal barcodes for plants do not work well for wood identification. Therefore, different combinations of markers are being tested. The only single marker, which is proving to be among the effective ones is marker that is popular for its ecological purposes – *trn*L intron. Hence, the *trn*L approach was applied to this study, among success in barcoding of wood it revealed its downfalls, meaning the length difference between the two sequences. Thus, the need for further optimization arises. However, despite the fact that isolation and PCR amplification of dead material has many downfalls and challenges that needs to be overcome, we have successfully isolated DNA from dead wood material using CTAB extraction method and PCR amplified it using *trn*L approach.

7. References

Aplin D. 2007. Tree-BOL to Barcode World's 100,00 Trees. Botanic Gardens Conservation International. Available from https://www.bgci.org/news-and-events/news/0463/ (accessed April 2018).

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.

Asif MJ, Cannon CH. 2005. DNA Extraction From Processed Wood: A Case Study for the Identification of an Endangered Timber Species (*Gonystyus bancanus*). Plant Molecular Biology Reporter **23**: 180-187.

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

Barbuto M, Galimberti A, Ferri E, Labra M, MalandraR, Galli P, Casiraghi M. 2010. DNA barcoding reveals fraudulent substitutions in shark seafood products: The Italian case of "palombo" *(Mustelus spp.)*. Food Research International **43**: 376-381.

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

Besansky NJ, Severson DW, Ferdig MT. 2003. DNA barcoding of parasies and invertebrate disease vectors: what you don't know can hurt you. Trends Parasitol **19**: 545-546.

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.

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

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.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin **19**: 11-15.

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.

Ferri G, Alù M, Corradini B, Beduschi G. 2009. Forensic botany: species identification of botanical trace evidence using a multigene barcoding approach. International journal of legal medicine **123**: 395-401.

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.

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

Galimberti A, De Mattia F, Losa A, Bruni I, Federici S, Casiraghi M, Martellos S, Labra M. 2013. DNA barcoding as a new tool for food traceability. Food Research International **50**: 55-63.

Gaston KJ, O'Neill MA. 2004. Automated species identification: why not? Philosophical Transactions of the Royal Society B: Biological Sciences **359**:655-667.

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.

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.

Hollingsworth PM. 2011. Refining the DNA barcode for land plants. Proceedings of the National Academy of Sciences **108**: 19451-19452.

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.

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, Erickson DL. 2008. DNA barcodes: Genes, genomics, and bioinformatics. Proceedings of the National Academy of Sciences **105**: 2761-2762.

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.

Kuch M, Rohland N, Betancourt JL, Latorre C, Steppan S, Poinar HN. 2002. Molecular analysis of an 11,700-year-old rodent midden from Atacama Desert, Chile. Molecular Ecology **11**: 913-924.

Lee SY, Ng WL, Mahat MN, Nazre M, Mohamed R. 2016. DNA Barcoding of the Endangered Aquilaria (Thymelaeaceae) and its Application in Specie Authenticaton of Agarwood Proucts Traded in the Market. PLoS ONE 11 (e0154631) DOI: 10.1371/journal.pone.0154631.

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.

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.

Meusier I, Singer GA, Landry JF, Hickey DA, Hebert PDN, 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.

Mora C, Titterson DP, Adl S, Simpson AGB, Worm B. 2011. How many species are there on Earth and in the ocean? PLoS Biology 9 (e1001127) DOI: 10.1371/journal.pbio.1001127.

Muellner AN, Schaefer H, Lahaye R. 2011. Evaluation of candidate DNA barcoding loci for economically important timber species of the mahogany family *(Maliaceaceae)*. Molecular Ecology Resources **11**: 450-460.

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

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

Park KT. Four new species of Torodora Meyrick and a new species of Antiochtha Meyrick from the Philippines (Lepidoptera, Lecithoceridae). Journal of Asia-Pacific Entomology **11**: 89-95.

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

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. Přírodovědécké fakulta, Masarykova univerzita, Brno. Available from https://is.muni.cz/do/rect/el/estud/prif/js13/genetika/web/index.html (accessed March 2018).

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.

Saunders GW. 2005. Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. Philosophical Transactions of the Royal Society B: Biological Sciences **360**: 1879-1888.

Savolainen P, Lundeberg J. 1999. Forensic evidence based on mtDNA from dog and wolf hairs. Journal of Forensic Science **44**: 77–81.

Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R. 2005. Towards te writing the encyclopedia of life: an introduction to DNA barcding. 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.

Smith SA, Donoghue MJ. 2008. Rates of molecular evolution are linked to life history in flowering plants. Science **322**: 86-89.

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.

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.

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.

Thomson KA, Newmaster SG. 2014. Molecular taxonomic tools provide more accurate estimates of species richness at less cost than traditional morphology-based taxonomic practises in vegetation survey. Biodiversity and Conservation **23**: 1411-1424.

Yu M, Jiao L, Guo J, Wiedenhoeft AC, He T, Jiang X, Yin Y. 2017. DNA barcoding of vouchered xylarium wood specimens of nine endangered *Dalbergia* species. Planta **246**: 1165-1176.

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.

Wernegreen JJ. 2012. Endosymbiosis. Current Biology **22**: 555-561.

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.

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

Williamson MH. 1996. Biological invasions. Chapman and Hall, London.

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