Egyptian Journal of Chemistry
http://ejchem.journals.ekb.eg/

The Power of DNA Barcoding for Plant Identification

Samira A. Osman*

Genetics and Cytology Department, National Research Centre, Giza, Egypt

Abstract

Plant taxonomy is considered an important branch of science; it needs highly professional taxonomists to classify plants depending on their morphological characteristics but this method has numerous disadvantages, it is time-consuming and highly dependent on plant growth stages and reproductive organs (such as tree species with a short bloom period or rarely blooming). Therefore, several molecular techniques (DNA fingerprinting, DNA barcoding, Next generation sequencing) have been used to overcome these obstacles. DNA barcoding is considered the bedrock to classify and discover different plant specimens by combining the strengths of molecular genetics, sequencing technologies and bioinformatics by using highly variable short regions of DNA. Four gene regions (*rbcL*, *matK*, *trnH-psbA* and *ITS*) have been utilized as the universal DNA barcodes for plants. The combination of two or more barcode regions improves the resolution level and provides perfect huge data for species identification. DNA barcoding faces many obstacles for discrimination of some plant specimens for example the presence of conserved ancestral alleles, high intraspecific variability and some other factors, As an outcome, it was proposed to use complete chloroplast sequences using next-generation sequencing (NGS) technique. NGS technique is highly expensive thus it was needful to use other new and cheap techniques such as genome skimming and twin track processes.

Keywords: Plant identification, Nuclear barcodes, Chloroplast barcodes, DNA mini-barcode, next-generation sequencing (NGS).

1. Introduction

Plant taxonomy is considered an important branch of science; the classification of plants is the first step in studying the biodiversity, breeding, conservation and development of populations and distinguishing different species. There are several ways to the taxonomic identification of plant specimens starting from morphological, PCR-based to sequence-based techniques. The traditional way for species discovery and identification is depending on the morphological characters' investigation. But this approach is not accurate for the classification of some plant species, because it is time-consuming, highly dependent on plant growth stages and reproductive organs (e.g. tree species with rarely or short bloom period), so this method needs highly professional taxonomists. The molecular techniques could overcome all these obstacles, where these techniques provide independent sets of data that could be used to determine the phylogeny of plants and determine plant systematic and identification **[1]**.

Biological diversity (biodiversity) involves genetic variability among species and ecosystems. Most biodiversity studies focused on species because species are considered the primary unit of biodiversity and it is easy to make comparisons among them quantitatively **[2]**.

Plant molecular characters are derived from three different sources, genetic fingerprints, structural genomic characteristics (e.g. gene order, gain or loss of genes, or non-coding regions), and finally, DNA sequences of specific coding or non-coding regions from one of the three plant genomes plastid, mitochondrial or nucleus **[3]**.

Different molecular markers (RAPD, ISSR, SSR, SCoT, IRAP…. etc) are used for performing DNA fingerprints by amplification of certain sequences of genomic DNA using polymerase chain reaction (PCR). DNA fingerprints give perfect results for discrimination of different samples within a defined group or samples having specific genes and studying their phylogenetic relationships or biodiversity but this method is imperfect for the identification of unknown or cryptic taxa. So, a sequence-based technique such as DNA barcoding was applied to overcome this obstacle **[4]**.

DNA barcoding is considered as the bedrock for the discrimination of species relationships and phylogeny at a molecular level by investigating sequence

*Corresponding author e-mail: s_nrc82@yahoo.com
Receive Date: 07 May 2022, Revise Date: 17 June 2023, Accept Date: 10 July 2023
DOI: 10.21608/EJCHEM.2023.137216.6049

variation in standardized gene loci which represents a small region from the complete genome **[5]**. In 2003, this technique was proposed as a tool for animal species discrimination, afterward it was used to identify other multicellular organisms (plants, fungi, protists) **[6]**.

2. Standard DNA Barcodes

DNA barcoding is a molecular technique that is performed by combining the strengths of molecular genetics, sequencing technologies and bioinformatics. This technique uses highly variable short regions of DNA to classify and discover different plant species in addition to provide fast and accurate species identification **[7]**. Although, there are no universal barcodes, the variable loci that are listed enable the scientists to differentiate plant samples from various taxonomic groups **[8]**. After a broad inventory of gene loci in the mitochondrial, plastid and nuclear genomes, four primary gene loci (*rbcL*, *matK*, *trnH-psbA* and *ITS*) were intensively be utilized as the standard DNA barcodes for plants applications **[9]**.

In animals, the nuclear DNA represents about 98% while mitochondrial DNA represents about 2% of total genomic DNA. Mitochondrial DNA contains a highly variable region *COI* (cytochrome oxidase subunit I) that is utilized as the animal barcode. *COI* is involved in the electron transport phase of respiration. The evolution of plant mitochondria is very slow compared to the animal mitochondria, so the region of the cytochrome oxidase (*CO1*) gene was used as a barcode marker for animals but not as a plant barcode to discriminate different plant specimens **[10]**.

In plants, DNA barcoding looks to be inherently more complicated than in animals **[11]**. Plant DNA is also, present in chloroplasts (cpDNA) besides nuclear and mitochondrial DNA. The related plant species which cannot be distinguished morphologically could be differentiated at the molecular level with the help of plant DNA barcodes loci.

Many loci and combinations thereof have been proposed as appropriate barcodes for plants. For example, the nuclear Internal Transcribed Spacer (*ITS*) loci of nuclear ribosomal genes and chloroplast DNA sequences were considered; the chloroplast genome consists of many barcode regions which classified as coding sequences (*matK*, *rpoB*, *accD*, *ndhJ*, *ccsA* ,and *rbcL*) and non-coding sequences (*trnH–psbA*, *atpF–atpH* and *psbK–psbI* spacers) as shown in **figures (1 and 2) [14]**.

The coding chloroplast barcodes *matK*, *ndhJ*, *accD* and *rpoB* encode proteins of maturase K, a nicotinamide dehydrogenase, β' subunit of acetyl-CoA carboxylase and β subunit of RNA polymerase respectively, while *ccsA* gene encodes a protein that involved in the cytochrome c biosynthesis and the *rbcL* gene encodes the large subunit of ribulose-bisphosphate carboxylase (EC 4.1.1.39) **[12]**. However, the non-coding chloroplast barcodes should be treated as suboptimal barcodes because of the prevalence of microsatellites **[13]**.

In this technique, the nucleotide sequence was analyzed by using different bioinformatics programs where the nucleotide sequences give huge data and each nucleotide site considers as highly specific genetic information. The resulting DNA barcoding data has allowed greatly improving the ability to index biodiversity and has helped to increase the understanding of species distributions. However, great effort is still necessary to establish a DNA barcode reference library for all species **[1]**.

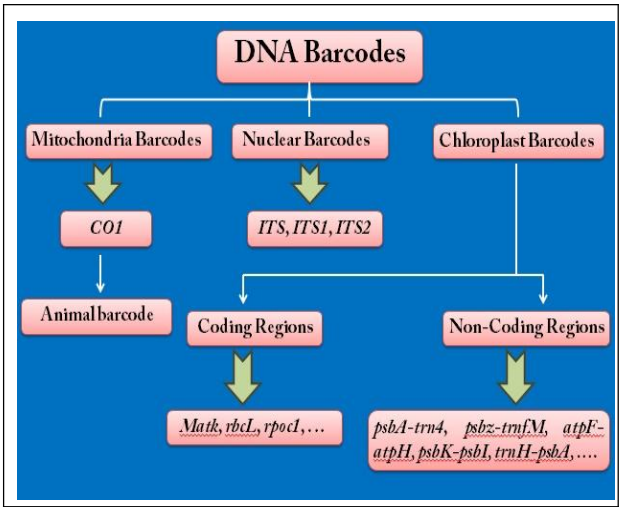

Figure (1): Schematic illustration of DNA barcode types.

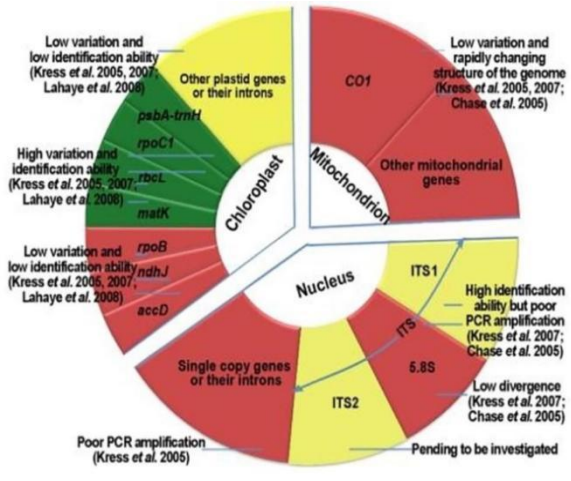

Figure (2): DNA barcode genes from Mitochondrial, chloroplast and nuclear genomes of plants. Green color represents a powerful barcodes, red color represents poor barcodes and yellow represent barcode regions yet to be investigated **[14]**.

2.1. Nuclear barcodes

2.1.1. Internal Transcribed Spacer (ITS) region

The nuclear genome contains an internal transcribed spacer (*ITS*), it is present as tandem repeats that encourage the amplification, detection and sequencing of nuclear ribosomal DNA (nr DNA) also this region gives the best result in PCR compared to other barcodes **[9]**.. *ITS* is considered a nonfunctional RNA sequence with a length of about 700 bp which includes 5.8S rRNA region with a length of about 163 or 164 bp. *ITS* exists between the structural ribosomal RNA during the maturation of rRNA and this spacer consider the product of maturation so any mutation in this spacer leads to the inhibition of mature small and large rRNA subunits. There are two types of this spacer (*ITS1* and *ITS2*) which are located between 25S and 18S rRNA coding regions. *ITS1* is located between 5.8S and 18S rRNA with a length of less than 300 bp while *ITS2* located between 5.8S and 25S rRNA with a length less than 250 bp as shown in **figure (3) [15]**.

ITS2 region is considered a universal barcode in animals and plants, it is utilized for the identification of Pan-eukaryote and Eukaryote. **Ceriaco *et al.*, [16]** mentioned that the *ITS2* barcode was utilized for distinguishing the morphologically similar species in addition to the contaminants present in the North American herbal products identified using combined *ITS2* and *rbcL* barcodes.

Indeed, some researchers elucidated that the nuclear *ITS* locus supplied better resolution than chloroplast barcode loci **[17]**. Some tropic evergreen shrubs and trees of Brazilian species from the family Sapotaceae, which are characterized by short-lived flowers (rarely bloomed), showed high resolution of *ITS* for discrimination of these species **[18]**.

Several applied research of medicinal plants appeared that the *ITS2* locus is better than *ITS1* and the combination of *ITS1* and *ITS2* for species identification **[19]**. The resolution of the *ITS1* and *ITS2* loci and their combination with other DNA barcodes were evaluated for discrimination of *Primula* species and taxa of the family Lauraceae **[20]**. In mosses and the group of legume plants from the subtribe Cassiinae, The *ITS1* region provided better identification of different taxa than *ITS2* and the combination of *ITS1* and *ITS2* **[21]**. Thus, the usability of barcode loci may vary, even with an equal rate and character of their evolution, relying on the task of the study.

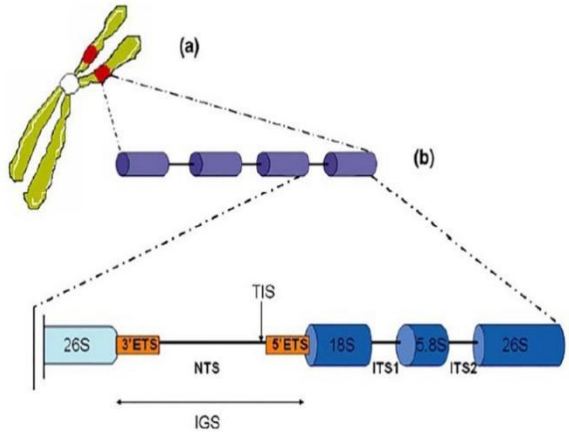

Figure (3): The universal structure of the rDNA regions in plants **[22]**.

(a) The location of rDNA region on the chromosome.
(b) Tandem arrays of the consecutive gene blocks (18S-5.8S-26S).
IGS: Intergenic Spacer.
ETS: External Transcribed Spacer.
NTS: Non-transcribed region
TIS: Transcription start site.
18S: The small subunit of rRNA.
5.8S and 26S: The large subunit of rRNA.
ITS: The internal transcribed spacer (ITS1 and ITS2).

2.2. Chloroplast barcodes

In the 1960s and early 1970s, it was confirmed that plastids are polyploidy, whereas proplastid has about 20 genome copies while chloroplast has about 100 copies **[23]**. The chloroplast genome is a circular double-stranded DNA molecule **[24]**.

The chloroplast genome contains two inverted repeats (IRs), a small single-copy region (SSC) and a large single-copy region (LSC) as shown in **figures (4 and 5)**. In plants, the length of IRs ranges in size from 5 to 70 Kb, it contains highly variable regions **[25]**.

Plant DNA barcoding studies were limited to the chloroplast genome based on the variation in the sequence of the non-coding region (*psbA-trnH*) and coding region (*matK*, *rbcL* and *rpoC1*). It was confirmed by previous studies that the chloroplast genome has the ability to identify and discriminate different plant specimens because this genome undergoes a lot of variations and substitutions **[26]**.

The phylogenetic analysis and molecular evolution of different *Panax* species were performed using chloroplast inter-genic spacer (IGS) such as *trnE-trnT*, *trnT-psbD*, *ndhF-rpl32* and *rpl14-rpl16* as chloroplast barcodes. Some non-coding regions (*atpF-atpH*, *psbK-psbI* and *trnH-psbA*) and coding regions (*rbcL* and *matK*) were used as barcodes to discriminate different *Orchidaceae* species from

Korea **[27]**. **Kress *et al.*, [28]** used the chloroplast inter-generic spacer *psbA-trnH* for identification of different *Dendrobium* species.

The intensive research was performed to discover the variable loci that provide more resolution than the universal DNA barcodes, when several of completely sequenced chloroplast genomes recorded in the GenBank increased. The *ycf1* gene is the second largest gene in the genome of chloroplast with an average of length about 6000 bp. This gene alone demonstrates better resolution than the combination of *trnH–psbA+matK +rbcL* for discrimination of 20 species from 67 families that comprised angiosperms, gymnosperms and mosses **[29]**.

2.2.1. The matk region

Maturase K (*matk*) gene presents in the chloroplast genome with a size 1500 bp and it encodes to maturase-like protein with 500 amino acids that assist with RNA editing. *Matk* gene embedded in the intron region of *trnk* while the two exon regions of *trnK* were flanked by the maturase (*matK*) protein (act as splicing factor) then it was cut down and keeping the *matk* gene intact during the splicing process **(Figure 6)**. *Matk* is a promising gene due to its high substitution rate and it manifests different numbers and sizes of indels (insertions and deletions) so it could be used in molecular systematic and determine the evolution of some plants to resolve some taxonomic problems **[10]**.

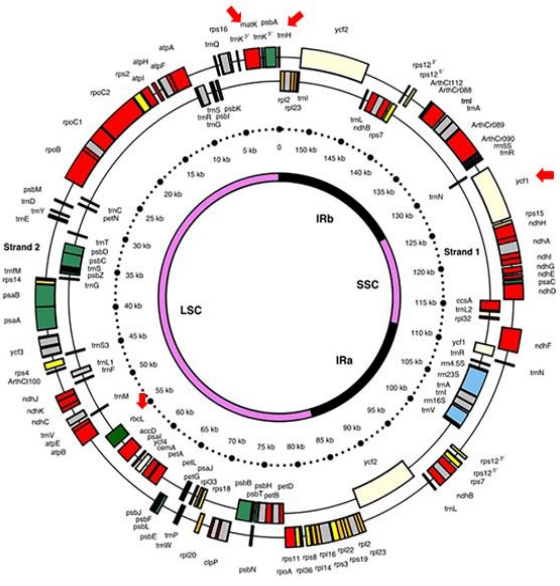

Figure (4): Chloroplast DNA (cpDNA) map of *Arabidopsis thaliana* (Original picture author: **prof. Emmanuel Douzery) [30]**.
- **IRA&IRB:** Inverted Repeat(A&B)
- **SSC:** Small Single Copy regions
- **LSC:** Large Single Copy regions

- **Yellow color** indicates small subunit ribosomal proteins.
- **Orange color** indicates large subunit ribosomal proteins.
- **Lemon color** indicates hypothetical chloroplast open reading frames.
- **Green color** indicates protein-coding genes either involved in photosynthetic reactions.
- **Red color** indicates other function.
- **Blue color** indicates ribosomal RNAs.
- **Black color** indicates transfer RNAs.
- **Gray color** indicates introns.

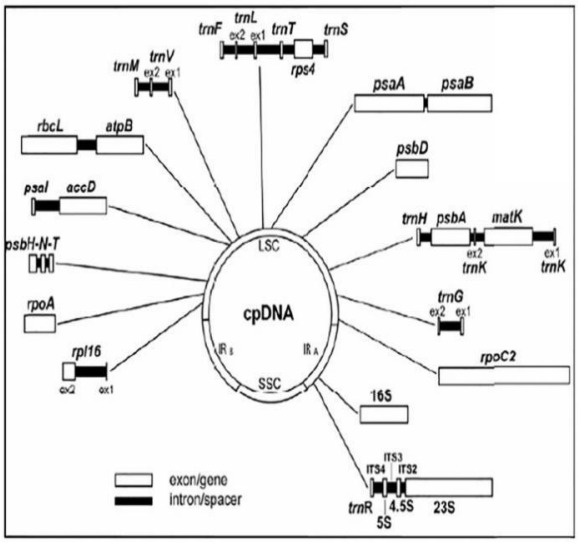

Figure (5): The positions of coding and non-coding loci on chloroplast genome **[3]**.
IRA&IRB: Inverted Repeat(A&B)
SSC: Small Single Copy regions
LSC: Large Single Copy regions

Johnson and Soltis [31] mentioned that the *matK* gene is a rapidly evolving gene because its rate of nucleotide substitution is higher than that of the large subunit of Rubisco (*rbcL*) by three folds while its amino acids substitution is higher than the substitution of *rbcL* amino acid by six folds This refers to the effectiveness of *matK* gene for resolving evolutionary relationships among plants at all taxonomic levels.

In the year 2010, the Linnaean Society of London used the analysis of the *matK* gene on a large scale. They use the available primer to obtain the match of generated *matK* sequence to make the possible modification and to simplify the barcode complications in the family Zingiberaceae **[32]**.

Lahaye *et al.* [33] used primers specific for *matk* gene amplification of 1667 angiosperms. Also, more than 90% of species in the Orchidaceae could be discriminated by *matK*. **Fazekas *et al.* [34]** used *matK* barcode for the identification of 92 species from 32 genera but only achieved a success rate of 56%.

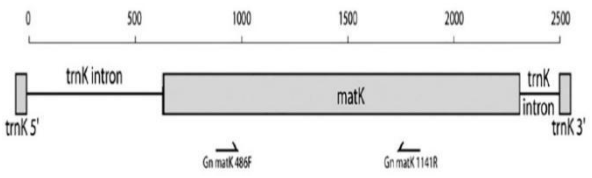

Figure (6): Relative place of *matK* loci. Coding loci were represented by boxes and connections showed spacer loci **[35]**.

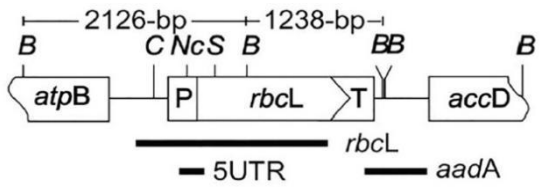

Figure (7): The region of the *rbcL* gene in the plastome of the tobacco plant **[14].**

2.2.2. The RuBisCo (rbcL) region

The *rbcL* region can be amplified, sequenced and aligned in most plants thus it considers an excellent DNA barcoding region in the plant at the level of family and genus so it can detect their evolutionary relationships. In the plant cells, the plastid has many chloroplasts; each chloroplast genome contains only one copy of the *rbcL* gene with a length of 600-750 bp. This gene consists of only one exon **(Figure 7)** and it encodes the polypeptide called RuBicCo **[36]**. Ribulose 1,5bisphosphate carboxylase oxygenase "RuBisCo" (EC: 4.1.1.39) is an important enzyme responsible for carbon fixation by converting atmospheric carbon dioxide to glucose (molecules rich with energy) in photosynthetic organisms. In plants, this enzyme has two types of protein subunits one of them with a large chain and another with a small chain. The large chain subunit is encoded by the *rbcL* gene that is present in the plant chloroplast genome while the small chain subunit is encoded by several related genes in the nucleus genome. The large chain contains the binding sites that form a dimer with a substrate (ribulose 1,5-bisphosphate) to form glucose. In some proteobacteria and dinoflagellates, this enzyme contains only a large chain subunit **[37]**.

Kress *et al.*, [35] showed that *rbcL* region length is a drawback where its double-stranded sequencing needs four primers **[38]**. However, despite its limitation *rbcL* is still being wildly used for plant barcoding because it gives a huge amount of data and its entire gene sequence could be recovered so *rbcL* barcode region in combination with other various chloroplast or nuclear barcode regions have the ability for accurate identification of plants. **Fouad et al., [39]** used *rbcL* locus sequence for determining the taxonomic relationships among twenty different plant species which were collected from Burg El Arab (lat. 30°54′N, long. 29°33′E), which belongs to the Mareotis sector of Mediterranean sand dunes coast west of Alexandria in Egypt.

2.2.3. The trnH-psbA region

The intergenic spacer *psbA-trnH* is considered an important DNA barcode because its end near to *psbA* is highly conserved but it appeared a big inversion near to *trnH* **(Figure 8)**. The conserved region is difficult to be used for the identification of different taxonomic groups that are highly diverged in their nucleotide variation **[40].**

The *psbA* gene encodes the photosystem II reaction center protein D1 which affects many plant physiological stages, light intensity photosystem and plant developmental stages. There are many studies since more than 20 years on the role of chloroplast *psbA* UTRs in the regulation of gene expression. This region has flanking primers that are involved in the amplification and sequencing through high molecular size. Also, this region has an effective role in analyzing the degraded DNA that is suggested to contain the desirable barcode. In some previous studies on 21 genera and 18 families in plants, *trnH-psbA* is effective for the identification compared to other barcodes *matK*+*rbcL* **[41].**

The *trnH–psbA* spacer is distinguished by a great number of inversions and indels this leads to a complicated alignment as in genus *Gentiana* **[17]**. In closely related species, this region can be varying in its length significantly. The *trnH–psbA* can provide a better resolution, when used alone or in combination with *ITS*, than *matK* and/or *rbcL* **[42]**. **Burgess *et al.*, [43]** documented that the *trnH–psbA* spacer can slightly improve the resolution for the combination of *matK* + *rbcL* region.

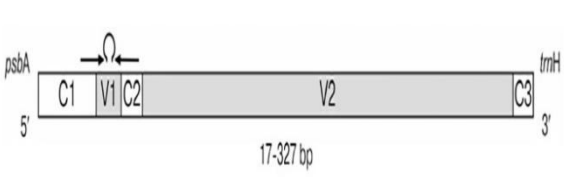

Figure (8): Schematic represented the arrangement of the *psbA-trnH* spacer, the variable regions (V1 and V2) and conserved regions (C1, C2 and C3) **[44].**

Recently, the two loci combination of *rbcL+matK* and *ITS* have been recommended as the essence barcode for land plants **[45]**. Also, several researches documented that this combination led to perfect species identification, although the resolution level for discrimination of some genera such as *Crocus*, *Berberis* and *Primula*, was insufficient **[43]**. Chinese scientists performed a comparative study among 6286 plant specimens from 141 genera and 1757 species using nucleotide sequences of various five DNA loci which are considered to be potential DNA barcodes. The combination of *MatK* and *rbcL* barcodes are considered species-specific in 49.7% cases **[46]**. **Abouseada et al., [47]** evaluated the genetic relationship among twelve wheat cultivars three of them from Egypt and two from Mexico while others cultivars from Morocco, Sudan, India, China, Pakistan, Syria and Australia using DNA barcoding based on the combination of *rbcL* and *matK* loci sequences.

Several studies found that the resolution of intrageneric was increased by the addition of either *ITS* region or *trnH–psbA* spacer in families Palmaceae or Labiatae respectively, **[48]**. Also, **Feliner and Rosselló, [49]** suggested that the well-studied *ITS* loci should be added to DNA barcodes. Many Chinese researchers analyzed *ITS* data for thousands of taxa from several families and reported that *ITS2* to be a highly effective DNA barcode **[50]**. As a result in 2011, it was suggested to add an alternative DNA barcodes (nuclear *ITS1* and *ITS2* loci and the chloroplast *trnH-psbA* spacer) to the list of standard DNA barcodes **[46]**. Obviously, the chloroplast barcodes usage had some limitations; it was observed that perfect results could be likely gained with highly variable low-copy-number nuclear genes, but, due to their different variability among some species and various other obstacles, none of them could be selected to be a DNA barcode **[51]**.

First community phylogenetic for the tree of 281 species of woody plants in Barro Colorado Island in Panama, depending on a super matrix analysis of combined data of *rbcL*, *matK* and *trnH-psbA* sequence, was published. That led to adding well-supported evolutionary components for knowing species diversity **[52].**

3. Obstacles overcoming and Future outlook of DNA barcoding

An attempt to utilize the nuclear barcodes to identify different species within the evolutionarily young group for instance *Hawaiian* species from two genera (*Cyrtandra* (family Gesneriaceae) and *Clermontia* (family Campanulaceae)) was failing, due to conserved ancestral alleles, high intraspecific variability and some other factors **[53]**. In spite of several attempts to utilize DNA barcodes and some additional chloroplast barcodes, no satisfactory species differentiation was progressed among the genus *Araucaria* from the New Caledonia islands **[54]**. As an outcome, it was suggested to utilize complete sequences of the chloroplast genome for taxa discrimination. A few years ago, scientists could not dream of it, but novel technologies such as **next-generation sequencing** (NGS) gave scientists with such a chance. The complete chloroplast genomes of 28 accessions from 11 New Caledonian *Araucaria* species were sequenced with NGS technology. These species were discriminated by chloroplast genomes (~147000 nucleotides) and 11 nuclear genes (>6000 nucleotides in total), they found that the differentiation of these species was the highest in the case of chloroplast genomes. This result did not satisfy the researchers because more than half of New Caledonian *Araucaria* species with multiple accessions were monophyletic according to the plastid or nuclear phylogenetic trees although the species were not monophyletic **[55]**. So the researchers concluded that the differentiation of young taxa needs several variable nuclear barcodes. Nevertheless, the utilization of complete chloroplast genomes was very charming.

There is another obstacle facing DNA barcoding in plants, this obstacle is the quantity and quality of DNA extracted from plants that depend on many factors such as plant's age (DNA quantity is very low in plants older than 50 years), methods of drying plant samples, pesticides used is storage process and finally, the taxonomic group of the plant. For instance, herbarium samples from the Boraginaceae family have a problem in DNA extraction and also medicinal herbal teas and other plant raw materials had degraded DNA **[56]**. Thus, DNA mini-barcodes (short nucleotide sequences less than 200 base in length) were progressed for the Sanger analysis of plants that contain damaged DNA such as stored plants and old herbarium samples. The most variable short DNA regions (DNA mini-barcodes) for these species were found after knowing the sequences of the complete chloroplast genome by Next generation sequences (NGS). For instance, the discrimination of *Panax ginseng* and *Panax notoginseng*, this performed depending on some variable regions as *rps16* gene (280 nucleotides), *ycf1a* gene (60 nucleotides), or *ycf1b* gene (100 nucleotides) with an identification power of 83.33%, 91.67% and 100% respectively **[57]**.

4. Next Generation Sequencing (NGS)

Next Generation Sequencing (NGS) is a recent technique that covers the whole genome sequences (sequencing depth) and it is considered an index reflecting how many times a nucleotide was sequenced. The higher the index was, the higher was

the chance to avoid mistakes during the assembly of the whole sequences of the genome. This technique is highly expensive so it was necessary to search for another new and cheap technique such as **the genome skimming** process. This process is characterized by a small sequence depth and low quality of DNA sequencing, yet, it remains suitable for a satisfactory assembly of repeats (organelle genomes and ribosomal genes). As a result by this process, ~1 Gb of sequences can be gained **[58]**. Genome skimming cannot determine rare variants so, it similar to the Sanger sequencing **[59]**.

There is another technique was suggested for the sequencing of low-copy nuclear genes **[60]**. Therefore, with the advance in sequencing techniques, the ideas for the utilization of a novel genome data format for DNA barcoding appeared. These ideas lead to the formulation of the so-called **a twin track process**, which had (1) the standard DNA barcode accumulation and their libraries and (2) enlarged barcode application by genomic skimming **[61]**. An enlarged barcode was achieved by NGS, which represents the sequences of the chloroplast genome and ribosomal genes together.

The next-generation sequencing technique had many important advantages firstly it had the ability to read sequences of chloroplast and nuclear markers of numerous specimens at one time this is important for the identification of herbal complements **[56]**. Secondary, NGS technique is used in the analysis of the herbarium **[59]**. This test permits researchers to obtain enlarged barcodes of old samples. Finally, the analysis of the DNA of museum samples (including herbarium) is specified, this is known by the term "museomics" **[62]**.

At the beginning of DNA barcoding studies, the scientists discussed the impossibility to utilize DNA barcodes for the differentiation of closely related species **[51]**. DNA barcoding could be used to differentiate the lower taxonomic levels (subspecies and varieties) and this is called "ultra-barcoding", this is performed when it became possible to use ribosomal genes and complete chloroplast sequences loci sequenced by NGS **[63]**. Until now, it is still a continuous search for the optimal DNA barcodes in plants.

5. Applications of DNA barcoding

Numerous studies mentioned that the applications of DNA barcoding on plants can be split into two categories (plants' taxonomic process and identification of unknown plant specimens). Also, DNA barcoding can be applied in other several fields **[64]** and **[65]** as follows:

5.1. Plant ecological forensics

DNA barcoding was utilized for plant identification from roots or leaves and cryptic life stages such as fern gametophytes**.**

5.2. Agricultural pests control

DNA barcoding is one of the new molecular techniques that have been used to determine the type of pests and control them, since many agricultural pests threaten the growth and productivity of most plants, these consider a source of concern for many farmers. Therefore, it was necessary to determine the pests at any stage of their life. In addition, the DNA barcoding technique contributes to identify and control fruit flies.

5.3. Disease vectors identification

DNA barcoding contributes to determine vector species. The vectors can cause dangerous infectious diseases in humans and animals. This technique plays an important role to know these diseases and deal with them. DNA barcoding for mosquitoes helps construct a reference barcode library. This library can help public health officials effectively control these diseases caused by a mosquito with very little use of insecticides.

5.4. Invasive foreign species determination

Identifying and controlling foreign species that can affect ecosystems and native species. Thus early discovery and regulatory measures are very important to stop the cross-border transport of foreign species, this can be performed by utilizing DNA barcoding technique for alien species identification.

5.5. Natural resources sustaining

Natural resource managers used DNA barcoding techniques and barcode libraries to monitor illegal commerce of products that are made of natural resources such as hardwood trees.

5.6. Endangered species protection

Due to bushmeat hunt in Africa, Primate Population decreased by 90%. DNA barcoding could be utilized to perform law enforcement on bushmeat in local markets to protect Primate from extinction.

5.7. Water quality observation

Water is very important for living beings, it is responsible for maintaining all living cell functions. So, it is necessary to study the quality of water. Several organisms live in water sources (streams, rivers and lakes). DNA barcoding could be used to identify and create a library for these species

especially for the species that are difficult to discriminate. Finally, DNA barcoding is a helpful technique for environmental agencies to create better policies that can ensure the safe supply of drinking water.

6. DNA barcode protocol

DNA barcode analysis comprises three major steps. These steps are sample collection, molecular analysis and bioinformatics as summarized in **figure (9).** The final data analysis will led to the discovery of numerous unknown plant species **[26]**.

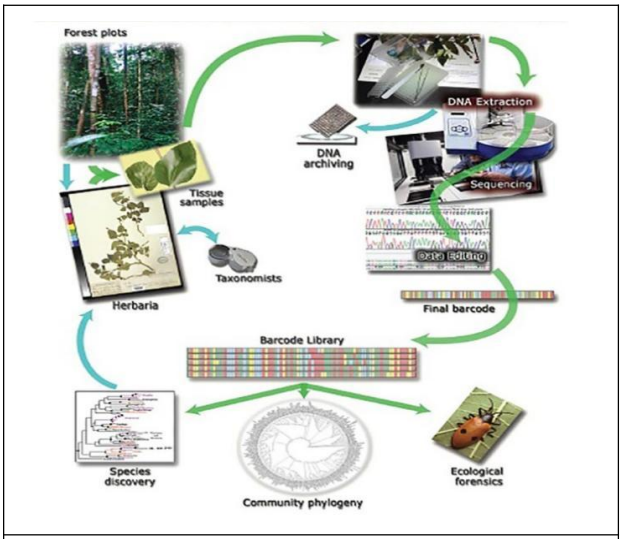

Figure (9): Graphic diagram summarized the steps of Plant DNA barcoding protocol **[66].**

6.1. Plant samples collection

The collected plant samples could be obtained from in herbaria or picked directly from the field (fresh life samples).

6.2. Plant genomic DNA Extraction

- Genomic DNA was extracted from young leaves of samples by DNA purification Kits or by manual methods such as CTAB.

- The extracted DNA was checked on agarose gel by electrophoresis and then spectrophotometry (NanoDrop 2000) afterward, diluted to 50ng/µl and used as a PCR template **[67]**.

6.3. In silico analysis and primers design

The primers design could be summarized in the following steps**:**

a- The desired barcode gene sequences of certain species (which were recorded in the barcode library (NCBI) by previous research) were retrievable from the National Center for Biotechnology Information (NCBI) database (GenBank). These sequences have already accession numbers were recorded in the gene library.

b- These gene sequences were downloaded and saved in FASTA format files after that the downloaded sequences were aligned by certain bioinformatics programs such as the MEGA program to determine the similar part among all aligned sequences

c- This similar part of sequences was used to design the primer by the online program Primer 3 (version 4) (http://bioinfo.ut.ee/primer 3-0.4.0).

d- Afterward the designed primer was tested *in silico* by aligning the retrievable complete desired barcode the gene sequence of the same studying species with the designed primer to insure that this primer was already a specific primer for the desired barcode and attached with barcode gene (understudying) by 100%.

6.4. Polymerase Chain Reaction

The standard DNA barcodes are characterized by easy amplification of polymerase chain reaction (PCR) and its length is short enough to permit a bidirectional read during the sequencing process **[34]**. The amplification of desired barcode region was performed in the Thermal Cycler apparatus by the following steps:

a- PCR mixture: genomic DNA (50 ng), dNTP (0.2 mM), designed primer (10 pM), MgCl, Taq Buffer (5x) and Go Taq DNA Polymerase (1U).

b- The used PCR program depends on the type of plant species and type of the desired barcode for instant the PCR program for amplification of matk gene in different *Triticum* species is 95°C for 4 min followed by 35 cycles of 95°C for 30 sec, 57°C for 1 min and 72°C for 1 min and a final extension at 72°C for 5 min **[68]**.

c- PCR products were loaded onto 1.5% agarose gel which stained ethidium bromide.

d- The gel photo was captured by gel documentation and Bands were analyzed by gel analyzer or TotalLab programs as shown in **figure (10).**

e- The resulting barcode fragment was cut and cleaned up from agarose gel to obtain the purified barcode gene ready to perform sequencing by Sanger sequencer.

6.5. Sequence Editing

The chromatogram (electropherogram or trace file) of the nucleotide sequences was obtained by Sanger sequencer. The formats of chromatograms file are ABI (Applied Biosystems sequence) and SCF, which require certain software to be opened and analyzed. This software may be free such as BioEdit (https://bioedit.software.informer.com/), Chromas (http://technelysium.com.au/wp/chromas/) andFinchTV(https://finchtv.software.informer.com/1.4/), while the commercial software includes

CodonCode (https://www.codoncode.com/), Geneious (https://www.geneious.com/) and ChromasPro(http://technelysium.com.au/wp/chromaspro/). By all these software, simple sequence editing and extensive analytical tools can be performed. The barcode gene is sequenced bi-directionally (forward and reverse) to recover the entire DNA sequence, when creating a DNA barcode library. The workflow of sequence editing is summarized in **figure (11 and 12).**

The isolated DNA barcode gene sequences have been deposited in NCBI and take certain accession numbers.

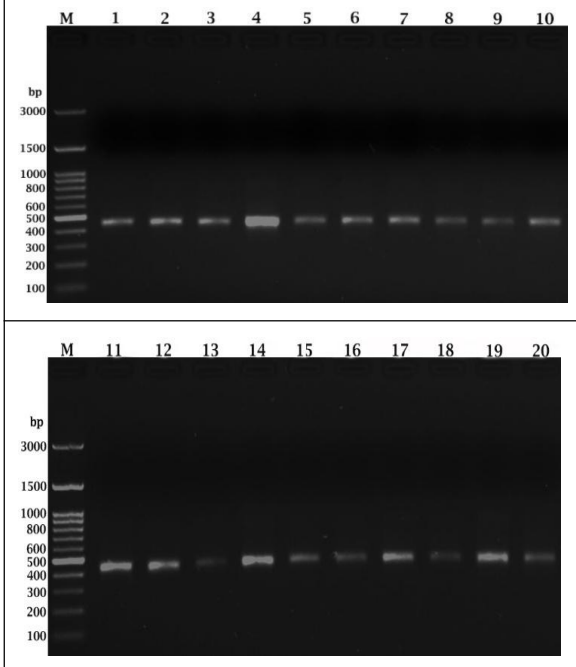

Figure (10): *matK* gene from twenty *Triticum* species **[68].**

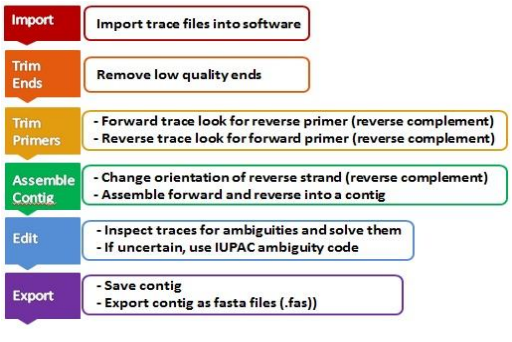

Figure (11): Scheme summarized the sequence editing workflow.

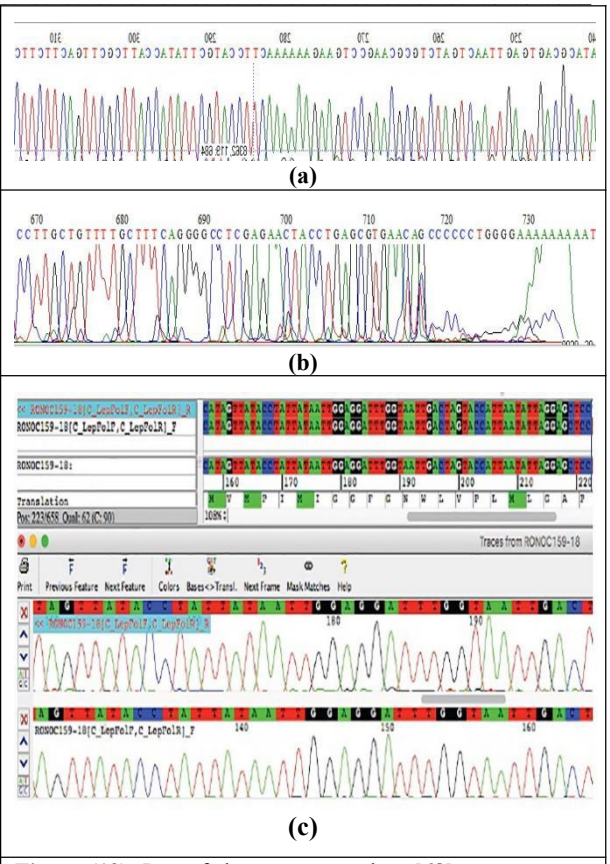

Figure (12): Part of chromatogram chart **[68].**
(a) Showing very good sequence.
(b) Showing bad sequence at the end of sequence.
(c) Showing a trim end and recover the entire DNA sequence using BioEdit software.

6.6. Phylogeny analysis

- The recovered nucleotide sequences were aligned by the Clustal W multiple sequence alignment using one of bioinformatics program (Mega program) as shown in **figure (13).**

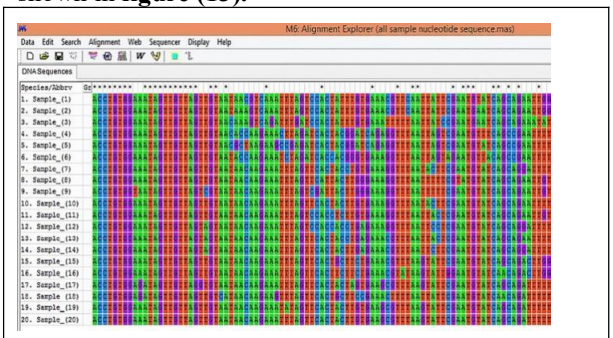

Figure (13): Part of *matK* sequence alignment in twenty different *Triticum* species from1 to 99 bases **[68].**

The phylogenetic relationships tree was provided for species identification or discovery; this was performed by several statistical data analyses such as Bootstrapping (1000 replications) and Pairwise distance for nucleotide sequences that align by ClusterW as mentioned in the previous step.

Also a huge data could be calculated by various bioinformatics programs such as the total length of nucleotide sequences, evolutionary divergence among sequences, nucleotide composition percentage and polymorphism calculation, Maximum Likelihood of Transition/Transversion Bias and Maximum Likelihood of Substitution Matrix **[69]**.

The amino acid sequences were obtained from the translation of the nucleotide sequences by using an online program such as ExPASy (https://web.expasy.org/translate).

Also the amino acid sequences were aligned by using the Clustal W multiple sequence alignment programs such as the MEGA program to construct the phylogenetic tree.

7. Conclusions

DNA barcoding technique combines the strengths of molecular genetics, sequencing and bioinformatics by using highly variable short regions of DNA. By studying the mitochondrial, plastid and nuclear genomes, four gene regions (*rbcL*, *matK*, *trnH-psbA* and *ITS*) were utilized as the universal DNA barcodes for plants. The combination of two or more barcode regions increases the resolution level and gives huge data for species identification. DNA barcoding faced several obstacles to differentiate numerous plant samples for instant, the presence of conserved ancestral alleles, high intraspecific variability and some other factors, As an outcome, it was proposed to use complete chloroplast sequences by next-generation sequencing (NGS) technique.

8. Acknowledgments

The author is grateful to all the researchers whom we cited in this review for their significant and valuable research.

9. References

[1] Vu H, Le L (2019) Bioinformatics Analysis on DNA Barcode Sequences for Species Identification: A Review. **Annual Research and Review in Biology 34(1):** 1-12. DOI: 10.9734/ARRB/2019/v34i130142.

[2] Mora C, Tittensor DP, Adl S, Simpson AGB, worm B (2011) How many species are there on Earth and in the ocean?. **PLoS Biology 9(8):** e1001127. https://doi.org/10.1371/journal.pbio.1001127.

[3] Stech M, Quandt D (2010) 20,000 species and five key markers. the status of molecular bryophyte phylogenetics. **Phytotaxa 9(1):** 196–228. DOI:10.11646/phytotaxa.9.1.11.

[4] Kim JS, Kim HT, Son S-W, Kim J-H (2015) Molecular identification of endangered Korean lady's slipper orchids (Cypripedium, Orchidaceae) and related taxa. **Botany 93(9):** 603-610. DOI:10.1139/cjb-2014-0255.

[5] Pei N, Chen B, Kress WJ (2017) Advances of community level plant DNA barcoding in China. **Frontiers in plant science 8(e84291):** 225. DOI:10.3389/fpls.2017.00225.

[6] Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. **Proceedings of the Royal Society B 270(1512):** 313-321. doi: 10.1098/rspb.2002.2218.

[7] Shafqat M, Qasim F, Nasrullah U, Ahmad M, Hussain M, Ali H, Noreen S, Qasim MW (2020) DNA Barcoding in Plants and Current Molecular Issues. **Scholars International Journal of Biochemistry 3(3):** 53-58. DOI: 10.36348/sijb.2020.v03i03.003.

[8] Goldstein PZ, DeSalle R (2019) Review and interpretation of trends in DNA barcoding. **Frontiers in Ecology and Evolution 7(302):** 1-11.

[9] Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S (2015) Plant DNA barcoding: from gene to genome. **Biological reviews 90(1):** 157-166. DOI: 10.1111/brv.12104.

[10] Andújar C, Arribas P, Gray C, Bruce C, Woodward G, Yu DW, Vogler AP (2018) Metabarcoding of freshwater invertebrates to detect the effects of a pesticide spill. **Molecular Ecology 27(1):** 146-166. https://doi.org/10.1111/mec.14410.

[11] Fazekas AJ, Kesanakurti PR, Burgess KS, Percy DM, Graham SW, Barrett SCH, Newmaster SG, Hajibabaei M, Husband BC (2009) Are plant species inherently harder to discriminate than animal species using DNA barcoding markers?. **Molecular Ecology Resources 9(1):** 130-139. DOI: 10.1111/j.1755-0998.2009.02652.x.

[12] Shneyer VS, Rodionov AV (2019) Plant DNA Barcodes. **Biology Bulletin Reviews 9(4):** 295-300. DOI:10.1134/S207908641904008X.

[13] Devey DS, Chase MW, Clarkson JJ (2009) A stuttering starts to plant DNA barcoding. Microsatellites present a previously overlooked problem in non-coding plastid regions. **Taxon 58(1):** 7-15. DOI:10.1002/TAX.581003.

[14] Chen S, Yao H, Han J, Liu C, Song J, Shi L, Zhu Y, Ma X, Gao T, Pang X, Luo K, Li Y, Li X, Jia X, Lin Y, Leon C (2010) Validation of the *ITS2* Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. **PLoS ONE 5(1):** e8613. https://doi.org/10.1371/journal.pone.0008613.

[15] Gonzalez IL, Chambers C, Gorski JL, Stambolian D, Schmickel RD, Sylvester JE (1990) Sequence and structure correlation of human ribosomal transcribed spacers. **Journal of molecular biology 212(1):** 27-35. DOI: 10.1016/0022-2836(90)90302-3.

[16] Ceriaco LM, Gutiérrez EE, Dubois A (2016) Photography-based taxonomy is inadequate, unnecessary, and potentially harmful for biological sciences. **Zootaxa 4196(3):** 435-445. DOI: 10.11646/zootaxa.4196.3.9.

[17] Liu J, Yan HF, Ge XJ (2016) The use of DNA barcoding on recently diverged species in the genus *Gentiana* (Gentianaceae) in China. **PLoS One 11(4):** e0153008. https://doi.org/10.1371/journal.pone.0153008.

[18] Vivas CV, Moraes RCS, Alves-Araujo A, Alves M, Mariano-Neto E, van den Berg C, Gaiotto FA (2014) DNA barcoding in Atlantic Forest plants: what is the best marker for Sapotaceae species identification?. **Genetics and Molecular Biology 37(4):** 662–670. https://doi.org/10.1590/S1415-47572014005000019.

[19] Wang X-Y, Zheng S-H, Liu Y, Han J-P (2016) ITS2, a better DNA barcode than ITS in identification of species in *Artemisia* L. **Chinese Herbal Medicines 8(4):** 352–358. https://doi.org/10.1016/S1674-6384(16)60062-X.

[20] Liu Z-F, Ci X-Q, Li L, Li H-W, Conran JG, Li J (2017) DNA barcoding evaluation and implications for phylogenetic relationships in Lauraceae from China. **PLoS One 124(4):** e0175788. https://doi.org/10.1371/journal.pone.0175788.

[21] Mishra P, Kumar A, Rodrigues V, Shukla AK, Sundaresan V (2016) Feasibility of nuclear ribosomal region *ITS1* over *ITS2* in barcoding taxonomically challenging genera of subtribe Cassiinae (Fabaceae). **PeerJ 4(4):** e2638. https://doi.org/10.7717/peerj.2638.

[22] Poczai P, Hyvönen J (2010) Nuclear ribosomal spacer regions in plant phylogenetics: problems and prospects. **Molecular Biology Reports 37(4):**1897–1912. doi: 10.1007/s11033-009-9630-3.

[23] Sugiura M (1992) The chloroplast genome. **Plant Molecular Biology 19(1):** 149-168. doi: 10.1007/BF00015612.

[24] Lopez-Juez E, Pyke K (2005) Plastids unleashed: Their development and their integration in plant development. **The International Journal of Developmental Biology 49(5-6):** 557–577. DOI: 10.1387/ijdb.051997el.

[25] Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK (2006) The complete chloroplast genome sequence of *Pelargonium* x *Hortorum*. Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. **Molecular Biology and Evolution 23(11):** 2175-2190. https://doi.org/10.1093/molbev/msl089.

[26] Creedy TJ, Norman H, Tang CQ, Qing Chin K, Andujar C, Arribas P, O'Connor R, Carvell C, Notton DG, Vogler AP (2020) A validated workflow for rapid taxonomic assignment and monitoring of a national fauna of bees (Apiformes) using high throughput DNA barcoding. **Molecular ecology resources 20(1):** 40-53. doi: 10.1111/1755-0998.13056.

[27] Kim HM, Oh S-H, Bhandari GS, Kim C-S, Park C-W (2014) DNA barcoding of Orchidaceae in Korea. **Molecular Ecology Resources 14(3):** 499-507. https://doi.org/10.1111/1755-0998.12207.

[28] Kress WJ, García-Robledo C, Uriarte M, Erickson DL (2015) DNA barcodes for ecology, evolution, and conservation. **Trends in ecology & evolution 30(1):** 25-35. DOI:10.1016/j.tree.2014.10.008.

[29] Song Y, Wang S, Ding Y, Xu J, Li MF, Zhu Sh, Chen N (2017) Chloroplast genomic resource of Paris for species discrimination. **Scientific Reports 7:** 3427. https://doi.org/10.1038/s41598-017-02083-7.

[30] Dobrogojski J, Adamiec M, Luciński R (2020) The chloroplast genome: a review. **Acta Physiologiae Plantarum** 42-98. https://doi.org/10.1007/s11738-020-03089-x.

[31] Johnson LA, Soltis DE (1994) *matK* DNA sequences and phylogenetic reconstruction in Saxifragaceae s. str . **Systematic Botany 19(1):**143-156. https://doi.org/10.2307/2419718.

[32] Selvaraj D, Sarma RK, Sathishkumar R (2008) Phylogenetic analysis of chloroplast *matK* gene from Zingiberaceae for plant DNA barcoding. **Bioinformation 3(1):** 24-27. doi: 10.6026/97320630003024.

[33] Lahaye R, van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough TG, Savolainen V (2008) DNA barcoding the floras of biodiversity hotspots. **Proceedings of the National Academy of Sciences 105(8):** 2923-2928.

[34] Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, Percy DM, Hajibabaei M, Barrett SCH (2008) Multiple Multilocus DNA Barcodes from the Plastid Genome Discriminate Plant Species Equally Well. **PLoS ONE 3(7):** e2802. DOI: 10.1371/journal.pone.0002802.

[35] 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 of United States of America 121(23):** 8369-8374. https://doi.org/10.1073/pnas.0503123102.

[36] Galmés J, flexas J, keys AJ, cifre J, Mitchell RAC, Madgwick PJ, Haslam RP, Medrano H, Parry MAJ (2005) Rubisco specificity factor tends to be larger in plant species from drier habitats and in species with persistent leaves. **Plant, Cell and Environment 28(5):** 571-579. https://doi.org/10.1111/j.1365-3040.2005.01300.x.

[37] Feller U, Anders I, Mae T (2008) Rubiscolytics: fate of Rubisco after its enzymatic function in a cell is terminated. **Journal of Experimental Botany 59 (7):** 1615-24. doi:10.1093/jxb/erm242. PMID 17975207.

[38] Bell NE, Quandt D, ÓBrien TJ, Newton AE (2007) Taxonomy and phylogeny in the earliest diverging pleurocarps. square holes and bifurcating pegs. **The Bryologist 110(3):** 533-560. https://doi.org/10.1639/0007-2745(2007)110[533:TAPITE]2.0.CO;2.

[39] Fouad AS, Hamed AB, Amer WM, Hafez RM (2022) Barcoding of Some Plant Species Using the *rbcL* Gene in the Mediterranean Oolitic Sand Dunes West of Alexandria, Egypt. Egyptian Journal of Botany **62(1):** 159-168. DOI: 10.21608/ejbo.2021.74134.1683.

[40] Štorchová H, Olson MS (2007) The architecture of the chloroplast psbA-trnH non-coding region in angiosperms. **Plant Systematics and Evolution 268(1-4):** 235-256. DOI 10.1007/s00606-007-0582-6.

[41] Erickson DL, Spouge J, Resch A, Weigt LA, Kress JW (2008) DNA barcoding in land plants: developing standards to quantify and maximize success. **Taxon 57(4):** 1304-1316. DOI:10.1002/tax.574020.

[42] Huan HV, Trang HM, Toan NV (2018) Identification of DNA barcode sequence and genetic relationship among some species of magnolia family. **Asian Journal of Plant Sciences 17(1):** 56–64. DOI:10.3923/ajps.2018.56.64.

[43] Burgess KS, Fazekas AJ, Kesanakurti PR, Graham SW, Husband BC, Newmaster SG, Percy DM, Hajibabaei M, Barrett SCH (2011) Discriminating plant species in a local temperate flora using the *rbcL* + *matK* DNA barcode. **Methods in Ecology and Evolution 2(4):** 333-340. https://doi.org/10.1111/j.2041-210X.2011.00092.x.

[44] Shaw J, Lickey E, Schilling E, Small R (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. **American Journal of Botany 94(3):** 275-288. ttps://www.researchgate.net/publication/51186853.

[45] CBOL (2009) Plant Working Group, A DNA barcoding for Land plants. **Proceedings of the National Academy of Sciences of the United States of America 106(31):** 12794-12797. https://doi.org/10.1073/pnas.0905845106.

[46] Li D-Z, Gao L-M, Li H-T, Wang H, Ge X-J, Liu J-Q, Chen Z-D, Zhou S-L, Chen S-L, Yang J-B, Fu C-X, Zeng C-X, Yan H-F, Zhu Y-J, Sun Y-S, Chen S-Y, Zhao L, Wang K, Yang T, Duan G-W (2011) Comparative analysis of a large dataset indicates that internal transcribed spacer (*ITS*) should be incorporated into the core barcode for seed plants. **Proceedings of the National Academy of Sciences of the United States of**

America 108(49): 19641-19646. https://doi.org/10.1073/pnas.1104551108.

[47] Abouseada HH, Mohamed A-S H, Teleb SS, Badr A, Tantawy ME, Ibrahim SD, Ellmouni FY, Ibrahim M (2023) Genetic diversity analysis in wheat cultivars using SCoT and ISSR markers, chloroplast DNA barcoding and grain SEM. **BMC Plant Biology** 23:193. https://doi.org/10.1186/s12870-023-04196-w.

[48] De Mattia F, Bruni I, Galimbert A, Cattaneo F, Casiraghi M, Labra M (2011) A comparative study of different DNA barcoding markers for the identification of some members of Lamiaceae. **Food Research International 44(3):** 693-702. https://doi.org/10.1016/j.foodres.2010.12.032.

[49] Feliner GN, Rossello JA (2007) Better the devil you know? Guidelines for insightful utilization of nrDNA *ITS* in species-level evolutionary studies in plants. **Molecular Phylogenetics and Evolution 44(2):** 911-919. doi: 10.1016/j.ympev.2007.01.013.

[50] Gao T, Yao H, Song J, Zhu Y, Liu Ch, Chen Sh (2010) Evaluating the feasibility of using candidate DNA barcodes in discriminating species of the large Asteraceae family, **BMC Ecology and Evolution 10:** 324. https://doi.org/10.1186/1471-2148-10-324.

[51] Shneyer VS (2009) DNA barcoding is a new approach in comparative genomics of plants. **Russian Journal of Genetics 45(11):** 1267-1278. DOI:10.1134/S1022795409110027.

[52] Whitfeld TJS, Kress WJ, Erickson DL, Weiblen GD (2012) Change in community phylogenetic structure during tropical forest succession: Evidence from New Guinea. **Ecography 35(9):** 821-830. doi: 10.1111/j.1600-0587.2011.07181.x.

[53] Pillon Y, Johansen J, Sakishima T, Chamala S, Barbazuk WB, Roalson EH, Price DK, Stacy EA (2013) Potential use of low-copy nuclear genes in DNA barcoding: a comparison with plastid genes in two Hawaiian plant radiations. **BMC Evolutionary Biology 13(1):** 35. https://doi.org/10.1186/1471-2148-13-35.

[54] Kranitz M, Biffin E, Clark A, Hollingsworth ML, Ruhsam M, Gardner MF, Thomas P, Mill RR, Ennos RA, Gaudeul M, Lowe AJ, Hollingsworth PM (2014) Evolutionary diversification of New Caledonian *Araucaria.* **PLoS One 9(10):** e110308. https://doi.org/10.1371/journal.pone.0110308.

[55] Ruhsam M, Rai HS, Mathews S, Ross TG, Graham SW, Raubeson LA, Mei W, Thomas PI, Gardner MF, Ennos RA, Hollingsworth PM (2015) Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in *Araucaria*?. **Molecular Ecology Resources 15(5):** 1067–1078. https://doi.org/10.1111/1755-0998.12375.

[56] Ivanova NV, Kuzmina ML, Braukmann TWA, Borisenko AV, Zakharov EV (2016) Authentication of herbal supplements using next generation sequencing. **PLoS One 11(5):** e0156426. doi: 10.1371/journal.pone.0156426.

[57] Dong W, Liu H, Xu C, Zuo Y, Chen Z, Zhou Sh (2014) A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: a case study on ginsengs. **BMC Genetics 15(1):** 138. doi: 10.1186/s12863-014-0138-z.

[58] Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A (2012) Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. **American Journal of Botany 99(2):** 349-364. doi: 10.3732/ajb.1100335.

[59] Bakker FT, Lei D, Yu J, Mohammadin S, Wei Z, DE Kerke SV, Gravendeel B, Nieuwenhuis M, Staats M, Alquezar-Planas DE, Holmer R (2016) Herbarium genomics: Plastome sequence assembly from a range of herbarium specimens using an iterative organelle genome assembly pipeline. **Biological Journal of the Linnean Society 117(1):** 33-43. https://doi.org/10.1111/bij.12642.

[60] Hollingsworth PM, Li D-Z, van der Bank M, Twyford AD (2016) Telling plant species apart with DNA: from barcodes to genomes. **Philosophical Transactions of The Royal Society B Biological Sciences 371(1702):** 20150338. DOI:10.1098/rstb.2015.0338.

[61] Coissac E, Hollingsworth PM, Lavergne S, Taberlet P (2016) From barcodes to genomes: extending the concept of DNA barcoding. **Molecular Ecology 25(7):** 1423-1428. https://doi.org/10.1111/mec.13549.

[62] Chomicki G, Renner SS (2015) Watermelon origin solved with molecular phylogenetics

including Linnaean material: another example of museomics. **New Phytologist 205(2):** 526-532. https://doi.org/10.1111/nph.13163.

[63] Kane N, Sveinsson S, Dempewolf H, Yang JY, Zhang D, Engels JMM, Cronk Q (2012) Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. **American Journal of Botany 99(2):** 320-329. https://doi.org/10.3732/ajb.1100570.

[64] Hollingsworth PM, Graham SW, Little DP (2011) Choosing and Using a Plant DNA Barcode. **PLoS ONE 6(5):** e19254. doi: 10.1371/journal.pone.0019254.

[65] Kaur S (2015) DNA Barcoding and Its Applications. **International Journal of Engineering Research and General Science 3(2):** 602-604.

[66] Kress WJ, Lopez IC, Erickson DL (2012) Generating plant DNA barcodes for trees in long-term forest dynamics plots. In: Kress W.J., Erickson D.L. (eds). DNA barcodes. Methods in Molecular Biology (Methods and Protocols), vol 858. Humana Press, Totowa, NJ. https://doi.org/10.1007/978-1-61779-591-6_22.

[67] Golovnina KA, Glushkov SA, Blinov AG, Mayorov VI, Adkison LR, Goncharov NP (2007) Molecular phylogeny of the genus *Triticum* L. **Plant Systematics and Evolution 264(3):** 195–216. doi:10.1007/s00606-006-0478-x.

[68] Osman SA, Ramadan WA (2019) DNA Barcoding of different *Triticum* species. **Bulletin of National Research Centre 43:** 174. https://doi.org/10.1186/s42269-019-0192-9.

[69] Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. **Molecular Biology and Evolution 30(12):** 2725-2729. https://doi.org/10.1093/molbev/mst197.