Current Trends in Pharmaceutical Research, 2022Vol 9 Issue 1
© Dibrugarh University www.dibru.ac.in/ctprISSN: 2319-4820 (Print)
2582-4783 (Online)

Review article

A CRITICAL REVIEW ON THE CHALLENGES AND ADVANCES IN DNA BARCODING FOR PLANT IDENTIFICATION

Farha Ahmed, Md. Kamaruz Zaman*

Department of Pharmaceutical Sciences, Faculty of Science and Engineering, Dibrugarh University, Dibrugarh 786004, Assam, India

Abstract

Background: *DNA barcoding is a novel and extremely prevalent approach of molecular categorization and identification of species using short genomic sequences. It supplementally solves current difficulties of classical taxonomy and phylogenetics allowingquick identification, effective taxonomic discrimination,and validated categorization. This ground breaking technology pioneered by Hebert et al. is being applied to plants for a wide range of applications.*
Objective: *The current literature critically reviews on plant barcoding its challenges, recent advancements, role of bioinformatics and novel computational methodologies and software tools involved.*
Methods: *A thorough search for manuscripts was conducted using a variety of platforms such as Google Scholar, PubMed ResearchGate, Science Direct, NCBI, SpringerLink.*
Results and Discussion: *Although DNA barcoding is swiftly gaining prominence as a game changer, it poses certain challenges when applied to plants. Because of the sluggish rate of genetic evolution in plant mitochondrial genomes, there is minimal variation between COI in closely related plant species. Plant hybridization and polyploidy are projected to have an impact on species identification by DNA barcoding. Plant material rich in polysaccharides, polyphenols, and other secondary metabolites increases DNA destruction, as shown in museum specimens. Amplification and sequencing of DNA might be difficult with such degraded samples.*
Conclusion: *In a large-scale effort to address these difficulties with conventional plant barcoding, advances in possible markers, sequencing, and computational technologies will reshape DNA barcoding in the future, making it an extensively utilised and useful tool.*

Keywords: Plants,DNA barcoding, Super barcoding, Ultra-barcoding, MEBarcoding, Next-generation sequencing, Bioinformatic

*Corresponding author's E-mail: kzaman71@dibru.ac.in

Introduction

Plants are an indispensable component of life on Earth, supplying mankind with breathing air, food, fuel, medicine, and other benefits. A thorough study of plants is required to increase agricultural production and sustainability, find novel medications, plan for and minimise the worst consequences of climate change, and get a better understanding of life in general. Many ecosystems are under peril as a result of a growing human population and a changing climate. As a result, identifying new or uncommon species and measuring their geographical distribution as part of larger biodiversity studies is becoming increasingly crucial[1,2].Conventional ways of plant identification include organoleptic methods (identification by the senses: taste, sight, smell, and touch), macroscopic and microscopic approaches (identification by structure, colour, and texture), and chemical profiling (e.g., TLC, HPLC-UV, HPLC-MS)[3].However, neither approach is easily capable of identifying species because the former requires skilled taxonomist for macroscopic and microscopic inspections. Such qualified subject matter specialists are in scarce (a concept called as the "taxonomic impediment"). Chemical profiles and indicators in the latter technique may be modified by physiological and storage factors[1,3]. The merging of alpha taxonomy with the extremely prevalent and unique notion of DNA barcoding, which uses a short genomic sequence to identify species, has accelerated the pace of alpha taxonomy. DNA barcoding solves current difficulties of classical taxonomy and phylogenetics as a supplemental but not full alternative to systematics study. Hebert et al. (2003) pioneered this ground-breaking technology by using the mitochondrial cytochrome c oxidase gene (COI) as a universal animal barcode to distinguish lepidopteran species.This recent marker-based strategy in molecular systematics research intends to provide a shared community resource of DNA sequences for quick identification, effective taxonomic discrimination, and validated categorization based on genetic data[4,5].Plant DNA barcoding is already being used in a wide range of applications. For example, barcoding techniques have been used toauthenticate a wide range of plant commodities, including medicinal herbs, household spices, berries, olive oil, tea,black pepper powder and commercial timber[6–8].These technologies have now been used to assess plant-pollinator interactions ,allergen monitoring [9] and in establishment of a novel tool for identifying dangerousplants by poisonous centres in the event of accidental ingestion[10].

Conventional DNA barcoding

The essential principles of DNA barcoding are standardisation, minimalism, and scalability. While COI fits these criteria for animal, thelow pace of genetic

evolution in plant mitochondrial genomes implies there is little variation between COI in closely related plant species, ruling out COI as a barcode identifier for them[11].

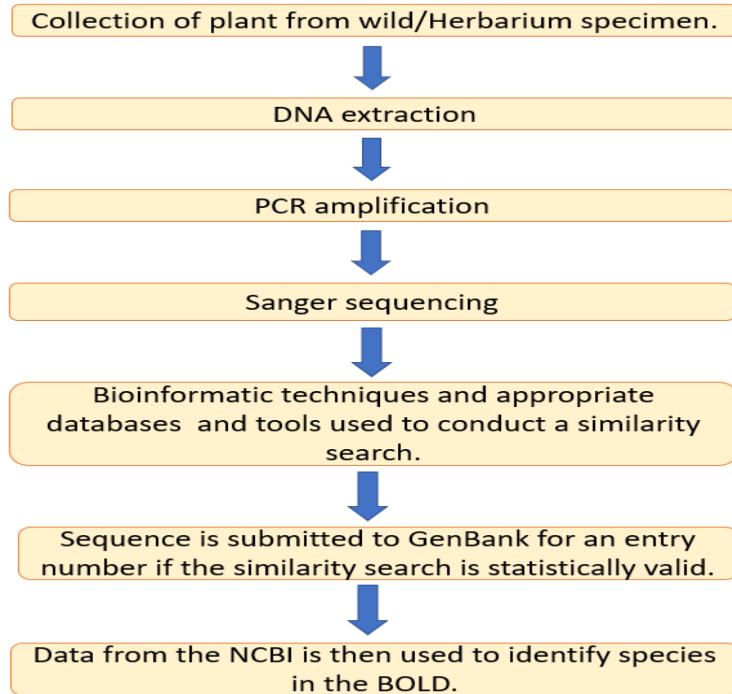

Figure 1: Schematic representation of steps involved in DNA barcoding.

Following the review of various potential markers, the matK and rbcL markers were declared as the core sequences of plant DNA barcodes in 2009 at the 3rd World DNA Barcode Conference, with ITS and trnH-psbA as complimentary sequences.Chen and colleagues recommended the ITS2 region as the DNA barcode and trnH-psbA as a complementary sequence for the identification of medicinal plant species after extensive testing and verification[12–14]. Gene deletion is a key limiting factor for single loci, inhibiting its usage as a universal DNA barcode, in addition to insufficient diversity and low PCR efficiency (typically owing to sequence variation in the primer binding areas).Multi-locus markers had been thought to be more effective in identifying species, however investigations have shown that they are also insufficient for universal plant identification[15]. MatK, rbcL, trnH-psbA, and ITS sequence are the most prevalent DNA marker combinations[12,16–18]. With a discriminating effectiveness of just 72 percent, the Consortium for the Barcode of Life-Plant Working Group (CBOL) has endorsed the two-locus combination of matK + rbcL as the best plant barcode. When compared to single-locus markers, combined barcodes enhance analytical problems, especially

when one of the target loci does not amplify [12,15].When compared to obtaining new samples from the wild, extracting DNA from herbarium specimens can be more difficult. The suitability of samples for DNA extraction in different herbaria differs depending on the storage conditions and how specimens were initially conserved. Newer material performs better than older material, and certain taxonomic groupings do worse than others. We advocate a mixed strategy for huge DNA barcode initiatives, which includes using herbarium specimens to immediately bulk up the number of samples available, and then filling in the gaps with additional collections of material for species that function less well from the herbarium[6,19,20]. DNA is extracted from collected samples using one or more of the procedures available, such as CTAB, SDS, PVP, Phenol-Chloroform, and so on[4,6].PCR amplification and Sanger sequencing of DNA barcode sequences from genomic DNA isolated from individual specimens are the traditional methods for obtaining DNA sequence data to produce a barcode for a species or specimen. This is for well-identified specimens for library creation, but the specimen does not need to be documented for species determination. For over three decades, Sanger sequencing technology, which can generate sequencing reads of up to 1000 bases, was the sole method utilised for DNA sequencing. Apart from the poor throughput, Sanger sequencing need a high-concentration DNA amplicon template (100–500 ng) to avoid biases and mistakes. Sanger sequencing also generates a separate electropherogram, or sequencing signal pattern, for each sequence created[21–23].After obtaining sequenced DNA data for the samples, bioinformatic techniques and appropriate databases and tools such as NCBI, BLAST, and others are used to conduct a similarity search. For both producing a new barcode and reporting a new record for an existing barcode, proper data analysis is regarded vital at that point. The sequence is submitted to GenBank for an entry number if the similarity search is statistically valid. Data from the NCBI is then used to identify species in the Barcode of Life Database[4].

Challenges

Apart from addressing taxonomic difficulties, DNA barcoding can answer issues about phylogenetic connections between species and phylogeographic diversity within species. These uses have occasionally been confused with DNA barcoding, producing doubt about the extent of DNA barcoding. There are three types of limitations: • Conceptual limitation: – Although every barcode area contains some phylogenetic information, DNA barcoding is not intended to reconstruct phylogenetic connections. • Genetic constraints: DNA barcodes may not give adequate precision to differentiate recently evolved species. Because the plant DNA barcode is made up of two plastid regions, hybridization and polyploidy have a predicted influence on species identification, which is possibly stronger in groups with high levels of apomixis.Polyploidy and mitochondrial or plastid introgression,

including ongoing or previous hybridization episodes, are not resolved by most DNA barcode markers. Heteroplasmy (the occurrence of two or more variations of the barcode area within an organism) might make reliable sequence recovery difficult.Organellar DNA sequences that are not functional frequently appear in the nuclear genome. If mistaken for the correct barcode sequence, these nuclear-mitochondrial sequences (NUMTs) and nuclear-plastid sequences (NUPTs) might cause data interpretation problems. •Methodological limitations: - In certain species, 'universal' primers for DNA barcoding may fail to amplify the target region[11,24]. When DNA is degraded, as with old museum specimens or tissue samples exposed to agents that degrade DNA (e.g., high temperatures), or when the plant material contains high amounts of polysaccharides, polyphenols, and other secondary metabolites, such as alkaloids and flavonoids, success in recovering pure, high molecular weight DNA is reduced. Shorter fragments are simpler to amplify from herbarium DNA, according to Sarkinen et al., who discovered a substantial negative association between amplicon size and PCR success. The relatively big genomic region rbcL (670 bp) was amplified in only 10% of the samples, the medium sized LEAFY intron (260 bp) was amplified in 24%, and the tiny was amplified in 78 percent of the samples. It is thus advised to design circumstances that yield extremely short amplicons, which are simpler to amplify, or to execute DNA ‘repair reactions' for badly damaged DNA[3,20,25].

RECENT BREAKTHROUGHS IN PLANT BARCODING

Species identification technologies are constantly improving. High speed and automation are now required. However, we cannot deny that approaches like barcoding are still in their infancy, and that much more progress still to be made. As a result, the DNA barcoding approach has received a lot of flak, but there are strategies for improving it. Furthermore, the DNA barcoding technology has a bright future ahead of it[26].Advances in potential markers, sequencing, and computational technologies are revolutionising DNA barcoding in a large-scale effort to tackle problems with classical barcoding[27].As a result, the innovative approaches discussed below can be used with current methodologies to provide a better and more noticeable barcoding result in plants**:**

Super barcoding

Species discrimination using whole-plastid-based barcodes has shown tremendous promise, especially for closely related taxa. The chloroplast genome comprises all of a plastid's DNA sequences, containing more genetic information for species identification than any single-locus marker now in use. The chloroplast genomes of 3452 plants were published on NCBI on October 27, 2019[28,29]. It's a significant step forward in the creation of DNA barcodes that use chloroplast genomes to identify and differentiate plants. It was proposed as a species-level DNA barcode because it can greatly increase resolution at lower taxonomic levels in plant

phylogenetic, phylogeographic, and population genetic analyses, facilitating the recovery of monophyletic lineages. The use of the cp-genome as a marker avoids issues like gene deletion and low PCR efficiency. The examination of this super-barcode also eliminates the sequence retrieval challenges that are common in regular barcoding investigations. The cp-genome is smaller than the nuclear genome and has higher interspecific and lower intraspecific divergence, making it more appropriate as a genome-based barcode[15,30–33]. Zhang et al. exploited highly variable sections of CP genomes, such as those found in *Dracaena* species, to accurately identify *Dracaena cochinchinensis,* the only source plant of rare traditional medicine dragon's blood, according to the Chinese Pharmacopoeia. The therapeutic effectiveness of dragon's blood will certainly be harmed by incorrect identification of *Dracaena* species[34–38].Although super-barcode has numerous advantages, it is ineffective for identifying plant species when DNA extraction is problematic[39].

Ultra-barcoding

"Ultra-barcoding" refers to the expansion of conventional DNA barcodes to complete plastomes and nrDNA sequences[40]. This method generates a massive quantity of data for each locus, making it significantly more sensitive than standard DNA barcoding and potentially providing the information needed to investigate variation below the species level. In plants, the plastid is the major target of this UBC method, although nuclear sequence is also produced and gives extra important information. Because ribosomal DNA (rDNA), like chloroplast DNA, is multicopy,it requires significantly less coverage than single-copy nuclear genes.Furthermore, while being multicopy, rDNA does not suffer from the same paralogy issues as other multicopy nuclear loci like transposons since the copies do not develop independently, owing to biased gene conversion and uneven crossing over. Because the paralogous copies evolve in lockstep, this trait is known as concerted evolution, and it implies that rDNA provides significant phylogenetic and even population-level signals[41–44].

The identification of a cryptic species in *P. yunnanensis*, a medicinally significant plant, was made possible by Ji et al. using this ultra-barcoding analysis. *P. yunnanensis* is presently divided into two genetic lineages, which correlate to the two phenotypes ("typical" and "high stem" form). The "high stem" variety, with its unusual morphologies and range, should be recognised as a previously unidentified species; it is now designated as *P. liiana sp.* nov.[40, 45]

Cost and the somewhat higher quality and quantity of DNA required, as well as the bioinformatics and computing resources necessary to deal with vast volumes of next-generation sequence data, are the key roadblocks for UBC. Nonetheless, we expect that, as technology and methodology advance, it will soon be possible to sequence and assemble complete plastid genomes for a wide range of uses[41].

Bar HRM Technology

The Bar-HRM approach combines high-resolution melting analysis with DNA barcoding. The following is how it works: after conducting the PCR, a 50–500 bp amplicon is gradually denatured by minor temperature rises, generating variations in fluorescence levels induced by the release of a fluorescent dye that is intercalated in double-stranded DNA. The reduction in fluorescence may be plotted against temperature rises to construct a melting curve by denaturing the DNA double strand. Several factors influence the melting curve, including the complementarity of the DNA double strand, the length of the amplicon, and the sequence and quantity of GCs.The Bar-HRM approach is quick and inexpensive, and it does not require the use of probes or sequencing, allowing for faster species identification. However, because it is a highly sensitive technology, its resolution power is dependent on the use of high-quality DNA[12,46–50]. This method, in combination with plant DNA barcoding, has been used to successfully authenticate ginseng food products (*Panax ginseng* C.A. Mey. and *Panax quinquefolius*L.) [50,51], identify olive oil [50,52–54]and wine [54], authenticate herbal medicines and accurately quantify adulterants in commercial herbal medicine products[55], and distinguish edible vegetables from poisonous plants for food safety[56].

DNA mini-Barcoding

Traditional barcoding methods are sometimes hindered by their inability to amplify and sequence degraded DNA, which is common in museum specimens and conserved and processed biological material (food items, rotting tissues). Because the DNA damage and degradation in museum collections is complicated and difficult to describe, approaches to repair DNA in vitro are wasteful and ineffective. Short sequences (less than 100 bp) are more stable in museum specimens. The adoption of a short or minimalist barcode considerably widens the applications of DNA barcoding[57–59].Mini-barcodes employ a shorter length of DNA, generally less than or equal to 200bp, that may be amplified faster than standard barcodes and have been demonstrated to be efficient for species-level identification in DNA-damaged specimens and in cases when obtaining a full-length barcode is problematic. The PCR success rate for DNA mini barcodes is greater since the amplicon length is shorter[4,20,59]. A short barcode sequence may also reliably predict components like average nucleotide composition, strand asymmetry patterns, and a high frequency of hydrophobic amino acid encoding codons. Furthermore, when compared to full-length barcodes, mini-barcodes have been demonstrated to offer measurements of sequence diversity and divergence at both the intra-specific and intra-generic levels in some circumstances[57,60,61].The design of DNA mini-barcodes must be screened and obtained, and the information must be gathered in databases such as GenBank, the European Molecular Biology Laboratory, or the DNA Data Bank from Japan for the development of DNA mini-

barcoding. The sequences must then be aligned in order to discover conserved and particular areas. Finally, to improve the identification power, the sequencing procedure for these must be bidirectional and the sequences must be joined in contig consensus sequences. However, because various length sequences are picked as mini-barcodes in different. The effectiveness of this approach is highly dependent on the use of species-specific primers, as well as their length and design, to avoid dimer and hairpin formation[46,62,63]. This approach has been applied to authenticate herbal supplements of gingko biloba using a 166-bp matK mini-barcode[64] and garlic (*Allium sativum* L.) supplements using a *trnL*^{UAA} mini-barcode[65].

Next generation sequencing approach

The automated Sanger method is considered as a 'first-generation' technology, and newer methods are referred to as next-generation sequencing (NGS)[66,67].Advances in next-generation sequencing technologies have transformed biological science since 2005. These next-generation sequencing (NGS) technologies have the potential to generate hundreds of thousands to tens of millions of sequencings reads in parallel. This high-throughput sequencing capability can generate sequence reads from fragmented genome libraries, a pool of cDNA library fragments derived through reverse transcription of RNA molecules, or a pool of PCR amplified molecules. Sequences are generated in all cases without the use of a vector-based cloning procedure, which is typically used to amplify and separate DNA templates[68,69].

For DNA barcoding initiatives, NGS technology might be used to overcome some of the short comings of Sanger-based sequencing. By utilising the improved throughput given by NGS technology, this innovative technique might possibly permit the creation of DNA barcodes more rapidly and at a lower total cost[21]. Currently, Roche Diagnostics' Pyrosequencing(previously454 Sequencing) and Illuminasequencing (formerly Solexa) are the two most widely utilised NGS methodologies[70]**.**

Pyrosequencing is a DNA sequencing method based on the notion of sequencing by synthesis. The method is based on a four-enzyme real-time bioluminescence monitoring of DNA synthesis employing a cascade that terminates in a measurable light signal after nucleotide incorporation (bioluminescence). The detection mechanism is based on the pyrophosphate generated when a nucleotide is inserted into the DNA-strand[71]**.**Pyrosequencingbegins with the DNA template being broken down into several short sequences (35–500 bp). The fragments are then ligated to adapters, which allow the tiny DNA fragments to connect to complementary adaptor strands attached to beads (one DNA fragment per bead). For clonal amplification, the bead-bound segments are subsequently exposed to emulsion PCR. Beads carrying amplified clonal DNA are then put in a

microfabricated array of wells at random, with just one bead in each well. Pyrosequencing is then applied to each bead at the same time[72–74]. The dNTPs (A, T, C, and G) are successively added to the polymerase enzyme during the sequencing procedure. PPi is released and converted to ATP utilising ATP sulfurylase and adenosine 5' phosphosulfate when a dNTP is effectively integrated into a nascent DNA strand in a microwell. The ATP is then utilised to drive luciferin to oxyluciferin conversion utilising luciferase, which releases a quantity of light proportionate to the amount of PPi generated. The chemiluminescent light signals emerging from each well are recorded using an imaging equipment.Apyrase degrades the leftover unincorporated dNTPs, ensuring that a reaction happens only when fresh dNTPs are added. The process is continued until the template is created. The intensity profiles of each bead have been created after the DNA strand has been replicated. This approach currently produces sequence reads with an average lifespan of 400 bp[70,72].

Illuminasequencing is a parallelized version of classical Sanger sequencing. To amplify clonal sequencing characteristics, the Solexa method uses bridge PCR (also known as "cluster PCR"). In a nutshell, an adaptor-flanked shotgun library is PCR amplified in vitro, but both forward and backward primers thickly coat the surface of a solid substrate, with a flexible linker linked at their 5′ ends. As a result, amplification products generated from any member of the template library remain tethered locally near their source. Each clonal cluster comprises 1,000 copies of a single template library member at the end of the PCR. Within each of the eight separate 'lanes' on a single flow-cell, many million clusters can be amplified to identifiable places [73,75,76].Because the DNA strands are clonally amplified within a relatively limited region, bridge amplification is utilised. Adding the dNTPs (all four at the same time) to the flowcell in cycles performs the sequencing reaction. Each dNTP has its own reversibly bonded fluorescent label and functions as a polymerase reaction terminator. During each polymerase reaction cycle, just a single dNTP is therefore incorporated. The coloured fluorescent signal generated in the flowcell after a laser excitation of each DNA "hotspot" is recorded using an imaging device. The integrated nucleotide's reversible terminator is subsequently cleaved, and the process is repeated until all of the template strands have been duplicated. The sequence is determined by the hue of each "hotspot" during each nucleotide incorporation cycle. This method currently produces sequence reads with an average length of roughly 40 bp[70].

High-throughput NGS methods have experienced a variety of problems since their launch in 2005. The first has been an enhancement in sequencing output in terms of read length (For all new platforms, read lengths are now much shorter) and accuracy (Base-calls made by the new platforms are, on average, 10 times less accurate than those obtained by Sanger sequencing). The sequencing experiment's

total output in relation to the cost and labour expended has been the second challenge.This third challenge comprises many kinds of PCR bias; nonetheless, all PCR-based NGS systems suffer from bias introduced during amplification. Many studies have shown that employing high template concentrations, smart primer selection, low cycle number, low annealing temperature, and mixed replicate reaction preparations, PCR bias may be significantly decreased[21,68,73].New technologies have the potential to profoundly alter the character of genomics-based research, especially when combined with the computer algorithms required to analyse the massive amounts of data generated by their huge sequencing output[68].

Microfluidic Enrichment Barcoding (MEBarcoding)

MEBarcoding is a high-throughput DNA barcoding method that employs the Fluidigm Access Array to amplify specified areas for 48 DNA samples and hundreds of PCR primer pairs (generating up to 23,040 PCR products) in a single thermal cycling process. Fluidigm Corporation's Access Array (1st generation) and Juno (2nd Generation) are the most widely used commercial microfluidic equipment for high-throughput sequencing library creation.Integrated fluidic circuits (IFC) are employed in these devices to exploit the chemical and fluid mechanics of chemicals used in PCR on a micromolecular level. By pushing DNA samples and PCR chemicals into tiny quantities (0.03 L) that interact in a customised thermal cycler, this method may alter them.All reagents are placed into a single device called an IFC, which is about the size of a 96-well PCR plate.In miniaturized and parallel PCRs, these equipmentenables for the simultaneous amplification of thousands of amplicons. This technology is a cost-effective way to amplify distinct target areas from several samples since it minimises not only the quantity of reagent needed in HTS library creation and focused amplification of barcode loci, but also instrument and sample handling, as well as technician time. However, this technology has two distinct drawbacks: 1) a high initial equipment cost, and 2) reduced sequencing success as compared to Sanger techniques. If a researcher does not have access to an equipment capable of performing microfluidic PCR, the first and possibly most pressing barrier is the high initial cost[77].

ROLE OF BIOINFORMATICS IN PLANT DNA BARCODING

Bioinformatics is frequently described as the use of computer approaches to comprehend and organise data relating to biological macromolecules. It has firmly established itself as a molecular biology discipline, including a wide variety of topics such as structural biology, genomics, and gene expression research [78]. Margaret Dayhoff (1925–1983), an American physical chemist, was one of the first to use computational approaches to biochemistry [79,80]. In order to take use of developments in next-generation sequencing technology for DNA barcoding projects, a user-friendly and successful bioinformatics pipeline is required. Traditional DNA barcoding is ineffective and time-consuming without the use of

bioinformatics.Bioinformatics tools and associated databases are required to assist at various stages of the barcoding process, including data collection, storage, analysis, visualisation, and correct administration [81]. The ultimate goal of a DNA barcoding bioinformatics process is to (1) create a "clean" or "reliable" digital representation of the DNA barcodes, and (2) utilise these DNA barcodes in conjunction with DNA sequence libraries to get information about the taxonomy of the unknown material [82].

Major Consortia for Plant Barcoding

iBOL (International Barcode of Life) In 2010, the International Barcode of Life initiative was activated for the first time. It was hosted by the University of Guelph's Biodiversity Institute of Ontario. The iBOL consortium's activities were carried out through its component nodes, which include 27 countries and 20 working groups. In five years, iBOL hopes to generate 5 million barcode records from 500,000 species. The primary goal of iBOL is to expand the geographic and taxonomic coverage of the barcode reference library - BOLD - by preserving the generated barcode records, allowing community access to the knowledge they represent, and developing new technologies to assure worldwide access to this data[81,83,84].

CBOL (Consortium for the Barcode of Life) CBOL is a global project dedicated to the advancement of DNA barcoding as a worldwide approach for identifying the plants and wildlife that make up the world's biodiversity. It was founded in May of 2004 and is funded by the Alfred P. Sloan Foundation. The secretariat of the CBOL is housed in the National Museum of American History in Washington, DC. CBOL now has approximately 130 members from 43 countries, all of whom are working to promote DNA barcoding via conferences, seminars, and trainings[81,85].

Computational Methodologies

Traditional PhylogeneticAnalysis: If the standard DNA barcoding approach is used, the following step is sequence analysis and the creation of phylogenetic trees when the matching query sequences have been obtained. Sequence analysis entails aligning query sequences with a reference collection of sequences. The multiple alignment stage is crucial since a faulty alignment would result in an erroneous phylogenetic tree, which will lead to misidentification of taxa, which is the fundamental purpose of DNA barcoding. ClustalW, T-Coffee, and MUSCLE, for example, are MSA programmes.After obtaining MSA, the alignment may be manually adjusted to improve alignment quality using BioEdit, Jalview, or any other alignment editor, and then phylogenetic trees can be constructed [81,86–88].

Trees can be built using either distance-based or character-based approaches:-

- Distance-based approaches rely on differences in DNA sequences between and within species. Though the method is suitable for large datasets and further analysis, as well as lineages with different branch lengths, it can be misleading in accurately assigning a query sequence to a taxon, because

there may be an overlap of inter and intraspecific distances between and within species due to different substitution rates, i.e., different rates of evolution[81,89].Two clustering-based algorithms, UPGMA and NJ, and two optimality-based algorithms, Fitch-Margoliash and minimal evolution, are among the distance-based approaches[90,91].

- Character-based techniques may be used to create classification rules based on an existing hierarchical organisation, and then categorise new information quickly without the need for complex phylogenetic procedures[92].The presence and absence of distinct diagnostic features, the four standard nucleotides (A, T, C, G), are used to classify species in this method.The capacity of the discovered character states to identify the query sequences must also be evaluated to ensure their dependability. Maximum probability and parsimony are two character-based techniques[81,93].

Following the alignment and analysis, the tree is built using either phenetic, cladistic, or tree-building techniques.MEGA version 5and PAUP version 4are some tree-building tools[81,94,95].

Novel computational methodologies

- DNA Metabarcoding

The fundamental purpose of DNA metabarcoding is to identify broad groups of species in a single environmental sample at the same time. This method combines high-throughput sequencing (HTS) with a polymerase reaction based on DNA amplification to provide unbiased amplification from a variety of DNA templates. Metabarcoding uses universal primers to identify numerous plant species at the same time. It has proven feasible to detect the content of medicinal plants in processed herbal products using a metabarcoding technique[46,84].De Boer used nrITS1 and nrITS2 DNA metabarcoding to identify orchid and other plant species present in 55 commercial products [12,96].

- Neutral network

The neural network is a mathematical/computational representation of a biological system. Back Propagation (BP) is a supervised learning approach that may be used to train multi-layer feed-forward neural networks. The approach entails training a network using a reference dataset of sequences, then utilising that trained network to identify query sequences. It also considers non-molecular characteristics, such as morphological characteristics or other behavioral information. Because it is a supervised learning process, this approach always allocates an input sequence to a known species, implying that the method is only helpful for identifying samples that have been preset[81, 97–99].

- Machine learning

This method, when used in conjunction with bioinformatics, is particularly useful for detecting species with non-coding barcodes, and it outperforms their previously published BP neural network approach. By employing graphical representations of DNA sequences through a DV-Curve or FJ-Curve technique, the method overcomes the difficulty of sequence alignment. When forming conclusions, it is less reliant on assumptions. It provides a significant performance advantage over the BP approach when dealing with large datasets since it decreases the data matrix size to a substantial amount[81,100].Using machine learning methodologies, researches were performed to authenticate the wood of eight endangered Dalbergia timber species[101] and quantify adulteration in marketed ayurvedic raw pharmaceuticals (e.g.*, S. alnifolia, T. arjuna*)[102].

- Composition Vector (CV) method

The composition vector (CV) method is a non-alignment methodology that is particularly well suited to non-protein-coding sequences used as barcodes. The tree topologies generated using the CV approach match those found using standard methods. The CV technique consists of four steps:(i) calculating the frequency for every k-string in the sequence, (ii) building a CV for the sequence, (iii) computing the distance between every two CVs to generate a distance matrix, and (iv) building a phylogenetic tree[81,103, 104]. Li et al. used a modified composition vectors (CVs) to perform correlation analysis on a plant dataset[105].

Tools used for plant DNA barcoding

The various software tools applied in DNA barcoding of plants are briefly summarised in table-1.

Table 1: List of software tools involved in plant DNA barcoding technique.

Tools	**Purpose**	**References**
SPIDER	SPeciesIDentity and Evolution in R includes a set of functions to examine and evaluate data at the species level, includes both molecular and morphological levels. Functions: generate significant summary statistics from DNA barcode data; evaluate specimen identification efficacy; testing; adjusting divergence threshold limits. A sliding window tool allows to design markers in degraded DNA.	[84,106]
ecoPrimers	Finds new barcode markers and their PCR primers; searches entire genomes for such markers without prior knowledge. Functions: Selects highly conserved primers	[107,108]

	from a trained set of sequences; evaluate an amplified region's ability to discriminate between taxa; Optimises two quality indices that measure taxonomic range and discrimination to select the most efficient markers from a set of reference sequences based on specific experimental constraints like marker length or taxa.	
MOTUs	"Molecular Operational Taxonomic Units" are sequence clusters formed by grouping the DNA sequences of a conserved gene or gene fragment. Sequence clusters reflect the genomes from which they were generated. At a variety of similarity cutoffs, a dataset of sequences categorised into MOTU, the cutoff value function as a parameter in the clustering process. jMOTU organises DNA barcode sequence data into MOTUs. Taxonerator: a software application inserts taxonomic annotation to the jMOTU-generated clustered barcode sequence data.	[81,109]
TaxI	Determines the sequence divergences between a query sequence (taxon to be barcoded) and each sequence in a collection of reference sequences supplied by the user.Because the assessment focuses on distinct pairwise alignments, this programme can operate with sequences with numerous insertions and deletions, which are difficult to match in huge sequence sets. After analysis of all potential pairs between query and reference files, the software uses the T-Coffee programme to make pairwise alignments**.**	[81,84,87, 89]
CLOTU	Programme for grouping and processing 454 amplicon data and taxonomic annotation; may also be used to analyse other forms of sequence data. Filtering and trimming, clustering, and taxonomic annotation are the 3 phases of the CLOTU method. Used to filter low-quality readings and detection of mismatches in primers and tags; detection option for homopolymers frequently acquired using 454 pyrosequencing	[81,110]
BRONX	"Barcode Recognition Obtained with Nucleotide	[81,111, 112]

	eXpose's" is a technique that does not need alignment. Finds small variable DNA segments and associated invariant flanking areas based on reference sequences. Character properties derived from the variable region estimations are then utilised to identify a query sequence. In comparison to previous algorithms, BRONX provides better and more accurate genus-level identifications.	
CAOS	"Characteristic Attributes Organization System" is computer-assisted approach to identify conserved character states in cladograms (trees) or groupings of categorical data. CAOS defines attribute checks at every node in a phylogenetic tree. Unlike decision tree algorithms, CAOS evaluates qualities that are diagnostically useful. It uses P-Gnome as diagnostic rules generator to searches through a data matrix and create diagnostic rule sets for each of the data matrix's pre-described elements P-Elf software may then categorise a file of query sequences according to criteria provided by P-Gnome.	[81, 92]
OFBG	"Oligonucleotide Frequency Barcode Generator" a novel species discrimination approach based on barcode loci's non-overlapping oligonucleotide frequency range (OFR).Three 3 C++ programmes used: DinuTrinu software creates two tab separated text files with the dinucleotide and trinucleotide frequencies of DNA sequences. These files used to create the lowest and maximum oligonucleotide frequency for each species using MS Excel macro "MinMax". DiBarcode and Tribarcode provide binary matrix of species resolution based on lowest and maximum oligonucleotide frequencies.	[81, 113]

Plant barcoding database

GenBank: GenBank is a large database with freely available nucleotide sequences for around 260 000 officially recognised species[114]. One of the most significant

sources of genetic data is GenBank, which is housed at the NCBI (National Center for Biotechnology Information). Through a variety of specialist databases, it gives DNA, RNA, and protein sequences. Unlike BOLD, GenBank does not keep the chromatograms of sequences given to the site, as well as metadata and picture collections. Visitors have no limits on how they can use or distribute GenBank data[46].

MMDBD: Medicinal Materials DNA Barcode Database (MMDBD) for retrieving data and searching for similarities. The Chinese Pharmacopoeia and the American Herbal Pharmacopoeia both mention approximately 1000 species of medicinal ingredients in MMDBD. It keeps track of DNA barcode sequences, basic information, and crucial medicinal ingredient references. The database's goals are to: (1) provide an organised and integrated web resource for medicinal species identification using DNA barcodes, (2) gather and integrate fundamental information on pharmaceuticals and their DNA barcodes, and (3) create online resources and tools for sequence comparison[115].

BOLD: The Barcode of Life Data System (BOLD) is a bioinformatics workbench that helps in DNA barcode record capture, storage, analysis, and publication.It bridges a typical bioinformatics gap by combining molecular, morphological, and distributional data.Several items must be submitted for a sequence to receive 'official' barcode status in BOLD: species name, voucher data (containing institution and catalogue information), collection record, specimen identifier, sequence of >500 bp, primer information, and raw sequence data files. Before making data public, BOLD administrators undertake quality checks on it after it has been published (i.e.,assurance that the sequence is not a contaminant, that it is a real functional copy, and that it is of sufficient quality)[116, 117]. For rbcL and matK sequences in plants, the BLAST method is used instead of the conventional BOLD identification engine[84].

BioBarcode: The BioBarcode database server system seeks to provide a cost-effective bioinformatics protocol that Asian researchers and research organisations interested in DNA barcoding may utilise for free. By providing specialised services, the BioBarcode promotes the rapid acquisition of biological species DNA sequence data that meets global standards, as well as useful tools that will make barcoding cheaper and faster in the biodiversity community, such as standardisation, depository, management, and analysis of DNA barcode data. A chromatogram viewer is included in the BioBarcode database, which increases DNA sequence analysis performance. The standards feature eight requirements that must be met for records to achieve official barcode status, which were adapted from CBOL and GenBank [84, 118].

Conclusion

In this paper,after briefly reviewing the classical DNA barcoding method in plants the loopholes and challenges posed by this technique could be well understood. It cannot be denied that plant barcoding is still in its early stages, with much more development to be done. As a result, the DNA barcoding method has drawn a lot of criticism, although there are ways to improve it. Researchers have devised methods like as super barcoding and ultra barcoding, which employ the whole chloroplast genome and nrDNA sequences as barcode identifiers and surpass single-locus markers. In damaged DNA samples, techniques like DNA Mini-barcoding use a shorter length of DNA, usually less than or equal to 200bp, that can be amplified faster than standard barcodes.HRM analysis and ME barcoding technologies are integrated with traditional DNA methods to reduce costs and technician time. Next-generation sequencing (NGS) high throughput technologies, having the ability to generate hundreds of thousands to tens of millions of sequencing reads in parallel, could be used to overcome some of the shortcomings of Sanger-based sequencing, allowing the creation of DNA barcodes to be done more quickly and at a reduced total cost.Without bioinformatics, traditional DNA barcoding is ineffectual and time-consuming. Bioinformatics tools and databases are needed to help with data collecting, storage, analysis, visualisation, and proper administration at various phases of the barcoding process Next-generation sequencing technology and bioinformatics advances will pave the way for a promising future for the DNA barcoding approach.

Conflict of Interest

The authors declare no conflicting interests.

Funding

Not applicable.

References

1. Cope JS, Corney D, Clark JY, et al. Plant species identification using digital morphometrics: A review. *Expert Systems with Applications* 2012; 39: 7562–7573.
2. Lee SH, Chan CS, Wilkin P, et al. Deep-plant: Plant identification with convolutional neural networks. In: *2015 IEEE international conference on image processing (ICIP)*. IEEE, 2015, pp. 452–456.
3. Techen N, Parveen I, Pan Z, et al. DNA barcoding of medicinal plant material for identification. *Current opinion in Biotechnology* 2014; 25: 103–110.
4. Ahmed SS. DNA Barcoding in Plants and Animals: A Critical Review.

5. Hebert PD, Cywinska A, Ball SL, et al. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B: Biological Sciences* 2003; 270: 313–321.
6. Vere N de, Rich TC, Trinder SA, et al. DNA barcoding for plants. In: *Plant Genotyping*. Springer, 2015, pp. 101–118.
7. Parvathy VA, Swetha VP, Sheeja TE, et al. DNA barcoding to detect chilli adulteration in traded black pepper powder. *Food Biotechnology* 2014; 28: 25–40.
8. Nithaniyal S, Newmaster SG, Ragupathy S, et al. DNA Barcode Authentication of Wood Samples of Threatened and Commercial Timber Trees within the Tropical Dry Evergreen Forest of India. *PLoS ONE* 2014; 9: e107669.
9. Bell KL, de Vere N, Keller A, et al. Pollen DNA barcoding: current applications and future prospects. *The 6th International Barcode of Life Conference* 2017; 01: 629–640.
10. Bruni I, De Mattia F, Galimberti A, et al. Identification of poisonous plants by DNA barcoding approach. *International Journal of Legal Medicine* 2010; 124: 595–603.
11. *The Global Taxonomy Initiative 2020: A Step-by-Step Guide for DNA Barcoding*. Centre for Biodiversity Genomics, University of Guelph: Secretariat of the Convention on Biological Diversity, Montreal, 2021.
12. Yu J, Wu X, Liu C, et al. Progress in the use of DNA barcodes in the identification and classification of medicinal plants. *Ecotoxicology and Environmental Safety* 2021; 208: 111691.
13. Group CPW, Hollingsworth PM, Forrest LL, et al. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences* 2009; 106: 12794–12797.
14. Chen S, Yao H, Han J, et al. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PloS one* 2010; 5: e8613.
15. Li X, Yang Y, Henry RJ, et al. Plant DNA barcoding: from gene to genome: Plant identification using DNA barcodes. *Biol Rev* 2015; 90: 157–166.
16. Newmaster SG, Grguric M, Shanmughanandhan D, et al. DNA barcoding detects contamination and substitution in North American herbal products. *BMC medicine* 2013; 11: 1–13.
17. FU Y-M, JIANG W-M, FU C-X. Identification of species within Tetrastigma (Miq.) Planch.(Vitaceae) based on DNA barcoding techniques. *Journal of Systematics and Evolution* 2011; 49: 237–245.
18. Purushothaman N, Newmaster SG, Ragupathy S, et al. A tiered barcode authentication tool to differentiate medicinal Cassia species in India. *Genet Mol Res* 2014; 13: 2959–68.
19. de Vere N, Rich TC, Ford CR, et al. DNA barcoding the native flowering plants and conifers of Wales. *PloS one* 2012; 7: e37945.

20. Särkinen T, Staats M, Richardson JE, et al. How to open the treasure chest? Optimising DNA extraction from herbarium specimens.

21. Shokralla S, Gibson JF, Nikbakht H, et al. Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. *Molecular ecology resources* 2014; 14: 892–901.

22. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences* 1977; 74: 5463–5467.

23. Polz MF, Cavanaugh CM. Bias in template-to-product ratios in multitemplate PCR. *Applied and environmental Microbiology* 1998; 64: 3724–3730.

24. Cowan RS, Fay MF. Challenges in the DNA Barcoding of Plant Material. In: Sucher NJ, Hennell JR, Carles MC (eds) *Plant DNA Fingerprinting and Barcoding*. Totowa, NJ: Humana Press, pp. 23–33.

25. Pusch CM, Giddings I, Scholz M. Repair of degraded duplex DNA from prehistoric samples using Escherichia coli DNA polymerase I and T4 DNA ligase. *Nucleic acids research* 1998; 26: 3189–3194.

26. Sarwat M, Yamdagni MM. DNA barcoding, microarrays and next generation sequencing: recent tools for genetic diversity estimation and authentication of medicinal plants. *Critical Reviews in Biotechnology* 2016; 36: 191–203.

27. Casiraghi M, Labra M, Ferri E, et al. DNA barcoding: a six-question tour to improve users' awareness about the method. *Briefings in bioinformatics* 2010; 11: 440–453.

28. Genome List - Genome - NCBI, https://www.ncbi.nlm.nih.gov/genome/browse/#!/overview/ (accessed 5 June 2022).

29. Govindaraghavan S, Hennell JR, Sucher NJ. From classical taxonomy to genome and metabolome: towards comprehensive quality standards for medicinal herb raw materials and extracts. *Fitoterapia* 2012; 83: 979–988.

30. Parks M, Cronn R, Liston A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC biology* 2009; 7: 1–17.

31. Huang CY, Grunheit N, Ahmadinejad N, et al. Mutational decay and age of chloroplast and mitochondrial genomes transferred recently to angiosperm nuclear chromosomes. *Plant Physiology* 2005; 138: 1723–1733.

32. Luo H, Sun Z, Arndt W, et al. Gene order phylogeny and the evolution of methanogens. *PLoS One* 2009; 4: e6069.

33. Luo H, Shi J, Arndt W, et al. Gene order phylogeny of the genus Prochlorococcus. *PLoS One* 2008; 3: e3837.

34. Adolt R, Pavlis J. Age structure and growth of Dracaena cinnabari populations on Socotra. *Trees* 2004; 18: 43–53.
35. Aslam J, Mujib A, Sharma MP. In vitro micropropagation of Dracaena sanderiana Sander ex Mast: An important indoor ornamental plant. *Saudi journal of biological sciences* 2013; 20: 63–68.
36. Gupta D, Bleakley B, Gupta RK. Dragon's blood: botany, chemistry and therapeutic uses. *Journal of ethnopharmacology* 2008; 115: 361–380.
37. Wilkin P, Suksathan P, Keeratikiat K, et al. A new threatened endemic species from central and northeastern Thailand, Dracaena jayniana (Asparagaceae: tribe Nolinoideae). *Kew Bulletin* 2012; 67: 697–705.
38. Chen XQ, Turland NJ. Dracaena vandelli ex Linnaeus. *Flora of China* 2000; 24: 215–217.
39. Wu L, Wu M, Cui N, et al. Plant super-barcode: a case study on genome-based identification for closely related species of Fritillaria. *Chinese medicine* 2021; 16: 1–11.
40. Ji Y, Liu C, Yang J, et al. Ultra-barcoding discovers a cryptic species in Paris yunnanensis (Melanthiaceae), a medicinally important plant. *Frontiers in plant science* 2020; 11: 411.
41. Kane N, Sveinsson S, Dempewolf H, et al. Ultra-barcoding in cacao (Theobroma spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *American Journal of Botany* 2012; 99: 320–329.
42. Nock CJ, Waters DL, Edwards MA, et al. Chloroplast genome sequences from total DNA for plant identification. *Plant biotechnology journal* 2011; 9: 328–333.
43. Steele PR, Pires JC. Biodiversity assessment: State-of-the-art techniques in phylogenomics and species identification. *American Journal of Botany* 2011; 98: 415–425.
44. Hillis DM, Moritz C, Porter CA, et al. Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science* 1991; 251: 308–310.
45. Li H. *The genus Paris (Trilliaceae)*. Science Press, 1998.
46. Gesto-Borroto R, Medina-Jiménez K, Lorence A, et al. Application of DNA barcoding for quality control of herbal drugs and their phytopharmaceuticals. *Revista Brasileira de Farmacognosia* 2021; 31: 127–141.
47. Palais RA, Liew MA, Wittwer CT. Quantitative heteroduplex analysis for single nucleotide polymorphism genotyping. *Analytical biochemistry* 2005; 346: 167–175.
48. Mishra P, Kumar A, Nagireddy A, et al. DNA barcoding: an efficient tool to overcome authentication challenges in the herbal market. *Plant biotechnology journal* 2016; 14: 8–21.

49. Reed GH, Kent JO, Wittwer CT. High-resolution DNA melting analysis for simple and efficient molecular diagnostics.
50. Simko I. High-Resolution DNA Melting Analysis in Plant Research. *Trends in Plant Science* 2016; 21: 528–537.
51. Jung J, Kim KH, Yang K, et al. Practical application of DNA markers for high-throughput authentication of Panax ginseng and Panax quinquefolius from commercial ginseng products. *Journal of Ginseng Research* 2014; 38: 123–129.
52. Ganopoulos I, Bazakos C, Madesis P, et al. Barcode DNA high-resolution melting (Bar-HRM) analysis as a novel close-tubed and accurate tool for olive oil forensic use. *Journal of the Science of Food and Agriculture* 2013; 93: 2281–2286.
53. Vietina M, Agrimonti C, Marmiroli N. Detection of plant oil DNA using high resolution melting (HRM) post PCR analysis: A tool for disclosure of olive oil adulteration. *Food Chemistry* 2013; 141: 3820–3826.
54. Pereira L, Gomes S, Barrias S, et al. Applying high-resolution melting (HRM) technology to olive oil and wine authenticity. *Food Research International* 2018; 103: 170–181.
55. Sun W, Li J, Xiong C, et al. The potential power of Bar-HRM technology in herbal medicine identification. *Frontiers in Plant Science* 2016; 7: 367.
56. Thongkhao K, Tungphatthong C, Phadungcharoen T, et al. The use of plant DNA barcoding coupled with HRM analysis to differentiate edible vegetables from poisonous plants for food safety. *Food Control* 2020; 109: 106896.
57. Hajibabaei M, McKenna C. DNA mini-barcodes. In: *DNA barcodes*. Springer, 2012, pp. 339–353.
58. Zimmermann J, Hajibabaei M, Blackburn DC, et al. DNA damage in preserved specimens and tissue samples: a molecular assessment. *Frontiers in Zoology* 2008; 5: 1–13.
59. Meusnier I, Singer GA, Landry J-F, et al. A universal DNA mini-barcode for biodiversity analysis. *BMC genomics* 2008; 9: 1–4.
60. Hajibabaei M, Smith MA, Janzen DH, et al. A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes* 2006; 6: 959–964.
61. Min XJ, Hickey DA. DNA barcodes provide a quick preview of mitochondrial genome composition. *PLOS one* 2007; 2: e325.
62. Gao Z, Liu Y, Wang X, et al. DNA mini-barcoding: a derived barcoding method for herbal molecular identification. *Frontiers in plant science* 2019; 987.
63. Hajibabaei M, Singer GA, Clare EL, et al. Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. *BMC biology* 2007; 5: 1–7.
64. Little DP. Authentication of Ginkgo biloba herbal dietary supplements using DNA barcoding. *Genome* 2014; 57: 513–516.

65. Paradiso L, Little DP. Authentication of garlic (*Allium sativum* L.) supplements using a *trnL* ^{UAA} mini-barcode. *Genome* 2021; 64: 1021–1028.
66. Metzker ML. Sequencing technologies—the next generation. *Nature reviews genetics* 2010; 11: 31–46.
67. Meyer CP, Paulay G. DNA Barcoding: Error Rates Based on Comprehensive Sampling. *PLoS Biol* 2005; 3: e422.
68. Shokralla S, Spall JL, Gibson JF, et al. Next-generation sequencing technologies for environmental DNA research. *Molecular ecology* 2012; 21: 1794–1805.
69. Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005; 437: 376–380.
70. Sucher NJ, Hennell JR, Carles MC. DNA fingerprinting, DNA barcoding, and next generation sequencing technology in plants. *Methods Mol Biol* 2012; 862: 13–22.
71. Ahmadian A, Ehn M, Hober S. Pyrosequencing: history, biochemistry and future. *Clinica chimica acta* 2006; 363: 83–94.
72. Ronaghi M, Karamohamed S, Pettersson B, et al. Real-time DNA sequencing using detection of pyrophosphate release. *Analytical biochemistry* 1996; 242: 84–89.
73. Shendure J, Ji H. Next-generation DNA sequencing. *Nature biotechnology* 2008; 26: 1135–1145.
74. Dressman D, Yan H, Traverso G, et al. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proceedings of the National Academy of Sciences* 2003; 100: 8817–8822.
75. Illumina I. Illumina sequencing technology. *Technology Spotlight*.
76. Adessi C, Matton G, Ayala G, et al. Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic acids research* 2000; 28: e87–e87.
77. Gostel MR, Zúñiga JD, Kress WJ, et al. Microfluidic Enrichment Barcoding (MEBarcoding): a new method for high throughput plant DNA barcoding. *Scientific reports* 2020; 10: 1–13.
78. Luscombe NM, Greenbaum D, Gerstein M. What is bioinformatics? A proposed definition and overview of the field. *Methods of information in medicine* 2001; 40: 346–358.
79. Moody G. *Digital code of life: how bioinformatics is revolutionizing science, medicine, and business*. John Wiley & Sons, 2004.
80. Gauthier J, Vincent AT, Charette SJ, et al. A brief history of bioinformatics. *Briefings in Bioinformatics* 2019; 20: 1981–1996.

81. Bhargava M, Sharma A. DNA barcoding in plants: evolution and applications of in silico approaches and resources. *Molecular phylogenetics and evolution* 2013; 67: 631–641.
82. Wilson J-J, Sing K-W, Jaturas N. DNA barcoding: Bioinformatics workflows for beginners. *Encyclopedia of Bioinformatics and Computational Biology* 2019; 985–995.
83. The International Barcode of Life Consortium. *International Barcode of Life*, https://ibol.org/ (accessed 31 May 2022).
84. Paulchamy C. Bioinformatics Resources for DNA barcoding. 2012.
85. Schindel DE, Miller SE. DNA barcoding and the Consortium for the Barcode of Life. *Success Stories in Implementation of the Programmes of Work on Dry and Sub-Humid Lands and the Global Taxonomy Initiative*.
86. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research* 1994; 22: 4673–4680.
87. Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of molecular biology* 2000; 302: 205–217.
88. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* 2004; 32: 1792–1797.
89. Steinke D, Vences M, Salzburger W, et al. TaxI: a software tool for DNA barcoding using distance methods. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2005; 360: 1975–1980.
90. Peng C. Distance based methods in phylogenetic tree construction. *NEURAL PARALLEL AND SCIENTIFIC COMPUTATIIONS* 2007; 15: 547.
91. Xiong J. *Essential bioinformatics*. Cambridge University Press, 2006.
92. Sarkar IN, Planet PJ, Desalle ROB. CAOS software for use in character-based DNA barcoding. *Molecular ecology resources* 2008; 8: 1256–1259.
93. Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of molecular evolution* 1981; 17: 368–376.
94. Tamura K, Peterson D, Peterson N, et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* 2011; 28: 2731–2739.
95. Swofford DL, Sullivan J. Phylogeny inference based on parsimony and other methods using PAUP*. *The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny, cáp* 2003; 7: 160–206.
96. Raclariu AC, Mocan A, Popa MO, et al. Veronica officinalis product authentication using DNA metabarcoding and HPLC-MS reveals widespread adulteration with Veronica chamaedrys. *Frontiers in pharmacology* 2017; 8: 378.

97. Zhang AB, Sikes DS, Muster C, et al. Inferring species membership using DNA sequences with back-propagation neural networks. *Systematic Biology* 2008; 57: 202–215.
98. PARKER D. Optimal algorithm for adaptive networks: Second order backprop, second order direct propagation, and second order hebian learning. In: *Proceedings of The IEEE First International Conference on Neural Networks, San Diego, 1987*. 1987.
99. Smith M. *Neural networks for statistical modeling*. Thomson Learning, 1993.
100. Zhang A, Feng J, Ward RD, et al. A new method for species identification via protein-coding and non-coding DNA barcodes by combining machine learning with bioinformatic methods. *PLoS One* 2012; 7: e30986.
101. He T, Jiao L, Yu M, et al. DNA barcoding authentication for the wood of eight endangered Dalbergia timber species using machine learning approaches. *Holzforschung* 2019; 73: 277–285.
102. Dev SA, Unnikrishnan R, Jayaraj R, et al. Quantification of adulteration in traded ayurvedic raw drugs employing machine learning approaches with DNA barcode database. *3 Biotech* 2021; 11: 1–16.
103. Chan RH, Wang RW, Yeung HM. Composition vector method for phylogenetics-a review. In: *Proc. 9th Int. Symp. Operations Research and Its Applications (ORSC & APORC, Chengdu, China, 2010)*. Citeseer, 2010, p. 13.
104. Qi J, Luo H, Hao B. CVTree: a phylogenetic tree reconstruction tool based on whole genomes. *Nucleic acids research* 2004; 32: W45–W47.
105. Li CP, Yu ZG, Han GS, et al. Analyzing multi-locus plant barcoding datasets with a composition vector method based on adjustable weighted distance.
106. Brown SDJ, Collins RA, Boyer S, et al. Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources* 2012; 12: 562–565.
107. Riaz T, Shehzad W, Viari A, et al. ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research* 2011; 39: e145–e145.
108. Ficetola GF, Coissac E, Zundel S, et al. An in silico approach for the evaluation of DNA barcodes. *BMC genomics* 2010; 11: 1–10.
109. Jones M, Ghoorah A, Blaxter M. jMOTU and taxonerator: turning DNA barcode sequences into annotated operational taxonomic units. *PLoS one* 2011; 6: e19259.
110. Kumar S, Carlsen T, Mevik B-H, et al. CLOTU: an online pipeline for processing and clustering of 454 amplicon reads into OTUs followed by taxonomic annotation. *BMC bioinformatics* 2011; 12: 1–9.

111. Little DP. DNA barcode sequence identification incorporating taxonomic hierarchy and within taxon variability. *PLoS one* 2011; 6: e20552.
112. Sandionigi A, Galimberti A, Labra M, et al. Analytical approaches for DNA barcoding data–how to find a way for plants? *Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology* 2012; 146: 805–813.
113. Tyagi A, Bag SK, Shukla V, et al. Oligonucleotide frequencies of barcoding loci can discriminate species across kingdoms. *PLoS One* 2010; 5: e12330.
114. Benson DA, Cavanaugh M, Clark K, et al. GenBank. *Nucleic acids research* 2012; 41: D36–D42.
115. Lou S-K, Wong K-L, Li M, et al. An integrated web medicinal materials DNA database: MMDBD (Medicinal Materials DNA Barcode Database). *BMC genomics* 2010; 11: 1–8.
116. Ratnasingham S, Hebert PD. BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). *Molecular ecology notes* 2007; 7: 355–364.
117. Meiklejohn KA, Damaso N, Robertson JM. Assessment of BOLD and GenBank–Their accuracy and reliability for the identification of biological materials. *PloS one* 2019; 14: e0217084.
118. Lim J, Kim S-Y, Kim S, et al. BioBarcode: a general DNA barcoding database and server platform for Asian biodiversity resources. In: *BMC genomics*. Springer, 2009, pp. 1–10.

How to cite this article:

Ahmed F, Zaman K. A critical review on the challenges and advances in dna barcoding for plant identification, *Curr Trends Pharm Res,*2022;9 (1): 115-139.