DNA Barcoding for the Identification of Botanicals in Herbal Medicine and Dietary Supplements: Strengths and Limitations

Abstract

In the past decades, the use of traditional medicine has increased globally, leading to a booming herbal medicine and dietary supplement industry. The increased popularity of herbal products has led to a rise in demand for botanical raw materials. Accurate identification of medicinal herbs is a legal requirement in most countries and necessary for delivering a quality product that meets consumer expectations. Traditional identification methods include botanical taxonomy, macroscopic and microscopic examination, and chemical methods. Advances in the identification of biological species using DNA-based techniques have led to the development of a DNA marker-based platform for authentication of plant materials. DNA barcoding, in particular, has been proposed as a means to identify herbal ingredients and to detect adulteration. However, general barcoding techniques using universal primers have been shown to provide mixed results with regard to data accuracy. Further technological advances such as mini-barcodes, digital polymerase chain reaction, and next generation sequencing provide additional tools for the authentication of herbs, and may be successful in identifying processed ingredients used in finished herbal products. This review gives an overview on the strengths and limitations of DNA barcoding techniques for botanical ingredient identification. Based on the available information, we do not recommend the use of universal primers for DNA barcoding of processed plant material as a sole means of species identification, but suggest an approach combining DNA-based methods using genus- or species-specific primers, chemical analysis, and microscopic and macroscopic methods for the successful authentication of botanical ingredients used in the herbal dietary supplement industry.

Introduction

DNA barcodes, a term introduced by Hebert et al. [1], describe short genomic regions from the nuclear and/or organelle genome used to distinguish animal, plant, fungal, and bacterial species. The use of these short genomic regions for biological species discrimination is called DNA barcoding. Applications include tracking the illegal trade of endangered species of both animals and plants (i.e., biopiracy) [2–5], forensic analysis [6,7] identifying invasive species [8,9], plant identification at any stage of the life cycle (juvenile or mature) [10], identifying complex food webs by studying species diversity in the gut contents of animals [11], analyzing herbivore’s diet components [12], checking adulterations and substitutions in food products [13], and authentication of herbal medicine and identification of botanical ingredient adulterants [14–17].

The herbal products industry is a multibillion-dollar industry and an important part of the world’s economy. In a study funded by the Natural Products Foundation, estimates suggest that the total economic contribution of the dietary supplement industry to the U.S. economy is more than three times the annual consumer sales, or $61 billion dollars per year [18]. However, as the popularity of herbal dietary supplements has increased, so have reports of adulteration; this admixture, or substitution, of herbal products/ingredients with materials of substandard quality is a growing concern since it may lead to decreased efficacy and the occurrence of serious adverse effects. Price pressure, increased demand, limited availability of medicinal herbs, and greed...
of unscrupulous suppliers are some of the reasons for the intentional substitution of botanical ingredients. Adulteration is often carried out using a closely related species having similar active ingredients (although in some cases, the use of closely related species is acceptable in pharmacopeial monographs) to fool the traditionally used authentication methods, but can occur with completely different lower-cost substitutes as well. Adulteration can also be accidental, for example, due to the complex nomenclature of medicinal herbs or the unintentional misidentification of plant species collected in the wild. Worldwide, a variety of different names to indicate a specific plant species is currently in use: pharmaceutical names, scientific binomials, older scientific names called synonyms, and common names (see review by Barnes et al. [19]). Confusion often arises when different plants have the same common names, which may lead to situations where the local gatherer collects the wrong material. For example, the name fang ji is used for two Chinese herbs – han fang ji (Stephania tetranda) and guang fang ji (Aristolochia fangchi). Both materials are used in traditional Chinese medicine (TCM) for the treatment of similar ailments, but A. fangchi roots contain aristolochic acids that are nephrotoxic and can cause urothelial carcinoma [20]. Notably, reports from Belgium in the 1990s detailed lochic acids that are nephrotoxic and can cause urothelial carcinoma [20]. Notably, reports from Belgium in the 1990s detailed lochic acids that are nephrotoxic and can cause urothelial carcinoma [20]. Notably, reports from Belgium in the 1990s detailed.

Several different names to indicate a specific plant species is currently in use: pharmaceutical names, scientific binomials, older scientific names called synonyms, and common names (see review by Barnes et al. [19]). Confusion often arises when different plants have the same common names, which may lead to situations where the local gatherer collects the wrong material. For example, the name fang ji is used for two Chinese herbs – han fang ji (Stephania tetranda) and guang fang ji (Aristolochia fangchi). Both materials are used in traditional Chinese medicine (TCM) for the treatment of similar ailments, but A. fangchi roots contain aristolochic acids that are nephrotoxic and can cause urothelial carcinoma [20]. Notably, reports from Belgium in the 1990s detailed lochic acids that are nephrotoxic and can cause urothelial carcinoma [20]. Notably, reports from Belgium in the 1990s detailed lochic acids that are nephrotoxic and can cause urothelial carcinoma [20]. Notably, reports from Belgium in the 1990s detailed.

The concept of the barcode is an analogy to the combination and spacing of black and white lines that can be found on the package of almost any commercial item. This “barcode” is a representation of the unique item. It can be scanned and compared to a database to identify the item and consequently its price. Barcodes of plants do not consist of black and white lines, but, instead, consist of a unique combination of nucleotides of roughly 300–1000 base pairs (bp) in length that are ideally specific to a plant of interest. These DNA barcodes are agreed-upon DNA sequences from either the nuclear or organelle (chloroplast, mitochondria) genome. In the strictest term, “DNA barcoding” refers to the technique where millions of copies of these diagnostic sequences are made using universal primer sets in a polymerase chain reaction (PCR), and which are subsequently identified by a sequencing method. The universal primer sets used for PCR were designed to bind to conserved flanking regions of the diagnostic/species identifying sequence found in most plants. The unique nucleotide sequence (DNA barcode) between the flanking regions can be determined and used to compare to voucher sequences stored in either house or openly accessible databases, such as GenBank, using the BLAST (Basic Local Alignment Search Tool) algorithm [27]. The open databases are collections of sequences submitted by individual researchers, labs, and institutions but are not curated, and the accuracy has not been verified or confirmed. Frequently, several different sequences are reported from different sources with multiple possible barcode regions of varying lengths. For example, Table 1 S, Supporting Information, depicts publicly available barcode sequences and their sources [28]. The data used to standardize the targeted barcodes comes from a range of different sources with different levels of certainty. Some of the barcodes come from wild species from unknown locations, without vouchers, the collector or the date of collection identified, the part of the plant may not be known, and the number of nucleotides sequenced in the barcode region may be different between samples of the same species (Table 1 S, Supporting Information).

Given that a 99–100% match is used to confirm the species, a match of 98% or less gives only confidence to the genus level [29], the integrity and reliability of the reference materials are crucial to the accuracy of the technique. The best gene region for barcoding of land plants is a matter of debate, and molecular biologists have used a number of different single loci, or combinations of loci, to authenticate plant materials [30]. A recent development is the generation of “mini-DNA barcodes” (Fig. 1). They represent only a fraction of the agreed upon barcodes and are usually 100–200 bp in length. Amplification of mini-barcodes is done by primer sets that are less universal and more specific to either the genus of a plant or, if possible, unique to only one plant species of a genus.

Though DNA barcoding is independent of morphological characteristics and physical and seasonal variations, it has limitations as a stand-alone method for authentication of processed botanicals. Hence, the current review highlights the potentials and pitfalls of DNA barcoding with an emphasis on DNA barcoding methodology used for the identification of herbs and future directions for authentication of herbal ingredients using different techniques.

Table 1 S, Supporting Information, depicts publicly available barcode sequences and their sources [28]. The data used to standardize the targeted barcodes comes from a range of different sources with different levels of certainty. Some of the barcodes come from wild species from unknown locations, without vouchers, the collector or the date of collection identified, the part of the plant may not be known, and the number of nucleotides sequenced in the barcode region may be different between samples of the same species (Table 1 S, Supporting Information).

Given that a 99–100% match is used to confirm the species, a match of 98% or less gives only confidence to the genus level [29], the integrity and reliability of the reference materials are crucial to the accuracy of the technique. The best gene region for barcoding of land plants is a matter of debate, and molecular biologists have used a number of different single loci, or combinations of loci, to authenticate plant materials [30]. A recent development is the generation of “mini-DNA barcodes” (Fig. 1). They represent only a fraction of the agreed upon barcodes and are usually 100–200 bp in length. Amplification of mini-barcodes is done by primer sets that are less universal and more specific to either the genus of a plant or, if possible, unique to only one plant species of a genus.

Though DNA barcoding is independent of morphological characteristics and physical and seasonal variations, it has limitations as a stand-alone method for authentication of processed botanicals. Hence, the current review highlights the potentials and pitfalls of DNA barcoding with an emphasis on DNA barcoding methodology used for the identification of herbs and future directions for authentication of herbal ingredients using different techniques.
DNA may be obtained from leaves and flowers because of the early stages of DNA extraction [41, 44]. Levels of polysaccharides and polyphenols that must be removed may affect the quality and quantity of DNA. In addition, plant metabolites such as polysaccharides, flavonoids, polyphenols, and terpene lactones may hinder DNA isolation. The polysaccharides and certain secondary metabolites reportedly coprecipitate with the DNA and prevent the complete dissolution of DNA, thus requiring modified protocols for the isolation of DNA from materials containing polysaccharides and problematic secondary metabolites [40–43]. The most widely used approaches to extract genomic DNA are the cetyl trimethyl ammonium bromide (CTAB) method [33] and commercial DNA extraction kits [37]. However, CTAB/kit methods are not successful in isolating DNA from plants or plant parts that contain high amounts of secondary metabolites. Roots, rhizomes, and tubers may contain particularly high levels of polysaccharides and polyphenols that must be removed using protocols with added high concentrations of CTAB, polyvinylpyrrolidone (PVP), and β-mercaptoethanol (β-Me) during the early stages of DNA extraction [41, 44–46]. High-quality DNA may be obtained from leaves and flowers because of the low levels of interfering metabolites and fibers. Especially the DNA obtained from fresh and young leaves and flowers can be used to prepare a crude extract, and this solution can be used in direct PCR to amplify the target DNA barcode [47]. However, some plant leaves like tomato, cotton, and tea contain high concentrations of polyphenols and tannins, which also hinder DNA isolation and PCR amplification, thus rendering them unsuitable for direct PCR. Therefore, modified DNA extraction methods were developed for the isolation of DNA from tissues or plants containing high amounts of phenolic compounds and tannins [45, 48]. Similarly dried stems, roots, and fruits may not be suitable for direct PCR as they contain secondary metabolites that inhibit PCR amplification while fresh tissues obtained right after the collection of medicinal herbs are preferred, most of the herbal materials are available for analysis either in dried or powdered form, or, after further processing, as capsules, tablets, or liquids. In modern phytomedicines, the plant DNA is often either removed or degraded during the manufacturing processes of herbal products. Hence, the DNA extracted from capsules, tablets, and liquid extracts appear often as a smear on an analytical agarose gel due to fragmentation, or no band appears due to complete removal of DNA (Fig. 2). Wallace et al. [49] tested 95 natural health products using a standard DNA barcoding technique with multiple markers and primer sets. The authors were unable to retrieve DNA barcodes from 25% of the tested plant products of which 2 were plant roots and 14 were capsules, tablets, or caplets. Their results demonstrated that DNA amplification could be easily accomplished in botanical materials (33 in a total of 35 samples of teas and roots) as compared to pharmaceutical formulations (19 out of 33 samples of tablets, capsules and caplets). The difference in the results could be attributed to the fact that herbal product formulations often contain excipients (such as fillers, diluents, binders, glidants, lubricants, pigments, and stabilizers) that may affect the DNA extraction, or that the primer sets failed to amplify the targeted region. Costa et al. [50] evaluated the possible effect of four different pharmacological excipients on DNA isolation. All the tested excipients, talc, silica, iron oxide, and titanium dioxide, exhibited adsorbent properties that affected the extraction of DNA from natural products [50]. However, with slight modifications in the DNA extraction protocol, it is sometimes possible to obtain useful DNA as a template for PCR from a powder, capsules, or tablets [51]. In contrast to this, the processes by which tinctures and extracts are made, which can include extensive heat treatment, filtration, extractive distillation, or supercritical fluid extraction [52], thus degrading or removing the plant DNA completely, often make these materials unsuitable for DNA barcoding analysis [53].

Loci selected as DNA barcodes for herbal materials

The selection of a universal barcode region for identification of all land plants has proven to be quite challenging. Though the technique has successfully been used in the identification of animal species using cytochrome oxidase I (COI) from the mitochondrial genome as a universal barcode [1], barcoding plants is more difficult for many reasons. The slow evolutionary rate of the plant mitochondrial genome means that the mitochondrial gene regions, including the COI region, do not sufficiently distinguish plant species. Therefore, relatively fast-evolving plastid and nuclear genomes were proposed as alternative barcodes for plants [54–57]. The most common regions are matK, rbcL, ITS, ITS2, psbA-trnH, atpF-atpH, ycf5, psbK-I, psbM, trnD, cod, nad1, trnl-L, rpoB, rpoC1, and rps16 [54–57]. These regions have a relatively fast evolutionary rate compared to mitochondrial genes, can dis-
tiguous the species based on differences in the genetic code, and have conserved regions flanking the ends of the DNA sequence for the binding of universal primers. None of the individual plant DNA barcodes described to date have both differentiating regions and universal primer regions. Hence, a multiplex plant barcode with combinations of two or three loci was recommended [55, 56]. The Consortium for the Barcode of Life (CBOL) Plant Working group [56] suggested matK + rbcL as the preferred plant barcode combination. The matK locus is difficult to amplify in some plant genera since the universal primer-binding site is not perfectly conserved and, therefore, may not be useful in a two-locus combination in some plants. Another suggestion was the rbcL + trnH-psbA combination, but due to the high variability of the trnH-psbA sequence, it was difficult to align, and thus the two-locus barcode approach was found to be problematic for some of the plants. To overcome this problem, a tiered approach was suggested by Newmaster et al. [54]. The method utilizes the easily amplifiable and alignable rbcL region as a scaffold on which data from highly variable non-coding regions such as ITS2 or the trnH-psbA region are employed for identification of plant species. Using this tiered approach, approximately 75–80% of the tested plant species can reportedly be barcoded [54, 58].

Numerous studies have been published to identify medicinal plants using various suggested barcode loci. Techen et al. [30] reviewed the various barcode loci and methods used for the identification of medicinal plants. The most extensive study of DNA barcoding of medicinal plants was accomplished by Chen et al. [20]. Seven DNA regions, psbA-trnH, matK, rbcL, rpoC1, ycf5, ITS (consisting of both ITS1 and ITS2), and ITS2, were evaluated for identification of more than 6600 samples of medicinal plants (fresh leaves) and their closely related species [20]. Their data suggested that the nuclear ITS2 locus could identify 92.7% of the tested species (8557 medicinal plants and closely related samples belonging to 5905 species from 1010 diverse genera of 219 families in 7 phyla-angiosperms, gymnosperms, ferns, mosses, liverworts, algae, and fungi) and proposed ITS2 as the core barcode for medicinal plants [20]. Subsequently, ITS2 has been tested across a broad range of plant taxa with a large sample size and confirmed as an effective barcode for plants by the China Plant Barcode of Life (BOL) Group [59]. Chen and colleagues subsequently built a TCM barcode platform, called the Traditional Chinese Medicine Database using a two-locus barcode system containing ITS2 and psbA-trnH sequences [47]. This database contains barcodes belonging to more than 23,000 medicinal plant species and known adulterants [47]. Additional investigations on different plant groups confirmed the effectiveness of ITS2 for the identification of medicinal plant species. For example, 24 medicinal plants from the Fabaceae family and their adulterants were identified using ITS2, with a success rate of 37–97% [60]. Sun and Chen [61] successfully used the ITS2 barcode to distinguish 19 cortex herbs listed in the Chinese Pharmacopoeia, and Pang et al. [62] also reported ITS2 as a suitable barcode for the identification of herbal materials. Zhu et al. [63] recommended ITS2 locus to discriminate between Glehniae radix and its common adulterants. The major advantage of ITS2 as a barcode for the identification of herbal supplements is its short length (200–230 bp on average). In most herbal products and dietary supplements, the DNA is highly degraded into pieces of less than 500 nucleotides in length due to various processing methods and, consequently, universal long barcodes ranging from 600 to 800 bp, on average, may not be amplified. Hence, short-length barcodes that can be easily retrieved from dried, powdered form, or sometimes even from extracts were recommended [51]. Despite this advantage, the ITS2 locus was not found to be suitable for global identification of plants because of the presence of multiple copies of ITS2 within one individual in all plant species. The multiple ITS2 copies, which are not always homogenized by concerted evolution, led to the incorrect identification of species due to their similarity with the copies of the more closely related species. Furthermore, the same problem might arise in hybrids due to the biparental inheritance of ITS2. Another disadvantage of ITS is the technical problems in amplification and sequencing that can arise due to the presence of DNA from other species (e.g., fungi, which coexist with plants as endophytes and/or mycorrhizal symbionts) [64, 65]. Recently, Cheng et al. [66] reported that the most frequently used primer pairs (ITS1 + ITS4) for ITS were originally designed for fungi [67]. The less plant-specific primer sets led to low PCR and sequencing success rates in some of the plant groups, for instance, <50% PCR success for algae [68, 69] and ferns [20], 57.6% for gymnosperms [70], and 88.0% for angiosperms [70]. Therefore, Cheng et al. [66] designed universal...
plant-specific ITS primers for both plant DNA barcoding and plant systematics. Recently, the term “mini-barcodes” was introduced for short-length DNA markers used for the identification of botanical ingredients from processed herbal supplements [51, 71]. Mini-barcodes are short (<200 bp) sequences of DNA from standardized matK and rbcL barcode regions, which have been used, e.g., to identify and authenticate herbal supplements made from saw palmetto (Serenoa repens) fruit [71], Ginkgo biloba [51] leaf, and devil’s claw (Harpagophytum procumbens and Harpagophytum zeyheri) root and rhizome [72] herbal dietary supplements available in the markets of North America. The advantages of mini-barcodes are the easy retrieval of DNA markers even from processed dietary materials due to their small amplicon length, and their ability to distinguish closely related species because of the genus/species specificity.

**PCR amplification**

PCR, first devised by Mullis [73], is a molecular method by which a single copy or a few copies of a piece of DNA is amplified and thousands to millions of copies of a particular DNA sequence are generated. The method employs a heat-stable DNA polymerase, the nucleotides, template DNA, and DNA oligonucleotides (also called DNA primers). The general barcoding technique uses universal primers for rapid identification of plant species [57, 58, 74, 75]. The universal primer sets recommended for barcoding of plant species are selected to amplify DNA from four genomic regions, namely ITS/ITS2 from the nuclear genome, and matK, rbcL, and trnH-psbA from the chloroplast genome. Inherent biases in the analysis can lead to false positives and false negatives [76]. Biases can occur when the sequence at one of the universal priming sites varies sufficiently to prevent efficient annealing, when different species have a different number of copies of the chosen region leading to over or underestimation unrelated to the actual composition of plant material, when the target DNA is degraded or fragmented, or from the manner of extraction and preparation of the DNA for sequencing [65, 76]. Soares et al. [77] and Costa et al. [78] demonstrated that the reliability of DNA barcode methods in complex plant samples varied depending on which of the commercial DNA extraction kits was used for the initial sample preparations. The efficiency of the PCR reaction itself can impose a bias on the barcoding results. Differences in the melting temperatures of the primers could lead to a reduced amplification rate and the affinity of universal primers to template DNA of all known and unknown organisms, and a balanced melting temperature of the primer pairs are two important criteria to produce robust amplification [79]. Further, the presence of inhibitory secondary metabolites and inactive ingredients in tablets and capsules can also reduce the efficiency of PCR amplification or lead to false negative results. The use of excipient materials made from wheat, rice, or soy is common in manufacturing processes to thousands to millions of copies of a particular DNA sequence are generated. In digital PCR, the samples are diluted (1:5–1:50000) in a suitable buffer to an extent that there is approximately one DNA template molecule per µL. The goal is to dilute the DNA to the extent that a few samples contain only molecules of low abundant DNA. The term “low abundant DNA” refers to DNA that originated either from the medicinal plant material present within a large amount of DNA from filler material (e.g., rice flour) or DNA derived from small amounts of adulterating plant material. By preparing several PCRs using one molecule/µL DNA solution as a template, chances are that DNA molecules of low abundance get amplified and, consequently, detected. The number of PCRs required depends on the expected frequency of the DNA to be detected. For a sample containing 10% of the low abundant DNA, 2 out of 20 PCRs may result in the amplification of low abundant DNA. The analysis showed 9 (24.3%) of the 37 herbal supplements required digital PCR to separate excipient DNA from possible ginkgo DNA. In the study, digital PCR produced amplicons of ginkgo, rice, and an unidentifiable species [51].

**Sequencing methods**

The conventional method used for generating DNA sequence data to obtain a barcode from PCR amplicons is Sanger’s di-deoxy method of sequencing [81]. Sanger’s sequencing technology is capable of generating sequencing reads of up to 1000 bases and has been the approach used for DNA sequencing in most of the DNA barcode analyses published to date. The inherent limitations of Sanger-based DNA sequencing are low throughput and the requirement for high concentrations of DNA (100–500 ng) to avoid biases and errors [82]. Moreover, the method provides two sequencing signal patterns, or electropherograms, for each sequence generated [83]. Hence, the Sanger sequencing method is suitable for herbal materials that contain only a single medicinal plant. If the herbal preparation (e.g., a dietary supplement) contains multiple plant species or excipients, co-amplification of barcode sequences from other material than the intended one can occur due to the nature of the universal primers during the PCR amplification step. This leads to the production of multiple/overlapping sequencing peaks and, consequently, a failure of sequencing because the correct DNA sequence of the barcode cannot be determined [Fig. 3]. Moreover, multiple sequences may also create confusion in the identification of the “true” barcode and other sequences. Additionally, many plants have symbiotic associations with bacteria, algae, or fungi [64, 65]. The occurrence of multiple copies of fungal ITS barcodes creates difficulties in direct Sanger sequencing. Most of these situations can lead to ambiguity or false information when the Sanger sequencing method is employed in generating DNA barcodes, and may result in repeated or failed sequencing attempts. The poor read quality can be improved by processes upstream to sequencing, for example cloning in a suitable bacterial or microbrial host. The DNA fragments are ligated with a vector and cloned in bacteria. Several bacterial clones are then sequenced to identify the different DNA sequences. However, cloning introduces biases against extreme base composition (e.g., stretches with high guanine and cytosine contents), inverted repeats, and genes not accepted by the bacterial cloning host [84]. To overcome the limitations of Sanger-based sequencing for DNA barcoding of processed or mixed samples, a high-throughput sequencing method called next-generation sequencing (NGS) has been used [85]. The NGS technology allows parallel sequencing of multiple DNA fragments from various DNA templates in a single reaction [85]. It can generate up to one million DNA sequences that are
up to 700 bases in length in a single sequencing run, though the base length is highly variable depending on the NGS platform/technology being used. The NGS platforms were originally developed to generate DNA sequence information from whole genomes or large environmental samples. For example, the whole chloroplast sequence of *Ceratophyllum demersum* was obtained by Moore et al. [86] using the 454 Life Sciences sequencing platform and complete plastomes of 37 *Pinus* species were assembled by Parks et al. [87] on a multiplex Illumina sequencing platform. The NGS method was also useful to verify the contents of multiple ingredient herbal products by parallel sequencing their barcodes [88]. This method prevents overlaying sequence peaks as found in Sanger sequencing and therefore facilitates the sorting of DNA barcodes, and, consequently, the identification of mixed plant material [65,89]. NGS proved to be superior to Sanger sequencing in a comparative analysis that included 15 commercial plant material [65,89]. NGS proved to be superior to Sanger sequencing, but excipient DNA was detected in two supplements instead. The NGS method using the ITS2 locus they could identify the DNA source(s). Single peaks (B) with a low background are desired to determine a sample’s DNA sequence. (Color figure available online only.)

The trade sample identities were validated and confirmed by these species-specific DNA barcodes [17]. Ginseng samples (*Panax* spp.) were tested for their authenticity by both Zuo et al. [93] and Wallace et al. [49]. Zuo et al. [93] analyzed DNA from fresh leaves of 95 ginseng samples, representing all of the species in the genus *Panax*. The analysis showed that the combination of *psbA-trnH* and ITS was able to identify all of the species and clusters in the genus. Wallace et al. [49] tested 41 commercial ginseng samples (*raw materials* and finished products) in the North American market and found that the core barcodes *matK* and *rbcL* required additional data from ITS for successful species identification. Stocek et al. [94] analyzed commercial tea samples (*Camellia sinensis*) with 90% success identification rates using *rbcL* and *matK* barcode loci and reported 33% adulterations in herbal teas. Black cohosh (*Actaea racemosa*) samples were analyzed by Baker et al. [95], and with a mini-barcode approach using the *matK* locus they could identify 75% of the tested samples. They found that 25% of the tested black cohosh samples were adulterated. Little and Jeanson [71] used mini-barcodes from the *rbcL* and *matK* regions to authenticate saw palmetto (*S. repens*) herbal dietary supplements. The analysis of these tested supplements demonstrated that 85% contained saw palmetto and that 6% of the supplements contain related species (*Aecolorhaphaewrightii* or an unidentified species) that cannot be sold legally as herbal dietary supplements in the United States [70]. Similarly, a mini-barcode assay of ginkgo (*G. biloba*) dietary supplements by Little [51] revealed that of the 40 supplements tested, 83.8% contained identifiable *G. biloba* DNA, and six supplements (16.2%) contained fillers without any detectable *G. biloba* DNA. Substitution of raw herbal materials in

**Fig. 3** Electropherograms showing sequencing signals obtained with Sanger sequencing. The overlaying peaks (A) make it difficult to determine the real sequence. It is an indication that the sequenced template consists of mixed DNA. Additional steps (digital PCR or cloning) are recommended to identify the DNA source(s). Single peaks (B) with a low background are desired to determine a sample’s DNA sequence. (Color figure available online only.)

**Accuracy of DNA Barcoding Techniques for Authentication of Botanicals in Herbal Products**

Most of the botanical ingredient DNA barcoding studies published to date focused on the identification of a universal marker or suitable barcode locus/loci for herbal raw material authentication. Only a small number of published papers discussed the use of DNA barcoding for the identification and authentication of botanicals from finished herbal products and dietary supplements. Srirama et al. [17] reported an investigation in which 25 *Phyllanthus* samples used as raw herbal drugs in the Indian market were assessed for their authenticity. Their analysis revealed that six different species of *Phyllanthus* were available on the market based on morphological studies. Seventy-six percent of the market samples contained *Phyllanthus amarus* as the predominant species (>95%) and the remaining 24% included five different species, namely *P. debilis*, *P. fraternus*, *P. urinaria*, *P. maderaspatensis*, and *P. kozhikodanum*. Species-specific DNA barcode signatures were developed for the tested *Phyllanthus* species using the chloroplast DNA region *psbA-trnH*. The trade sample identities were validated and confirmed by these species-specific DNA barcodes [17]. Ginseng samples (*Panax* spp.) were tested for their authenticity by both Zuo et al. [93] and Wallace et al. [49]. Zuo et al. [93] analyzed DNA from fresh leaves of 95 ginseng samples, representing all of the species in the genus *Panax*. The analysis showed that the combination of *psbA-trnH* and ITS was able to identify all of the species and clusters in the genus. Wallace et al. [49] tested 41 commercial ginseng samples (*raw materials* and finished products) in the North American market and found that the core barcodes *matK* and *rbcL* required additional data from ITS for successful species identification. Stocek et al. [94] analyzed commercial tea samples (*Camellia sinensis*) with 90% success identification rates using *rbcL* and *matK* barcode loci and reported 33% adulterations in herbal teas. Black cohosh (*Actaea racemosa*) samples were analyzed by Baker et al. [95], and with a mini-barcode approach using the *matK* locus they could identify 75% of the tested samples. They found that 25% of the tested black cohosh samples were adulterated. Little and Jeanson [71] used mini-barcodes from the *rbcL* and *matK* regions to authenticate saw palmetto (*S. repens*) herbal dietary supplements. The analysis of these tested supplements demonstrated that 85% contained saw palmetto and that 6% of the supplements contain related species (*Aecolorhaphaewrightii* or an unidentified species) that cannot be legally sold as herbal dietary supplements in the United States [70]. Similarly, a mini-barcode assay of ginkgo (*G. biloba*) dietary supplements by Little [51] revealed that of the 40 supplements tested, 83.8% contained identifiable *G. biloba* DNA, and six supplements (16.2%) contained fillers without any detectable *G. biloba* DNA. Substitution of raw herbal materials in
local Indian markets was shown for Sida cordifolia [96] and in 50% of Cassia fistula and Senna spp. [97]. Palhares et al. [98] analyzed 257 dried or powdered samples from 8 medicinal plant species approved by the World Health Organization (WHO) for the production of herbal drugs sold in Brazilian markets. These included witch hazel (Hamamelis virginiana) leaves, chamomile (Matricaria recutita) flowers, espinheira sânta (Maytenus ilicifolia) leaves, guaco (Mikania glomerata) leaves, Asian ginseng (P. ginseng) roots, passion flower (Passiflora incarnata) leaves, boldo (Peumus boldus) leaves, and valerian (V. officinalis) roots. The DNA barcoding analysis using matK, rbcL, and ITS2 regions confirmed species belonging to the correct genus in 42% of the samples. For the remainder, results suggested that the level of substitutions might be as high as 71% [98], although some of this is due to the misapplication of the common name, e.g., 100% of the samples presented as P. ginseng were actually from the genus Pfaffia, also called “Brazilian ginseng.” Recently, Han et al. [47] investigated 1436 samples representing 295 medicinal species from 7 primary TCM markets in China. Their results indicated that of the 1260 samples, approximately 4.2% were identified as being adulterated and they suggested a regulatory platform based on DNA barcoding should be established for TCM market supervision. However, verification of the accuracy of the results using established compendial methods, e.g., those listed in the European Pharmacopoeia or the United States Pharmacopeia (USP), was not performed in any of these studies.

The first investigation into a larger set of diverse dietary supplements using DNA barcoding was published by Newmaster et al. in 2013 [99]. In the study, the authenticity of 44 (41 capsules, 2 powders, and 1 tablet) single ingredient herbal products from 12 companies was evaluated. The validity of the approach was evaluated by including a second set of samples grown from commercially available seeds in horticultural greenhouses, including all those plants listed on the herbal product labels and some closely related species. The authors of the study were able to recover DNA from 40 out of 44 samples. According to the results, 14 samples out of 40 were correctly labelled, 14 contained the correct species but included additional DNA from either another species or from an excitipient, and 12 contained only DNA from other species or excitipients. A number of concerns about the study were raised by DNA experts and members of the American Botanical Council [100], e.g., the results were not confirmed by orthogonal methods, such as chemical analysis, and the DNA barcoding method was not validated to the standard required of the industry. According to the authors, products contained solely crude powdered raw material; however, attempts to identify the contents using botanical microscopy were unsuccessful due to the lack of recognizable plant fragments (S. Newmaster personal communication, July 8, 2014). Also, although the products were manufactured and sold, the authors did not take into account the normal manufacturing process for commercial dietary supplement products. Another issue is the acceptable presence of other species. Monographs for herbal raw materials, such as those in the European Pharmacopoeia and the United States Pharmacopoeia, allow a certain amount (e.g., 2% in USP) of foreign organic matter. DNA from foreign organic matter such as other plant species could be accidentally introduced at any stage of processing, for example, at the time of collection of herbal plant material, during storage, drying, grinding, at various stages of the product manufacturing process, or during the analysis in the quality control laboratory (Fig. 4). Therefore, it is normal to detect DNA from additional species in crude herbal raw materials.

The results and methodology published by Newmaster et al. [99] prompted the New York State Attorney General (NYAG) to launch his own investigation into the quality of herbal dietary supplements using DNA barcoding (details about the method have not been made publicly available) [101, 102]. The results from this investigation suggested that out of 24 commercial products [labeled to contain Echinacea (Echinacea spp.), garlic (Allium sativum), ginkgo (G. biloba), ginseng (Panax spp.), saw palmetto (S. repens), St. John’s wort (Hypericum perforatum), or valerian (V. officinalis)] analyzed, only 5 contained DNA of the labeled species. The results led the NYAG to demand that the four retailers selling the supplements, GNC, Target, Walgreens, and Walmart, remove the products from their shelves [101]. The accuracy of the results was immediately questioned, mostly because the majority of the products were made from herbal extracts, where DNA was probably fragmented or degraded, and DNA barcoding (i.e., the amplification of the several hundred nucleotide long complete barcodes) has not been shown to provide useful results for these cases [103]. In addition, the investigation found DNA from Oryza, Allium, and Dracaena species in 19, 9, and 7 of the samples analyzed, respectively, strongly suggesting cross-contamination. Another puzzling result was the occurrence of saw palmetto DNA in one of the valerian samples, again raising the question about cross-contamination. And while the usefulness of full-length DNA barcoding in providing reliable information of what types of genera or species might be present in a sample containing crude or raw botanical materials is generally recognized, the full-length barcoding approach suffers from the misconception that DNA is always uniformly preserved and available [72, 104]. As mentioned before, DNA quality is affected by heat, freeze/thaw cycles, fungal or bacterial contamination, irradiation, and a number of chemicals. In addition, most of the DNA is typically removed during metabolite extraction, and any DNA that does remain in an herbal extract will usually be fragmented. Additional purification steps used in commercial extraction, such as column chromatography, may eliminate any remaining DNA altogether. Therefore, the relatively long genomic regions required for universal DNA barcoding are no longer present in most botanical extracts. According to Little [72], universal DNA barcoding rarely provides accurate species identification, since distinction among closely related species is difficult, often impossible, and that the diagnostic features examined may not be distinctive enough for a given species.

---

**Fig. 4** Various steps in the processing of plant material during which exogenous DNA may be introduced into a sample. (Color figure available online only.)
Method Validation

A critical step in ensuring reliable results from any analytical method is the method validation. Guidelines for the validation of methods to identify botanical ingredients have been published, e.g., by AOAC International [105]. The guidelines were written with chemical methods in mind, so transferring these to DNA barcoding may not be straightforward. The method of choice has to be able to distinguish the species of interest from its adulterants, which may include different plant parts from the same species. It also needs to be able to determine at what level of contamination the method can detect an adulterant. This is done by preparing mixtures of the target species with known amounts of adulterant. These requirements can usually be fulfilled with raw herbal material, except that DNA barcoding will not be able to determine the plant part from which the article is derived from. However, method validations become complicated when finished products are evaluated. DNA barcoding methods are most often “validated” by using fresh or dried raw plant material in order to determine if the chosen DNA barcode region can be amplified [72, 98, 99] and distinguished from related species. The usefulness of such validation in finished products is questionable since processing methods may alter or eliminate the DNA, excipients and secondary metabolites may interfere with DNA extraction and amplification, and the presence of DNA from excipients and fillers may lead to erroneous results. Therefore, the development of guidelines to address the method validation requirements for DNA barcoding, in particular with regard to finished products, is much needed and one of the larger issues to be resolved.

Limitations and Future Challenges of DNA Barcoding for Authentication of Herbal Products

The major limitations to DNA barcoding of herbal products are related to the quality of DNA, primer affinity, PCR amplification, and sequencing of amplicons. Plant DNA is a relatively stable molecule and can be easily extracted from fresh or dried plant material using simple DNA extraction methods. However, the manufacturing process of herbal products that involves extensive heat treatment, irradiation, distillation, filtration, UV light exposure, and/or supercritical fluid extraction leads to either complete removal of DNA or degradation of DNA into smaller fragments [52]. Hence, DNA barcoding is not feasible for processed herbal products such as extracts and tinctures in which the DNA is not present at all or highly degraded. Therefore, the stage at which DNA barcoding analysis should be performed is very important. For example, in the case of Asian ginseng (P. ginseng), the roots are often mixed or substituted with undeclared P. ginseng leaves. Roots and leaves both contain high levels of ginsenosides, but exhibit a different chemical profile. In such cases of substitution with the same species parts, DNA barcoding will not be able to detect the adulteration. Similarly, the approach will not detect if exhausted material is used, i.e., material where the putative active components have been removed previously via extraction. Likewise, if the prescribed plant is collected in the wrong season, the efficacy of the herbal material decreases and DNA barcoding will fail to detect the substitution with such low quality material. To overcome these limitations of DNA-based analysis, chemical profiling or other analytical chemistry methods should be adopted for authentication of herbal products. Thus, DNA barcoding should go hand-in-hand with chemical analysis and macroscopic and/or microscopic evaluation to tackle the adulteration problems prevailing in the herbal industry. Another limitation of DNA barcoding is the universal primer set affinity to the excipient DNA or DNA of another adulterating or substituted species. Generally, the excipients are added after the processing of the herbal material is completed and, hence, the filler’s DNA remains intact. Consequently, DNA barcode primers are likely to preferentially amplify DNA from excipients, possibly yielding a false negative result for the herbal species that was intended to be detected. PCR bias can partly be overcome by NGS, digital PCR [51, 71], or by cloning specific PCR products into vectors and sequencing the amplicon. Moreover, it is preferable to design species/genus-specific primers for short barcode sequences (< 200 bp), so-called mini-barcodes, for successful amplification of species-specific barcodes, rather than using the longer universal barcodes. The shortcomings of Sanger-based sequencing methods can be overcome by NGS techniques for sequencing of herbal products containing multiple plant species or admixtures. DNA barcoding, like any other analytical method, has its own limitations, and hence, we view it as an additional identification tool for the authentication of herbal products in combination with other established methods.

One of the major challenges of DNA barcoding for authentication of herbal products is the lack of reference libraries and voucher specimens linked to reference DNA sequences in the GenBank database. The creation of a mini-barcode reference library or HerbBOL (also suggested by Mishra et al. [107]), containing all of the authentic reference barcode sequences linked to the respective taxonomically validated herbarium vouchers, would be a useful tool to ensure access to reliable DNA barcodes. The use of a barcode reference library could provide a basis for using DNA technologies as a cGMP-compliant approach for the authentication of herbal products and dietary supplements in the future.

Conclusions

The DNA sequencing technologies to identify medicinal plant species in herbal products and dietary supplements is a highly reliable and promising tool under specific conditions, such as the correct stage of analysis when the DNA could be detected, primer affinity for successful PCR amplification, and absence of contaminating DNA. The detection of adulteration of botanical ingredients could be improved if DNA barcoding is routinely and appropriately used for authentication of herbal materials. It is important to apply the most appropriate method to efficiently detect and identify the analyzed raw or processed material.
However, the inherent limitations of the DNA barcoding methods make it unsuitable as a stand-alone tool for identifying and authenticating the herbal plant species. Therefore, we advocate the addition of DNA barcoding to the other existing analytical methods for authentication of botanical ingredients in herbal medicines and dietary supplements.

Supporting information
Reference materials for the DNA barcoding of H. perforatum L. (St. John’s wort) are available as Supporting Information.

Acknowledgements
This research was funded, in part, by the Food and Drug Administration grant no. 1U01FD004246–05. We thank Jon Parcher for his revision of the manuscript and suggestions.

Conflict of Interest
The authors declare no conflicts of interest.

References
1 Hebert PDN, Cywinska A, Ball SL, de Waard JR. Biological identifications through DNA barcodes. Proc R Soc Lond B 2003; 270: 313–321
31 Murray M, Thompson WF. Rapid isolation of high molecular weight plant DNA. Nucleic Acid Res 1980; 8: 4321–4325
38 Xin Z, Chen J. A high throughput DNA extraction method with high yield and quality. Plant Methods 2012; 8: 26–32
genomic DNA extraction protocol for large scale genetic analyses of plant biological systems. Plant Gene 2015; 2: 43–45
40 Porebski S, Bailey LG, Baum BR. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol Biol Report 1997; 15: 8–15
43 Sahu SK, Thangaraj M, Kathiresan K. DNA extraction protocol for plants with high levels of secondary metabolites and polysaccharides without using liquid nitrogen and phenol. ISRN Mol Biol 2012; 2012: 1–6
45 Peterson DG, Boehm KS, Stack SM. Isolation of milligram quantities of nuclear DNA from tomato (Lycopersicon esculentum), a plant containing high levels of polyphenolic compounds. Plant Mol Biol Rep 1997; 15: 148–153
48 Kitterman FRH, Shattuck VI. An effective method of DNA isolation from the mature leaves of Gossypium species that contain large amounts of phenolic terpenoids and tannins. Prep Biochem 1983; 13: 347–359
51 Little DP. Authentication of Ginkgo biloba herbal dietary supplements using DNA barcoding. Genome 2014; 57: 513–516
64 Sucher NJ, Hennell JR, Carles MC. DNA fingerprinting, DNA barcoding, and next generation sequencing technology in plants. Methods Mol Biol 2012; 862: 13–22


