DNA-Based Authentication of Botanicals and Plant-Derived Dietary Supplements: Where Have We Been and Where Are We Going?

Planta Med 2015; 81: 687–695

This electronic reprint is provided for non-commercial and personal use only: this reprint may be forwarded to individual colleagues or may be used on the author’s homepage. This reprint is not provided for distribution in repositories, including social and scientific networks and platforms.
DNA-Based Authentication of Botanicals and Plant-Derived Dietary Supplements: Where Have We Been and Where Are We Going?

Abstract

Herbal medicines and botanicals have long been used as sole or additional medical aids worldwide. Currently, billions of dollars are spent on botanicals and related products, but minimal regulation exists regarding their purity, integrity, and efficacy. Cases of adulteration and contamination have led to severe illness and even death in some cases. Identifying the plant material in botanicals and phytomedicines using organoleptic means or through microscopic observation of plant parts is not trivial, and plants are often misidentified. Recently, DNA-based methods have been applied to these products because DNA is not changed by growth conditions unlike the chemical constituents of many active pharmaceutical agents. In recent years, DNA barcoding methods, which are used to identify species diversity in the Tree of Life, have been also applied to botanicals and plant-derived dietary supplements. In this review, we recount the history of DNA-based methods for identification of botanicals and discuss some of the difficulties in defining a specific barcode or codes to use. In addition, we describe how next generation sequencing technologies have enabled new techniques that can be applied to identifying these products with greater authority and resolution. Lastly, we present case histories where dietary supplements, decoctions, and other products have been shown to contain materials other than the main ingredient stipulated on the label. We conclude that there is a fundamental need for greater quality control in this industry, which if not self-imposed, that may result from legislation.

Supporting information available online at http://www.thieme-connect.de/products

Introduction

Plants have been used for their medicinal applications for millennia, and continue to be used. The World Health Organization (WHO) and others have reported that millions of people use medicinal plants worldwide in both developing and developed countries. In developing countries, about 80% of the population relies on plants for their primary healthcare [1–3]. In China, 30–50% of the medicines consumed are derived from plants. In Germany, 90% of the people use natural medicines at some time during their life, whereas in other European countries, over 50% of the population has done so [1–3]. More than half of all adults in the US use dietary supplements, which are used by healthy people to add to their diets [4,5]. The most recent data indicate that in 2012, 17.9% of all US adults used botanical supplements [6] and that 70% of the Canadian population tried them at least once [2,3]. The global herbal market has expanded even during the recent economic recession. Furthermore, people without medical insurance are more likely than most people to use botanicals [4]. Billions of dollars are being spent on products used as dietary supplements or phytomedicines, either in capsules containing dried plant material or as extracts. The global market for these products was approximately US$60 billion in 2000 [2] and is expected to increase dramatically, reaching US $107 billion by the year 2017 [7]. Phytomedicines represent a major share of the pharmaceutical market in the US and Europe. On a global basis, Europe is the largest market for herbal supple-
As the medicinal plant market has grown, incidents of contamination (with insecticides, pesticides, synthetic drugs, heavy metals) or adulteration (substituting one plant for another either purposefully or through misidentification) are continually reported, resulting in increased concerns about the safety, effectiveness, and quality of herbal products. In addition, the safety and efficacy of these products are directly linked to quality control of the raw materials from which all herbal preparations start [8–10] (Fig. 1). To date, not all starting materials are routinely tested for authenticity or adulteration. Guidelines for the registration of herbal products have been published by the WHO [3,9], and a number of countries have developed their own guidelines to ensure the quality control of herbal medicines, including the importance of adhering to good agricultural and collection practices (GACP), good manufacturing practices (GMP), and good laboratory practices (GLP) (see other WHO publications). According to the WHO [3], of the 92 countries who responded to a survey on regulating herbal medicine, 65% of them have laws and regulations. Seventy-three percent of 103 responding countries allow herbal medicines to be sold with claims; the most frequent are medical and health claims [3]. Generally, countries in the WHO European region have a high level of commitment to the quality, safety, and effectiveness of herbal medicines. In the US, the Food and Drug Administration (FDA) does not regulate these products because they are exempt. In its place, Congress approved the Dietary Supplement Health and Education Act (DSHEA) in 1994, which classified herbal products as dietary supplements [4,10] and stipulated that supplements can be sold without prescription and do not have to be proven efficacious. DSHEA also prohibits manufacturers and distributors from marketing products that are adulterated or misbranded, and dietary supplements must be produced according to GMP. In 2003, the FDA proposed the use of current good manufacturing practices (CGMP) for further quality control of these products [4,10]. Nevertheless, the potential for fraud and contamination has not diminished over the years, and so other means have had to be developed to analyze the starting plant material, whether fresh or dry, and the products derived from them. The products are assumed to be safe until proven otherwise, and in the US, the burden of proof rests with the Federal government, not with the manufacturers or distributors. Botanicals are part of a complementary and alternative medicine portfolio used worldwide, but not without risk. For example, herbal medicines are responsible for 9% of drug-induced liver injury, and in Asian countries, this figure ranges from 19% to 63% [11]. Thus, manufacturers, distributors, and the Federal government have a critical need to identify botanical adulterants efficiently and confidently.

**DNA Markers for Identifying Plant Species**
Numerous articles and pharmacopeia have been published about the methods for demonstrating the authenticity and purity of herbal products or botanicals [12]. Most products are purchased by consumers as powders in capsules or as extracts, but manufacturers often purchase the unprocessed plants from suppliers (Fig. 1C,D). Adulterants may be detected at this point using classical botanical methods based on morphological and anatomical features of the plants [13]. For example, microscopic methods can be used to examine the plants and plant powders for characteristic structures such as stomata, hairs, scales, xylem cells, etc. However, today few people have sufficient training in plant anatomy or classical plant taxonomy to perform these types of analyses accurately, and so these more traditional methods have become less frequently employed, although most pharmacopeias, worldwide, include them. Moreover, for extracts or for herbal products for which the active pharmaceutical ingredient (API) is known, techniques such as GC/MS, HPLC, 2D-NMR, or LC/MS/MS are not only used to verify the presence, but also the quantity of
the API(s) [10]. Nevertheless, because many dietary supplements and botanicals are composed of dried or powdered plant material and the API may not be known, authentication in the last few years has been mostly based on the analysis of DNA markers. Since our earlier review on the topic of molecular methods for the authentication of botanicals and detection of potential contaminants and adulterants [14], the number of review articles addressing this topic has marked increased [15–18]. Many of the techniques described earlier, which we referred to as the use of DNA markers to validate and authenticate plant material [14], have been superseded by DNA barcoding. Some of these earlier described methods are summarized in Table 1S, Supporting Information. Hebert et al. [19] were the first to use the term “DNA-barcoding” for short, orthologous DNA sequences that discriminate organisms comprising the Tree of Life – plants, animals, fungi, bacteria, and archaea – to the species level. The ideal locus for barcoding should possess high universality yet discriminate among taxa that are not only separated by hundreds of millions of years of evolutionary divergence, but also taxa separated by a few million years [20, 21]. For animals, the universal animal DNA barcode, a mitochondrial gene encoding cytochrome c oxidase 1 (CO1), is widely used to differentiate diverse animal species [22, 23]. Relative to nuclear genes, animal mtDNA exhibits rapid evolution, limited recombination, lacks introns, and has a high copy number, all of which makes it suitable as a single-locus barcode for most animal groups.

The extent of species diversity in the Tree of Life monitored by plant biologists is huge, spanning some 490 million years since the divergence of vascular plants from mosses. However, no single DNA sequence has so far been found to be ideal for plant species identification. In plants, mitochondrial genes evolve at about one-third of the rate of chloroplast genes [24] and consequently exhibit very little variation [25], so other candidate genes have been studied, especially those from the chloroplast [16]. Several plastid loci, particularly the psbA-trnH spacer, rpoC1, matK, rbcL, and combinations such as matK + rbcL [17, 20], have been widely utilized to differentiate a broad range of species of angiosperms, gymnosperms, and cryptograms [20,25]. Universal primers can be used for the plastid DNA sequences described above, and although many of these can discriminate down to the species level, they may not be able to do so for all phyla. For example, matK is not useful for amplifying DNA from many gymnosperms and is extremely ineffective for cryptogams. Also, at this time, the expense, expertise, and time required to analyze and differentiate plants among so many phyla using multiple genes is greater than what is needed for single-locus barcode analysis. Accordingly, nuclear as well as plastid loci have also been considered barcodes, but again the requirement for universal primers to amplify PCR products has restricted the choice to a limited number of DNA sequences. The most commonly used barcodes are the internal transcribed spacer (ITS) of ribosomal DNA, parts of the 5S, and the 5.8S region [26] (Fig. 2). Nonetheless, the use of one or two loci for which universal primers can be designed as barcodes has resulted in incredible progress towards understanding plant relationships.

The aim of systematics is to reconstruct the true phylogenetic or evolutionary relationships among taxa. The fact that different phylogenetic trees, all with strong support, can be constructed using different single genes on the same taxa is a strong argument against the single-locus barcode method. Indeed, true phylogenetic relationships should be constructed using large data sets comprised of multiple loci of independent genes, each evolving at different rates [27,28]. In vertebrates, either ultra-conserved sequences [29,30] or highly conserved sequences [28] were very effective in resolving clades across a diversity of evolutionary timescales. The key to the success of this approach, called anchored hybrid enrichment [28], is that probes are designed to target conserved regions and their flanking variable regions, which allows discrimination across diverse evolutionary scales.

Similarly, plant scientists are working toward the use of multiple loci as employed for identifying bacterial species [multilocus sequence analysis (MLSA)], which is performed by concatenating various housekeeping gene sequences. This is a very powerful tool for distinguishing microbial species. Such an approach has not been widely undertaken for plants, not only because of the paucity of utilizeable nuclear loci, but also because plastid gene barcodes, which can be biparentally inherited, often show too much interspecific variation. In any case, the use of multiple plastid gene sequences yields results that are similar to MLSA for bacteria. Also, combinations of nuclear and plastid loci, e.g., ITS2 + trnH-psbA, often discriminate species better than many plastid barcodes used alone [25]. However, adding plastid DNA barcodes to the analysis does not always significantly enhance the resolution given by ITS2 alone [20,21].

Targeted Enrichment Approaches

Next generation sequencing (NGS) platforms eliminate many of the problems (lack of knowledge of genome size, genome sequence, etc.) faced by genetic/genomic plant community members who work outside the major model and crop species and, hence, have quickly become a powerful phylogenomic tool. Target enrichment is an alternative approach to the restriction enzyme sequencing approaches discussed below (i.e. RAD, GBS) and could be easily applied to medicinal plants for which genomic resources are largely not available. This method requires identification of primers that target conserved orthologous se-
quences (COS), an approach similar to sequence enrichment of ultraconserved elements in vertebrates [28, 29]. Once COS are identified, bait probes are designed and the source (sample) DNA hybridized, pooled, and sequenced on an NGS platform. This approach has been used effectively in plants in closely related taxa as well as plants exhibiting wide evolutionary distances. Target enrichment would be valuable in detecting adulterants from the bona-fide product because herbal medicines often include plants derived from closely related taxa. Indeed, many botanicals have sister taxa whose geographic distributions overlap and may be difficult to discriminate using classical botanical features. Additionally, although some botanicals may consist of related species, only one or a few taxa have bioactivity. The *Echinacea* species complex is a good example where confusion over taxonomy and differences in activity among the species exist [31–33]. Similarly, *Ganoderma lucidum*, a popular mushroom used in herbal medicine, is very difficult to distinguish from other *Ganoderma* species using morphological criteria, and mitochondrial data do not satisfactorily circumscribe the species complex. However, it is important to distinguish among them because several *Ganoderma* species are cytotoxic to drug-sensitive and drug-resistant cells of small cell lung cancer, whereas most *Ganoderma* species are not [34].

Targeted enrichment approaches include both chloro-plast and nuclear genomes. The advantage of the chloroplast genome is that it has long been used in systematics and is a small genome with highly conserved regions. NGS chloroplast enrichment studies effectively resolved 107 *Pinus* species [35] and more recently, it has been developed to work with a broad range of taxa. By developing probes from conserved regions from a wide diversity of taxa, the universality of sequences and power of discrimination can be increased. Stull et al. [36] utilized plastid genomes from 22 eudicotyledonous species representing about 75% of angiosperm diversity to design 55,000 bait probes capable of capturing the breadth of eudicot diversity. Although coverage depth varied considerably across the 24 taxa assessed, the probe set recovered essentially complete plastomes of all taxa. These studies indicate that plastid probe enrichment, particularly probes designed to target highly conserved plastome regions, are suitable for assessment across wide phylogenetic distances [36, 37].

**Restriction Enzyme Sequencing Approaches**

Before the development of NGS platforms, biologists were hampered by the lack of sequence information for developing probes to explore the genetics and genomics of their organism. The amplified fragment length polymorphism ( AFLP) method [38] for DNA fingerprinting has been employed across a vast array of taxa across the Tree of Life. The AFLP method meets the CBOL criteria [20] for universality because primers are based on restriction sites and can be applied to any taxon with minimal developmental effort and without any prior sequence knowledge needed. The AFLP method detects single nucleotide polymorphisms (SNPs) between taxa, and this restriction enzyme approach has been adapted to NGS platforms. Two methods have been developed and presently are used primarily in the plant breeding community to detect SNPs among closely related taxa and interrelated individuals. These include restriction site-associated DNA sequencing (RAD or RADseq) and genotyping by sequencing (GBS). RAD sequencing [39] begins with the digestion of genomic DNA with a restriction enzyme, ligation of the “P1” primer adaptor (consisting of a forward amplification site, an Illumina sequencing primer site, and a barcode), to the restriction site, shearing the ligated DNA to size-fractionate it, ligation of the “P2” to the reverse complement of the “P1” primer site, and then selectively amplifying the RAD tags. On the other hand, GBS utilizes restriction enzymes that leave 2–3 base-pair overhangs and avoids cutting in the repetitive regions of a genome [40]. Barcode and common adapters are ligated to the DNA-digested sticky ends, size-selected, amplified by PCR, cleaned-up, and quality controlled, and finally subjected to multiplexed sequencing on the Illumina platform. Neither RAD nor GBS requires a reference genome, but both require bioinformatics computers or cloud computing and bioinformatics skills.

Presently, and likely for several more years, whole genome sequencing will be impractical for identifying and discriminating among botanical samples because sequencing to a sufficient depth can be cost-prohibitive and de novo genome assembly is not trivial. Instead, NGS applied to reduced genomes is a viable alternative that is easily implemented. Several genome reduction methods are currently employed. These include methods designed to reduce the repetitive regions through 1) the use of a thermo stable duplex-specific nuclease [41]; 2) digestion with methylation-sensitive restriction enzymes [40]; or 3) a single restriction enzyme ligated to an adaptor (RESCAN, [42]). Perhaps ironically, the high copy number fraction, including ribosomal and cpDNA that have been extensively used in molecular systematic studies, are largely excluded by these methods.

**Whole Chloroplast Sequencing**

A number of plant biologists have started to use whole chloroplast (cp) genome sequencing, also known as super-barcoding, to analyze phylogenetic relationships among some hard-to-resolve plant groups [43]. Some plants of medicinal interest have already been examined this way [44, 45], but mainly to assess their phylogeny. Cp genome sequencing is akin to what several bacterial taxonomists have proposed should be done to provide better insight into the separation of one bacterial species from another, especially for problematic families such as the Burkholderiaceae [46]. The vast number of finished or draft bacterial genomes in various databases makes this approach possible. The number of cp genomes in GenBank has significantly increased from 1986 to 2012 [17]. However, unlike the case for bacterial genomes, the number of available plastomes, approximately 411, is limited. Decreasing sequencing costs and NGS technologies will make this strategy more popular and common. The targeted enrichment strategy for plastids [36], discussed above, which recovers essentially whole cp genomes, establishes this as a viable strategy and provides bait sequences for a wide diversity of taxa. It is likely that the super barcoding/targeted enrichment approach using NGS will be useful for botanical authentication and validation. Also, urgently needed are the bioinformatics resources and training to make this information available and utilisable.
Authenticating and Validating Botanicals, Herbals, and Dietary Supplements: Considerations

The discrimination required for the identification of plants used in herbal medicines and dietary supplements may not have to be as detailed as the evidence needed for elucidating the Tree of Life. For one, species boundaries for many plant families are still uncertain because of incomplete data, and thus more analysis is needed to fill in the gaps. For another, herbal medicines do not originate from as many phyla as the plants that comprise the Tree of Life. In the cases where the botanical is a gymnosperm, e.g., ginkgo, species-specific primers can be employed to identify the presence or absence of the plant in a dietary supplement or herbal medicine [14]. Also, the major goals for DNA barcoding botanicals and related products are to determine their authenticity (the identity of plant species that is sold or used for an extract) and purity (whether or not other species are substituted or accidently added to the starting material). Companies and vendors want to identify the plants used as starting products for herbal medicines, but require simple methods to authenticate and validate them. Similarly, clinical researchers need to verify that the dietary supplements or botanicals they are studying contain the material specified by the label. Moreover, due to the shear bulk and diversity of material used in the manufacturing of these products as well as the fact that people who are not trained in plant biology will perform much of the testing, the methods have to be cheap, fast, and easy-to-use. Hence, the reasons for using a limited number of barcodes, the universality of the primers to amplify them, and reliable and reproducible methods of DNA isolation are clear. However, because DNA barcoding of botanicals requires comparison with closely related species, the need for phylogenetic analysis on a larger scale still stands [47]. The lack of phylogenies and knowledge of an herbal product’s relationship to other species severely hinders accurate authentication of plants used as botanicals.

Because plant materials used for herbal products contain numerous phenolics, flavonoids, and other secondary compounds, which comprise the vast majority of extracts sold to the public, difficulties arise with isolating high quality DNA for authentication. Also, the starting plant material may have been dried in the sun or heated in such a way that any isolated DNA is degraded. In cases where the DNA is very fragmented, nuclear loci are probably better for distinguishing one plant species from another and potentially from adulterants, although opinions exist on both sides about the robustness of ITS2 in amplifying DNA from herbarium specimens [25;47]. The difference in results may depend on the quality of preservation and the age of the herbarium specimens (Fig. 1B) [25]; D.F.C. Moraes and A.M. Hirsch, unpubl.). In any case, homemade methods have been used to repair degraded DNA from dietary supplements [48], and today kits can be purchased for this purpose. Also, NGS platforms with their short reads, sequence depth, and accuracy in base calling, particularly with paired-end reads, have significantly diminished the impact of degraded DNA in obtaining useful sequence information.

Application of DNA Methods to Botanicals

Early on, Mihalov et al. [49] identified two different types of ginseng by amplifying a 650-bp ITS fragment and the plastid sequence for rbcl, and Lau et al. [50] employed ITS2 primers to authenticate medicinal Dendrobium species. LeRoy et al. [48] utilized ITS1 as well as the same 650-bp ITS region used by [49] to differentiate two different legume dietary supplements obtained from both fresh tissues and commercial samples. A problem, however, with using the larger ITS region is that it may pick up divergent paralogous ITS sequences that exist in some plants [51]; many of these may be highly polymorphic [52,53]. For example, the presence of multiple copies of ITS made the authentication of a medically useful plant such as green tea challenging [54]. Moreover, the possibility exists that some ITS sequences may be derived from fungal contaminants or endophytes. The presence of these “extra” ITS regions in the starting material or within the powdered dietary supplements and whether they are fungal-derived must be addressed. Fungal endophytes frequently contribute to the phytomedicinal profile of a plant [55], but contamination by unwanted fungi may lead to poor quality starting material due to disease and damage. Some of the problems of divergent paralogous ITS sequences may be resolved by using ITS2 [26,56–60]. PCR products derived from ITS2 are generally small in size (100–400 bp) and thus are useful for the frequently fragmented DNA that is isolated from dried plant material.

Information regarding differences in ITS sequences can also be resolved through a technique known as single strand conformational polymorphism (SSCP). SSCP was originally developed for the analysis of allelic differences in humans [61] and has been widely applied to other studies that require differentiation between DNA sequences, particularly the microbial community structure [62], SSCP has also been shown to be useful for the differentiation of plants, and the technique can be used to authenticate botanicals and identify the presence of contaminating plant material. Similar to the other types of barcoding described earlier, SSCP discriminates between genotypes by taking advantage of the polymorphism present between species in specific DNA markers. In SSCP, the DNA marker is amplified by PCR, denatured or digested to generate single-stranded DNA (ssDNA), cooled to facilitate the formation of secondary structure, and then run on an acrylamide gel. The secondary structure of the ssDNA is determined by the nucleotide sequence; therefore, the variation found between species in certain markers will result in differences in the DNA’s migration on the gel. SSCP can be used to authenticate products by comparing the migration of the ssDNA to that of an authenticated sample. Furthermore, the existence of contaminants can be detected by the presence of bands that do not migrate as expected. Bands on the gel can be subsequently excised and sequenced, which allows an additional level of authentication and the identification of the contaminating material by barcoding.

SSCP was first used in the study of botanicals by Kojoma et al. [63] to differentiate between three species of cinnamon using the trnl-trnf IGS chloroplast region and the trnl intron. Lum and Hirsch [57] found the ITS2 region more useful than the ITS1 for this procedure and also utilized matK to differentiate several plants, including two closely related species of licorice, from each other. They further applied the method to commercial samples of dried plant material and confirmed by sequencing that the appropriately sized bands were of the products specified and that any additional bands present were due to contaminating plant material [57]. SSCP in combination with capillary electrophoresis, which eliminates the need for polyacrylamide gels, was used by Wu et al. [64] to authenticate olive oil. These authors found that the rbcl gene could discriminate between seven oil plants.
and that SSCP analysis could be used directly on oil samples to authenticate their content [64].

High-Resolution Melting

High-resolution melting (HRM) is a method developed to distinguish between sequence variants amplified by PCR and has been applied to a wide variety of tasks. These include detecting adulterants of botanicals and foods, assessing methylation levels and allele-specific gene expression, and variant detection in marker-assisted crop breeding [65–71]. This simple technique is suitable for applications in which primers can be developed that target genes or genomic regions for which sequence variation exists between the two samples, e.g., the bona fide botanical and the contaminant. To obtain the melting curve, fluorescence is plotted as a function of temperature as the thermal cyclers temperature increases and the double-stranded DNA dissociates [65]. The shape and melting temperature (Tm) of the curve are functions of the GC/AT ratio of the amplified region; the Tm of a melting domain increases with greater GC content. Sequences with Tm of less than 2°C can easily be distinguished [65], and the shape of the melting curve can discriminate between two amplicons with a Tm of less than 0.3°C [68]. High-resolution instruments have the capability of distinguishing between amplicons with a Tm of 0.0°C [72]. In genotyping applications, deletions/insertions are easily distinguished because they have large effects on melting temperatures, whereas heterozygotes are recognized by changes in the shape of the melting curves. Successfully differentiating between homozygotes with different base substitutions (SNPs), however, is dependent on the nature of the substitution, precision of the instrument, and melting analysis software [72].

HRM has capitalized on the widespread use of barcodes and conserved sequences to distinguish among the adulterants of a variety of foods including olive oil [70] and legumes [73–75]; the amount of the adulterant is typically detected to a level of 1% of the admixture. When applying HRM to botanicals, primer sets can be designed for detecting between unknown or known contaminants [68]. In the former case, a universal barcode can be designed to detect the unknown adulterants, while a sequence-specific primer pair can be designed if the botanical is commonly contaminated with a known species. Although the development of the primers and determination of the level of sensitivity to the adulterant requires up-front development to design primers for a given taxa, once established, this method has the advantage of being quick and requires no subsequent sequencing. Perhaps surprisingly, HRM has only been applied sparingly to detect adulterants in botanicals. High-resolution melting was used to identify Helleborus niger L. by targeting chloroplast trnL-trnF and matK regions [76]. Although data are limited, assays using universal primers appear to be less sensitive in detecting adulterants than those that target specific adulterant taxa. Assays using universal primers targeting the matK gene detected a level of 0.1% contamination in an admixture of H. niger and an unknown species, but by using primers specific to the common adulterant Veratrum nigrum, the assay detected a level of 0.0005% contamination of the H. niger/V. nigrum admixture [76].

Cases of Possible Adulteration or Substitution and the Need for Databases of Medicinal Plants

Several recent studies on medicinal plants have addressed the issue of identification, particularly as it refers to commercial products, and the need to increase the number of DNA sequences available in genome databases for commercially important plants. Some of these studies are summarized in Table 25. Supporting Information. At the present time, the number of DNA sequences available for herbal and botanical products is insufficient. For example, a study of herbal teas, which are classified as food and not as dietary supplements, showed that although rbcl or matK barcodes were available for most of the teas, many of the labeled ingredients were not detected, whereas other materials that were not listed on the label were found [77]. The authors of this report queried both GenBank (NCBI) and the BOLD [78] databases and found it difficult to make a successful match. They attributed their lack of success to the limited information in the databases with respect to the ingredients listed on the labels. Another study on saw palmetto dietary supplements showed that 6% of the samples were either mislabeled or misidentified [79]. The supplements contained DNA from related palms, which cannot be legally sold in the US. Also, Wallace et al. [80] studied both plant and animal commercial products and utilized both GenBank and a local barcode library to identify some of the species. They found that Korean ginseng frequently was identified as American ginseng. Thus, an important issue for barcoding is the comprehensiveness and accessibility of the barcode databases. Currently, the most common sources for matching plant names with DNA barcodes is either GenBank or BOLD. Although GenBank contains a large number of DNA sequences from a wide variety of plants, the database is self-archived and, hence, sequence errors are common and often persist. BOLD at this point in time does not include a large enough number of DNA sequences for commercially important plants. However, it has the advantage of having voucher specimens and storing supporting information. Other databases in development are a Medicinal Materials DNA Barcode Database (MMDBD) [81] and one being produced by the China Barcode of Life group [59]. The lack of adequate database support influences numerous studies, even when using a high-throughput analysis (HTS) based on Second Generation sequencing technology. One such study analyzed the components of traditional Chinese medicine (TCM) that consisted of either plant or animal material or both [82]. Although HTS may be more efficient than many of the DNA-based methods described herein and offers a deeper coverage of more samples, the poor state of much of the material used for TCM in particular and phytomedicines in general results in highly degraded DNA. Thus, in the HTS study, high-quality DNA was isolated from only 54% (15/28) of the samples analyzed [82], and it was not clear whether all of the components of the samples were actually identified. In addition, for these samples, the materials making up some of these medicines were on the CITES (Convention on International Trade in Endangered Species) list or contained plants that were toxic or potentially allergenic (Table 25, Supporting Information). Such findings bring up serious concerns with regard to both legality and safety. Despite the fact that mislabeling (or including unlabeled material) has long been an issue that industry has been aware of, another recent study points out that these issues still persist. Newman et al. [83] used rbcl and ITS2 barcodes to analyze 44 different herbal products as well as 50 leaf samples corresponding to the contents of the dietary supplements examined. Although DNA
barcodes were recovered from 91% of the herbal products, the authors were unable to authenticate more than 50% of them. In addition, 33% of the samples contained products not listed on the labels. The authors concluded that most of the herbal products were of poor quality, containing ingredients not on the label or substitutions, contaminants, or fillers. Rather than using GenBank or BOLD to identify the herbal products, the authors generated their own barcode library, the sequences of which were later deposited in BOLD and GenBank. At the time of analysis, too few commercial plant DNA sequences existed in BOLD, making accurate identifications potentially difficult. A library of vouched specimens was also generated. This research [83] created a lot of attention, including being described in the New York Times and other media sources, partly because of the large number of samples that were shown not to contain the labeled herbal product [84]. The study also generated a strong negative response from a group that focuses on botanicals and was criticized on several issues [12]. However, based in part on publication [83], the attorney general of New York State, in early 2015, called for the removal of certain herbal supplements from retail pharmacies, as reported in the New York Times, because DNA barcoding showed that most of the tested products did not contain the herbs listed on their labels [85]. What is not clear, however, was if these products were actually supposed to consist of dried plant material or were they extracts? If the latter, many of these are added to an inert filler, which is often derived from plant material. As mentioned earlier in this review, many analytical methods such as GC/MS, HPLC, 2D-NMR, or LC/MS/MS are used to deduce the presence or amount of an active pharmaceutical ingredient (API) within an herbal supplement.

In any case, DNA barcoding, and very likely NGS-facilitated barcoding, is here to stay and currently, no other standardized methods, high-throughput or otherwise, exist to determine the authenticity of supplements that contain plant material to a 100% level. It is our strong recommendation that reliable information resources become more freely available not only for preparing high-quality DNA, but also for enabling accurate identification, authentication, and determination of the purity of the materials used to prepare dietary supplements and herbal medications and perhaps even more importantly, to test them rigorously before they are packaged into capsules or decoctions. Moreover, it is critical that databases containing DNA sequences of plants used as botanicals, phytomedicines, and dietary supplements be curated and supported by voucher specimens and ancillary materials. The lack of sufficient sequence information, resources, and back-up material for botanicals and medicinal plants in public databases remains a significant problem.

At this time, it is not clear how quickly industry will take on the task of authenticating the starting materials or the products it sells. Concern about the integrity of herbal products is on the rise, and if industry does not take the lead on testing botanicals and dietary supplements, who will? DSHEA took the dietary supplement industry away from overview by the FDA in the US, and it seems unlikely that this will change [86]. In the meantime, we are dealing with an imperfect world with regard to the integrity of herbal products. Although many reputable companies produce high-quality dietary supplements and herbal medicines, the opportunities for accidental harm or for the consumer wasting his or her money on an ineffective product still exist. The components of herbal products are not strictly regulated and as such consumers have very little knowledge of the potential adverse effects that some of them may have. An alarming range of adulterants and/or contaminants has been reported in the literature, with adverse effects ranging from mild to severe and in extreme cases, death [11,87–92]. With such consequences in mind, it behooves producers of botanicals, dietary supplements, and herbal medicines to invest more into research towards improving the quality and efficacy of these products. In addition, each ingredient in a dietary supplement or phytomedicine should be rigorously tested for safety before and after marketing [93] by multiple techniques following strict scientific guidelines. Moreover, industry and academics must partner on developing accurate and reliable assays for testing not only the integrity of the starting material for botanicals, but also on developing procedures for monitoring the purity of the finished products. Because much of the botanical/phytomedicine market is based on the perception of product quality and integrity, industry support is critical to resolving these problems to ensure the quality of the initial and final constituents. DNA-based authentication of the ingredients incorporated into these products is perhaps the most efficient and economical method by which adverse effects of adulterants and plant or microbial contaminants may be minimized. Indeed, this is an important first step in determining the safety of herbal products and the plant material from which they are derived.

Supporting information

Some examples of DNA-based techniques and specifically of DNA barcoding-based studies to evaluate the authenticity of herbal products are available as Supporting Information.

Acknowledgements

Research on botanicals and dietary supplements in the Hirsch lab is funded in part by the Shanbrom Family Foundation. D.F.C. Moraes thanks the CNPq (National Council for Scientific and Technological Development) and the Brazilian program “Science without Borders” for a research fellowship. We are grateful to Drs. Stefan J. Kirchanski and Maskit Maymon for their valuable comments on the manuscript and to Dr. Barry Prigge for his help with obtaining herbarium specimens.

Conflict of Interest

The authors have declared that they have no conflicting interests.

References

6 Peregoy JA, Clarke TC, Jones L, Stussman BJ, Nahin RL. Regional variation in use of complementary health approaches by U.S. adults. NCHS Data Brief 2014; 1–8
7 Global Industry Analysts, Inc. Herbal supplements and remedies – global strategic business report. Available at: http://www.prweb.com/re-
Moraes DFC et al. DNA-Based Authentication of... Planta Med 2015; 81: 687–695


59 China Plant BOL Group. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proc Natl Acad Sci U S A 2011; 108: 19641–19646


71 Bosmal I, Ganopoulos I, Madesis P, Tsaftaris A. Microsatellite and DNA barcode regions typing combined with High Resolution Melting (HRM) analysis for food forensic uses: a case study on lentils (Lens culinaris). Food Res Int 2012; 46: 141–147

72 Ganopoulos I, Madesis P, Darzenta N, Argiriou A, Tsaftaris A. BarCode high resolution melting (Bar-HRM) analysis for detection and quantification of PDO “Fava Santorinis” (Lathyrus clymenum) adulterants. Food Chem 2012; 133: 505–512

73 Madesis P, Ganopoulos I, Argiriou A, Tsaftaris A. The application of Bar-HRM (Barcode DNA-High Resolution Melting) analysis for authenticity testing and quantitative detection of bean crops (Leguminosae) without prior DNA purification. Food Control 2012; 25: 576–582


