5 The DNA Toolkit
A Practical User's Guide to Genetic Methods of Botanical Authentication

Danica T. Harbaugh Reynaud

CONTENTS

INTRODUCTION

Accurate botanical species authentication and detection of adulterants are arguably the greatest challenges for raw materials and natural product manufacturers today, especially given the requirement for 100% identity testing in the Food and Drug Administration (FDA)'s dietary supplement cGMPs (21 CFR part 111). After all, the safety and quality of foods, herbal supplements, and other natural products depends on using ingredients that are (1) the correct identity and (2) free of contaminants. Although there are a wide range of technologies developed for authentication and adulterant detection, from microscopy and organolepsis, high-performance thin layer chromatography (HPTLC) and thin layer chromatography (TLC), and Fourier transform infrared (FTIR) and near infrared (NIR) to name a few, still major challenges remain especially regarding differentiation between closely related species and identification of multiple species in a mixture. As a result, there is widespread adulteration in the natural products industry. The combination of an increase in economically motivated adulteration and number of ingredients being sold as extracts and blends rather than in their whole unprocessed form makes the task of identification and detection of contaminants more challenging than ever. As illustrated in

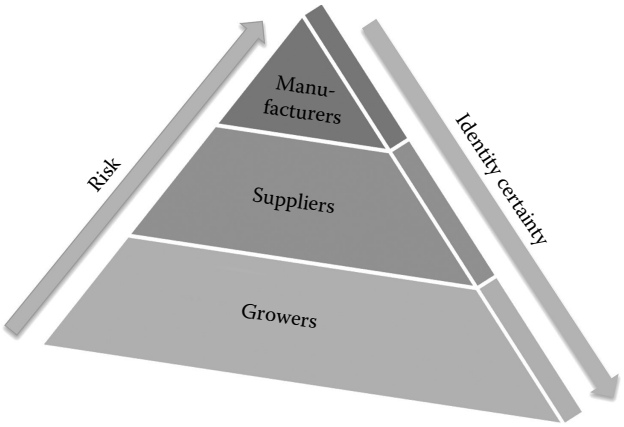

FIGURE 5.1 Risk-Identity Pyramid. The levels on the pyramid represent organizations along the supply chain for natural products, from the growers who farm the plants, to the suppliers who may process, blend, and/or distribute the plant materials, to the manufacturers who produce products using the plant ingredients. On the sides of the pyramid, the arrows indicate the inverse relationship between the certainty that these organizations have in identifying their materials.

the Risk-Identity Pyramid in Figure 5.1, there is an inverse relationship between our ability to confidently identify materials and the risks associated with manufacturing using adulterated ones, as materials are further processed and distributed through suppliers to manufacturers, they become more difficult to identify and are more prone to adulteration. Therefore, having a reliable toolkit that includes the most cutting-edge methods is critical to keep up with the changing times and ensure the safety and efficacy of natural products, especially for product manufacturers.

Authentication using deoxyribonucleic acid (DNA) sequence data is beginning to reveal just how widespread the adulteration issue is. For example, out of 250 randomly selected samples tested at AuthenTechnologies LLC. (Richmond, CA) during 2013 using DNA, a total of 22.4% of the tests were confirmed to be adulterated. Although the figure of 22.4% seems high, we estimate that the real percentage of adulterated materials is actually much greater, especially for blended or more processed materials. As Table 5.1 illustrates, most of the adulterated samples were routine quality assurance/quality control (QA/QC) samples that were identified as being substituted by closely related species (12.8%); this figure was more than double the amount that were adulterated by distantly related substitutes (5.2%). Table 5.2 lists the genera in which we identified adulteration by closely related species. Although some of these taxa are well known for having issues with adulteration, for example, *Cinnamomum* and *Cordyceps* where inferior species are often substituted for more rare and expensive ones, many others were surprising. It is believed that many instances of adulteration by close relatives are due to accidental substitution by species that may look similar or are taxonomically challenging and do not have distinct species boundaries (i.e., *Epimedium* and *Taraxacum*), sometimes as a result of hybridization (i.e., *Glycyrrhiza* and *Passiflora*).

TABLE 5.1
Results from 250 Randomly Selected DNA Species Authentication Tests

Sample Type	Results	Percentage of Test Articles (%)
Routine QA/QC	Correct species identity confirmed	77.6
Routine QA/QC	Closely related substitute identified	12.8
Routine QA/QC	Distantly related substitute identified	5.2
Routine QA/QC	Fungal contamination detected	2.0
Routine QA/QC	Unknown substitute detected	0.8
Out of specification	Distantly related substitute identified	1.6
	Total	100.0

The table includes the type of sample such as a routine QA/QC test (with no indication that the sample was of the incorrect identity) and those that were submitted with prior knowledge of the sample being out of specification using another method. Additionally, the second column describes different categories of results from the DNA authentication tests using DNA sequencing methods. The last column indicates the percentage of test articles that had each of the types of results.

TABLE 5.2
Genera Adulterated by Closely Related Species

Genus	Common Name(s)	Family
Baptisia	Wild Indigo	Fabaceae
Cinnamomum	Cinnamon	Lauraceae
Cistus	Rockrose	Cistaceae
Cordyceps	Cordyceps	Clavicipitaceae
Eleutherococcus	Siberian Ginseng	Araliaceae
Epimedium	Horny Goat Weed	Berberidaceae
Eupatorium/Eutrochium	Gravel root	Asteraceae
Fucus	Bladderwrack	Fucaceae
Glycyrrhiza	Licorice	Fabaceae
Myrica	Bayberry	Myricaceae
Origanum	Oregano, Marjoram	Lamiaceae
Passiflora	Passionflower	Passifloraceae
Plantago	Plantain	Plantaginaceae
Quercus	Oak	Fagaceae
Rhamnus	Buckthorn	Rhamnaceae
Schrophularia	Figwort	Schrophulariaceae
Taraxacum	Dandelion	Asteraceae
Thymus	Thyme, Savory	Lamiaceae
Tribulus	Puncture Vine	Zygophyllaceae
Verbascum	Mullein	Schrophulariaceae
Ziziphus	Jujube	Rhamnaceae
Verbena	Vervain	Verbenaceae

TABLE 5.3
Species Adulterated by Distantly Related Species

Labeled Species	Labeled Family	Labeled Common Name	Identified Species	Identified Family	Identified Common Name
Coriandrum sativum	Apiaceae	Coriander	*Apium graveolens*	Apiaceae	Celery
Eupatorium purpureum	Asteraceae	Gravel root	*Collinsonia canadensis*	Lamiaceae	Stone root
Euphrasia officinalis	Orobanchaceae	Eyebright	*Odontites* sp.	Orobanchaceae	Red Bartisia
Panax quinquefolius	Araliaceae	American Ginseng	*Astragalus membranaceus*	Fabaceae	Astragalus
Petroselinum crispum	Apiaceae	Parsley root	*Pastinaca sativa*	Apiaceae	Parsnip
Picrorhiza kurroa	Plantaginaceae	Kutki	*Musa acuminata*	Musaceae	Banana
Fallopia multiflora	Polygonaceae	Chinese Knotweed	*Silybum marianum*	Asteraceae	Milk Thistle
Ptychopetalum olacoides	Olacaceae	Muira Puama	*Croton echioides*	Euphorbiaceae	Muira Puama
Viburnum opulus	Adoxaceae	Cramp bark	*Acer tataricum*	Sapindaceae	Tatar Maple

In Table 5.3, we have listed examples from the tests that identified the presence of distantly related substitutes. Interestingly, none of the 17 samples were adulterated with their commonly known adulterant. This illustrates the need to use testing that not only detects and identifies close relatives and common adulterants, but also has the ability to detect *unexpected* ones as well. Examples from some of the tests that revealed distantly related adulterants that were unexpected includes substitution of coriander (*Coriandrum sativum*) with celery (*Apium graveolens*) both in the Apiaceae. A particularly unusual one was adulteration of *Picrohiza* with banana root (*Musa acuminata*). We also determined that a sample labeled as *Panax ginseng* was really *Astragalus membranaceus.* Another particularly interesting one was the substitution of Muira Puama (*Ptychopetalum olacoides*) with a plant from a distantly related family, *Croton echioides*, both commonly known as Muira Puama, and with similar aphrodisiac effects. To date, all Muira Puama samples tested at AuthenTechnologies using DNA have been identified as the adulterant, *C. echioides*, including commercially available reference materials. This illustrates the importance of using validated methods of identification, certified reference materials that have been properly authenticated using a valid method, as well as methods that are able to detect and identify unexpected adulterants.

There are myriad DNA-based technologies that have been used widely for decades by academics and law enforcement agencies, from defining species taxonomies to identifying criminals. Yet the natural products community has been slow to adopt DNA-based methods of authentication of ingredients. However, the realization that alternative methods

may be unsatisfactory for identification of some species and detection of adulterants, and increasing pressures to comply with federal regulations has pushed more companies to utilize DNA-based species authentication. The first contract-testing laboratory in the United States to specialize in DNA-based technologies for species identification and adulterant detection, AuthenTechnologies LLC, was founded in 2010. Since that time DNA-based methods have quickly become a preferred test for routine quality control across a wide array of natural product ingredient suppliers and manufacturers. As a result, the National Institute of Standards and Technology (NIST) launched the first line of DNA-authenticated Standard Reference Materials for botanical species in 2013 in partnership with AuthenTechnologies. Further development and validation of DNA-based methods through partnership between AuthenTechnologies, the National Institutes of Health (Office of Dietary Supplements and National Center for Complementary and Alternative Medicine), and the FDA's Center for Food Safety and Nutrition is helping to accelerate the use and acceptance of DNA as a routine method in the natural products industry.

DNA is the building block of life and provides a suite of powerful and sensitive methods that can be used for natural product authentication and detection of adulterants. DNA is in every cell of every living or once-living organism on Earth, and it is the only characteristic that remains intact and unchanged throughout its life and beyond. In addition to DNA's extreme specificity, discriminating between the most closely related species and even individuals depending on the specific technique utilized, there are numerous other advantages of using DNA methods for identification. For instance, DNA is not affected by the developmental stage of the organism, nor by the environmental conditions that it lived in, nor by the season it was harvested. DNA-based methods are reliable and powerful especially when applied to single organ specimens where diagnostic taxonomic characters are not present, when applied to powdered materials where the distinguishing characteristics are no longer visible, and when it is difficult to distinguish among closely related and/or morphologically or chemically similar species. Although the morphology and chemistry of an organism can change drastically throughout its life, the DNA remains stable and consistent. Therefore, variation in DNA markers can be interpreted only by the genetic patterns underlying them, which is the foundation of their identity, and can be used to better understand other characteristics, such as their morphology and chemistry.

DNA-based methods have been used widely for plant identification for at least the past two decades, and advancement in technologies in recent years has made the use of DNA faster and more reliable than ever. Some of the early forms of DNA fingerprinting include methods based on hybridization of random pieces of DNA such as restriction fragment length polymorphism (RFLP), amplification of arbitrary DNA such as amplified fragment length polymorphism (AFLP), or site-targeted polymerase chain reaction (PCR). However, recently a more highly reproducible and informative analysis is the comparison of gene sequences from a specific stretch of DNA, or gene, referred to here as DNA sequencing, or barcoding; DNA barcoding is the use of short DNA sequences for the identification of an organism [1,2]. The DNA barcoding method is also widely used in the authentication of animals, and the method has already been validated by the FDA for identifying fish [3,4]. This method has dominated the field of plant taxonomy for the past decade and is beginning to take root as the preferred

method of medicinal herb authentication [5,6]. Very recent advancements in sequencing technologies, such as next-generation sequencing (NGS), offer exciting new potential methods for genetic authentication, especially for degraded or mixed materials.

In this chapter, we will explore the use of DNA as the most generally useful basis for determining species identity, and in turn for authenticating unknown raw materials. Because genetic methods reveal the genetic history of the organism, DNA methods can be used to identify material of hybrid origin. This is especially important for medicinal and other commercially grown herbs, where humans have interfered with their reproduction through intentional or inadvertent crosses. Another use of genetic methods is to elucidate the speciation patterns, or phylogenetic patterns, of organisms; a *phylogeny* is a branching diagram or tree that shows the relationships between organisms, not unlike a family tree. Figure 5.2 illustrates an example of a phylogeny of black cohosh (*Actaea racemosa*) and its related species in the genus *Actaea*. By understanding the genetic relationships of species and their dispersal patterns throughout the globe, we are often able to understand the unique chemical profiles within them. In addition to providing great specificity, DNA is also extremely sensitive to contamination and can be used to identify multiple species in a mixture, whether intentional additives to a blend, or intentional or accidental adulteration. Depending on the specific method utilized, DNA is able to detect and identify up to *hundreds* of species in a single sample, even unexpected ones. Because they are highly sensitive, DNA sequences can identify many unexpected adulterants, contaminants, and additives that are present in low levels or chemically inert, including common fillers such as soy flour, as well as nonbotanical contaminants such as bacteria, fungi, and insects.

As with any class of methods, such as morphological or chemical, the specific technique used is differently suited for different types of materials and for detection and identification of adulterants. There are dozens of DNA-based methods, which vary in their utility for natural product ingredient authentication. In this chapter, we will briefly review the traditional use of DNA sequencing data for establishing the basis of species identity, summarize the different types of available genetic methods, and discuss their applicability to botanical authentication. Finally, we will provide a *DNA Toolkit* as a practical user's guide to aid in method selection and development. In this guide, we will discuss many of the aspects that are important to consider when applying these in a commercial setting such as the skill level, development time and cost, analysis time and cost, specificity, repeatability, and ability to detect mixtures, including unexpected and expected contaminants. Finally, we will indicate if there are any known commercial testing laboratories currently offering testing and discuss potential applications of using each technology.

Regardless of the specific DNA technology utilized for authentication or adulterant detection, the question remains whether DNA sequence methods should replace or complement chemical identification methods. Although DNA is well suited for taxonomic identification and for identifying species mixtures such as adulterants, it is unable to identify the plant part, as well as the presence or abundance of chemical components, both of which are critical for the quality and efficacy of the final product. Therefore, depending on the application, most DNA-based methods should be complemented by another analytical method that can, for instance, identify the plant part or the amount of specific marker compounds if necessary. However, DNA

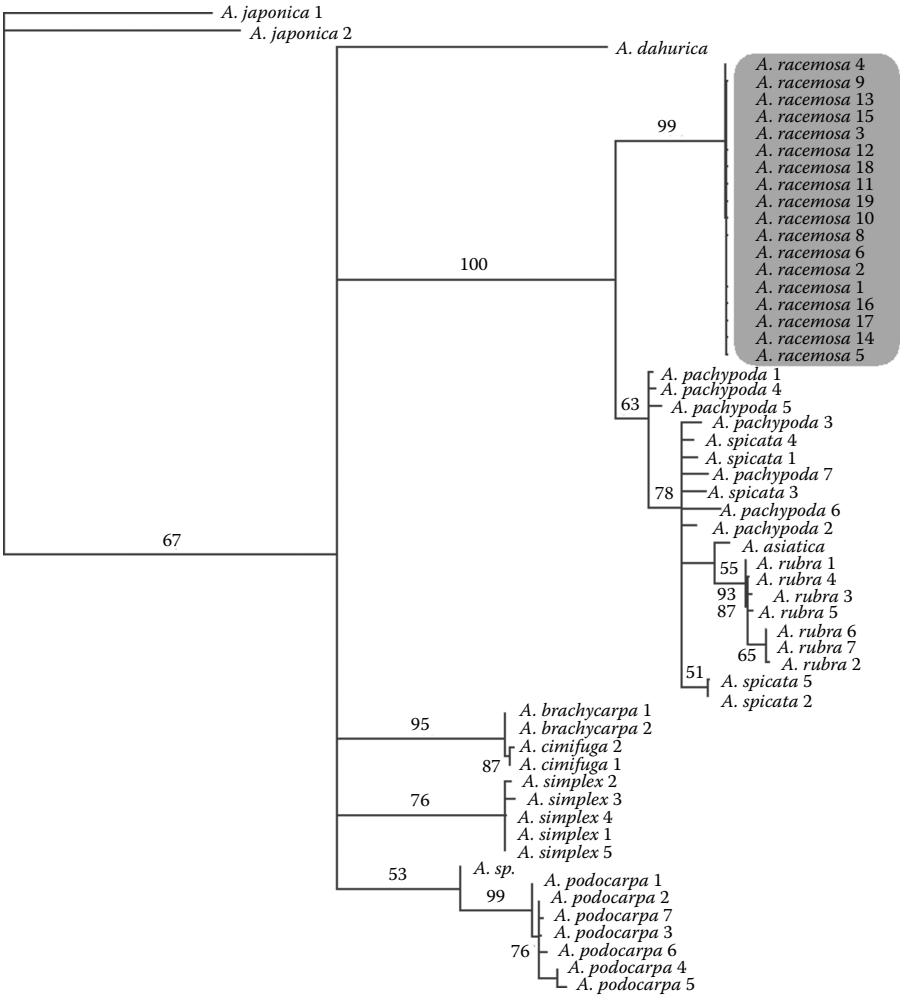

FIGURE 5.2 Phylogeny of black cohosh (*Actaea racemosa*) and related species. A branching diagram using DNA sequence data to elucidate the relationships between multiple specimens of black cohosh (*Actaea racemosa*) and its most closely related species. The numbers on the branches represent the bootstrap (BS) support values out of 100 and the branch lengths are proportional to the amount of genetic change. The numbers after the taxon names are sample numbers, representing independent vouchered collections. The gray box highlights the *Actaea racemosa* samples included in the inclusivity sampling frame, while those not highlighted are the close relatives included in the exclusivity sampling frame.

may be used to replace morphological and microscopic methods of identification, especially in processed materials, which are difficult to perform on a large scale and the quality of the results relies heavily upon the training of the person performing the test. Likewise, it may be used as an alternative to methods such as TLC, HPLC, FTIR, or NIR that are used simply as a species identification tool, and not for characterizing quality or potency. In most cases, the use of DNA may be best used as

the initial identity screen before other more costly chemical tests are performed to characterize the quality of the material. In this chapter, we will discuss where alternative identity estimation methods have fallen short, especially with distinguishing between closely related species, and how DNA-based technologies can alleviate these problems.

DNA Is the Basis of Species Identity

Taxonomic botanists have struggled over the centuries to classify plants into real and meaningful groups. Traditionally, species have been identified based on morphological characters. As in the Linnaean system of taxonomy, features of reproductive organs of the flower and the fruit have played central to determining the difference between species. However, due to the inherent variation within organisms and differences in observers' interpretation of this variation (i.e., *lumpers* vs. *splitters*), many species taxonomies have changed over the years. The complex taxonomic history of herbs and changing circumscriptions present a great challenge to the natural products industries that are required to reliably authenticate incoming lots of herbs and produce consistent products. Fortunately, in the past two decades, there has been an important technological advancement that has increased our knowledge of the identity of organisms and has lead to development of more reliable and robust species taxonomies: DNA sequencing. Today, DNA sequencing is the most commonly used method by taxonomic botanists to classify species or to reclassify species that were traditionally circumscribed based on alternative methods such as morphology or chemistry. Thousands of papers have been published on the use of DNA sequence data for species identification and authentication (e.g., see [2,7]). One major benefit of using DNA as the basis for species identity and authentication of natural products is that it is unbiased and can provide a more solid understanding of the identity and relationships of individuals or species that doesn't change—even if the names do.

As botanists unravel the genetic history of species, they are discovering that many of the traditionally used methods using morphology and chemistry do not correlate with the true identity of the species. The problem with basing a system on morphological or chemical characteristics alone is convergent evolution, whereby unrelated groups have similar features due to adaptation to the environment. For example, American cacti and African euphorbias look similar and are both succulent due to adaptations to arid environments, but they are very distantly related [8]. Likewise, identical or similar chemical compounds may arise separately in unrelated lineages [9]. Although families or genera of plants may contain specific classes of compounds [10], these may not be continuously expressed or may be triggered by herbivory, damage, or be dependent on other environmental factors. As a result, certain metabolites that may be present in a given taxon are not shared with their most closely related ones. Inconsistent secondary metabolite profiles mean that the systematic value of chemical characters becomes a matter of interpretation in the same way as traditional morphological characters. The distribution of secondary metabolites has some value for taxonomy, but their occurrence reflects adaptations and particular life strategies embedded in a given phylogenetic framework. However, these data combined with DNA may be the most appropriate taxonomic system with DNA as the *taxonomic scaffold* by which other chemical and morphological characteristics are applied and used in the field for identification [11]. This integrative approach whereby the DNA blueprint is

correlated with the morphology and chemistry allows us to make more meaningful and reliable categories of plants that are useful to people, as well as help us to identify strains or lineages that have superior qualities, as in marker-aided selection, commonly used in animal and plant breeding, as well as for botanical drug discovery [12–17]. That said, phylogenetic analyses based on DNA sequences are extremely valuable tools for natural products manufacturers, developers, and researchers not only to authenticate ingredients, but also to understand the variation and patterns in secondary metabolites and activity, as well as other functional or qualitative characteristics of organisms [18], and should play a central role in research and validation studies utilizing *any* botanical identification method.

DNA TOOLKIT

DNA-based methods offer a number of excellent solutions for species identification and detection of adulterants [19]. As with any category of methods (i.e., morphological or chemical), DNA is not appropriate for all types of starting materials. Table 5.4

TABLE 5.4
Materials Appropriate for DNA Testing

Species: Any-botanical, animal, fungal, or bacterial
Form: Any-whole, cut, chopped, sliced, powdered
Part:
- *Botanical:* Whole, leaves, stems, roots, bark, flowers, stigmas, pollen, aerial parts, herbs, fruits, and seeds
- *Animal:* Living or once living tissues, organs (e.g., muscle, skin)
- *Fungal:* Mycelia and fruiting body
- *Bacterial:* Whole (living or dried)

Type: Fresh, dried, frozen, juiced, puréed

Examples of appropriate ingredients

Dried herbs and spices	Meat products
Fresh seeds and nuts	Liquid extracts
Fruit purées	Pressed oils (e.g., olive oil)
Pea and soy protein	
Juices and concentrates (liquid and dried)	
Grain flours	
Probiotic bacterial cultures (liquid and dried)	
Supplement capsules and tablets	
Food products (except highly refined)	

Examples of inappropriate ingredients

Dried extracts
Distilled oils (e.g., essential oil)
Plant resins
Milk
Oyster shells

lists examples of materials that can be analyzed using DNA. In general, DNA testing can be used on any species and most forms, as long as there is some cellular material intact. Examples include dried and fresh botanical, fungal, animal, and bacterial samples as well as purées, juices, and select supplement capsules and tablets and finished products such as foods that are not highly processed other materials that often contain DNA are pressed oils (i.e., olive oil) and liquid extracts+tinctures. Examples of materials that do not contain DNA are powdered herbal extracts, essential oils and resins, as well as animal milk and shells such as from oysters.

There are a wide range of DNA-based technologies each with their own advantages and shortcoming for a particular application such as the specificity; precision; repeatability; upfront development cost and time; ability to work on processed materials; difficulty of data analysis; and ability to detect unexpected adulterant, contaminant, or substitute species. Therefore, it is imperative to weigh each of these aspects when selecting the most appropriate method for a particular purpose. Therefore, in this section I will highlight several of the most widely used genetic methods for species identification studies and examine the features of each method that are most relevant to natural products researchers and manufacturers. Features of the genetic tools are rated using my own ordinal scale based on experiences as outlined in Table 5.5, to indicate the level of effort, time, and/or cost.

1. DNA hybridization
   a. *Description:* Comparison of random pieces of DNA such as in RFLP.
   b. *Skill:* 2.
   c. *Development time:* 3.
   d. *Development cost:* 3.
   e. *Analysis time:* 3.
   f. *Analysis cost:* 3.
   g. *Specificity:* 3.
   h. *Repeatability:* 3.
   i. *Degraded DNA analysis:* 1.
   j. *Mixture detection-expected:* 1.
   k. *Mixture detection-unexpected:* 1.
   l. *Commercial availability:* Unknown.
   m. *Discussion:* DNA hybridization techniques were once popular for plant genetic studies [20–23], but have been replaced by more reliable

TABLE 5.5
Scale Used for Rating DNA Tool Features

Rating	Description
0	Very low
1	Low
2	Moderate
3	High
4	Very high

technologies such as DNA sequencing analysis. In RFLP, a restriction enzyme is used to produce DNA fragments of different lengths. In theory, the patterns of DNA fragments vary between different strains or species. The fragments are separated through electrophoresis on an agarose gel. The DNA fragment profile is then transferred to a matrix and hybridized with a fluorescently labeled DNA probe and polymorphisms in fragments are detected by the presence or absence of fragments that fluoresce. Depending on their design, RFLP assays may be highly specific, accurate, and reproducible. It may be able to detect adulterants; however, their identity may not be determined. However, the RFLP profile obtained may be contingent on developmental stage, organ sampled, or environmental conditions depending on the enzyme(s) selected for the restriction reaction. The upfront development time and cost and ongoing sample analysis cost and time are very high. Additionally, it requires high-quality DNA, so it is not a good candidate for processed or degraded materials.

2. Arbitrary PCR
   a. *Description:* PCR amplification of DNA loci using arbitrary primers as in random amplified polymorphic DNA (RAPD), AFLP, and intersimple sequence repeat (ISSR) analysis.
   b. *Skill:* 2.
   c. *Development time:* 2.
   d. *Development cost:* 2.
   e. *Analysis time:* 2.
   f. *Analysis cost:* 1.
   g. *Specificity:* 2.
   h. *Repeatability:* 1.
   i. *Degraded DNA analysis:* 2.
   j. *Mixture detection-expected:* 0.
   k. *Mixture detection-unexpected:* 0.
   l. *Commercial availability:* Unknown.
   m. *Discussion:* PCR-based methods involve the use of primers that select a specific region of the genome and amplify them using a thermal cycler or PCR machine. In general, PCR-based methods are advantageous because only very small quantities of DNA are necessary for analysis. In arbitrary PCR methods, short arbitrary oligonucleotide primers are used to generate a large number of random PCR products. Polymorphism in the products is due to random mutations, which are visualized on a gel with a fluorescent DNA stain. In some methods, such as AFLP, the use of restriction enzymes is combined with PCR of random fragments. These methods require no knowledge of DNA sequences or relationships between individuals or species; therefore, there is little upfront development time and cost. Although these methods are efficient for screening multiple sites across the genome, and can differentiate between individuals within a single species, this category of methods can be prone to issues with repeatability. For instance,

degraded DNA can affect the amplification patterns. Additionally, they are not designed to identify multiple species in a mixture because it is difficult to distinguish between variation in the target and closely related species. Although this category of methods is relatively quick and affordable, it can lead to generation of data that is very complex and difficult if not impossible to analyze and understand patterns, including making species identifications. However, AFLP and RAPD studies have been performed on a number of medical plant groups, especially Traditional Chinese Medicinal herbs [24–29]. This category of methods is best utilized for analysis within a single species or within a set of closely related species, such as to understand relationships between populations, varieties, or cultivars.

3. Site-targeted PCR
   a. *Description:* Amplification and detection of particular loci using specific primers either through traditional PCR, as in cleaved amplified polymorphic sequence, direct amplification of length polymorphisms, amplification refractory mutation system, simple sequence repeats, sequence characterized amplified region, and through the use of quantitative PCR (qPCR) techniques.
   b. *Skill:* 3.
   c. *Development time:* 4.
   d. *Development cost:* 3.
   e. *Analysis time:* 2.
   f. *Analysis cost:* 1.
   g. *Specificity:* 4.
   h. *Repeatability:* 4.
   i. *Degraded DNA analysis:* 4.
   j. *Mixture detection-expected:* 4.
   k. *Mixture detection-unexpected:* 2.
   l. *Commercial availability:* Yes.
   m. *Discussion:* There are a wide variety of PCR-based techniques that utilize primers that amplify specific gene regions or loci. Each one of them varies slightly in their overall skill level, development time and cost, and analysis time and cost, specificity, and repeatability. However, in general, PCR-based methods are extremely reliable in that they are repeatable and can be used on a wide variety of starting materials including lightly processed and cooked finished products and some botanical extracts because they can be designed to amplify very small fragments of DNA. One major advantage of this technique is that it is generally robust against DNA degradation, as primers are designed to amplify small fragments of DNA (~100–200 bp). Additionally, it is highly sensitive and useful for screening multiple potential contaminants or adulterants quickly and affordably.

      Site-targeted PCR has been used to detect adulterants in several important Traditional Chinese Medicines [30,31]. Similar site-targeted

PCR methods have been used to detect species in botanical raw materials [32–35] and, as well as identify the origin of plant species in vegetable oils from olive to soybean oil [36–40] and fruit species in juices [41]. Additionally, one study identified plant species in liquid botanical extracts using this method [42]. Unless careful and extensive validation studies are performed however, to examine the specificity and sensitivity of the PCR primers, and necessary precautions are taken to eliminate contamination, these highly sensitive tests may lead to erroneous results.

The particular assay's ability to discriminate between different species of interest, as well as detect low levels of DNA in a mixture, depends heavily on the design of the assay. For instance, more *universal* primers can be used to detect a large set of organisms (i.e., all plants) or species-specific primers can be used for detection of one species or variety. However, the sensitivity of the primers to detect low levels of contaminants is affected by the primer design and specific PCR cycling parameters used. Additionally, if a particular plant species is excluded from the primer design, they may not be detected. Therefore, expected adulterants can reliably be detected with this method. However, if more universal primers are not also used, unexpected adulterants may be missed; these universal primers, however, tend to be less sensitive to the very specific ones. Although a lot of upfront cost and development time is required to design and validate the PCR primers, the cost and time to test samples and analyze the results once developed is minimal. One of the key components to designing a valid PCR-based method is to first validate the specificity of the gene regions, using DNA sequencing analysis, as described in DNA Sequencing: Sanger section.

In addition to species authentication, site-targeted PCR methods are used more widely in commercial applications for the detection of specific allergens, such as wheat or peanut. Additionally, similar site-targeted methods are used for the detection of genetically modified organisms (GMOs).

4. Chips and arrays
   a. *Description:* Simultaneous detection of a predetermined set of genes or species using one of a number of traditional microarray and emerging chip technologies.
   b. *Development time:* 4.
   c. *Development cost:* 4.
   d. *Analysis time:* 2.
   e. *Analysis cost:* 3.
   f. *Specificity:* 3.
   g. *Repeatability:* 3.
   h. *Degraded DNA analysis:* 3.
   i. *Mixture detection-expected:* 4.
   j. *Mixture detection-unexpected:* 0.

k. *Commercial availability:* Unknown.
l. *Discussion:* DNA chip and microarray technologies are extremely powerful in that they can be used to analyze a high number of genes in several samples simultaneously. There are a number of chip and array technologies, the specifics of which are outside the scope of this chapter. However, depending on their design, chips and arrays can be highly specific, accurate, and reproducible. Although they are relatively time-consuming and expensive to develop, the relative ease of performing routine testing once they are developed makes them suitable for automation. However, they are difficult to modify once developed. These technologies are typically resistant to problems with degraded DNA. Like site-targeted PCR, a predetermined set of species must be designed into the assay in order to detect them. Therefore, the specificity and the accuracy of the assays for detecting specific species can vary from assay to assay and must be validated extensively before being put into use. Applications of this technology include screening for the presence of target and adulterant species in incoming raw materials or finished products and have been used for authentication of medicinal herbs and in herbal drug research [43–47].

5. DNA sequencing: next generation
   a. *Description:* Amplification of a select gene followed by elucidation of the identity and arrangement of bases using one of a number of cutting-edge sequencing technologies called NGS, including pyrosequencing (454), MiSeq (Illumina), sequencing by ligation (SOLiD), and ion semiconductor (Ion Torrent), among others.
   b. *Development time:* 2.
   c. *Development cost:* 2.
   d. *Analysis time:* 4.
   e. *Analysis cost:* 4.
   f. *Specificity:* 4.
   g. *Repeatability:* 4.
   h. *Degraded DNA analysis:* 3.
   i. *Mixture detection-expected:* 4.
   j. *Mixture detection-unexpected:* 4.
   k. *Commercial availability:* Yes.
   l. *Discussion:* There are a number of NGS technologies that vary widely in their specific abilities and cost. However, in general, NGS methods produce huge quantities of data relatively rapidly. Although for the amount of DNA sequencing data they generate in a relatively short period of time the cost per base sequenced is quite low, many thousands if not millions of DNA base pairs can be sequenced with these methods, making the cost per sample much higher than other methods. However, depending on the application, NGS can be a particularly powerful tool and well worth the expense. For instance, NGS can produce upward of 10,000 short DNA sequences from a single sample, which can allow for identification of hundreds of species, even unexpected

ones. One of the drawbacks to NGS methods is that they require highly skilled analysts to perform data analysis and comparison to reference sequences. In general, NGS technologies are well suited for degraded DNA because they amplify and sequence relatively short fragments of DNA. If universal primers are used, there is little upfront development cost and time; however, the extensive time and money spent to perform each analysis and analyze the data makes NGS technologies not suitable for routine authentication testing, until bioinformatics pipelines are optimized. Additionally, before NGS methods are used, the specific gene regions used for identification must be validated for specificity as described below such as through using traditional Sanger sequencing, and the proper reference materials must be obtained. There are numerous potential applications to NGS using universal gene regions, including authentication of ingredients, complex blends, and finished products, as well as product de-formulation and detection of adulterants, allergens, and filth. This technology has also been useful for identification of complex microbial communities in environmental and plant samples [48] as well as for detection of plant and animal adulterants in Traditional Chinese Medicine [49]. This technology has been useful for label claims and detection of unlabeled ingredients, as illustrated in the Certificate of Analysis from NGS testing (Figure 5.3) of a finished dietary supplement capsule in which 16% of the DNA sequences originated from soy, which was not listed on the label.

6. DNA sequencing: Sanger
   a. *Description:* Amplification of a select gene region followed by elucidation of the identity and arrangement of bases using traditional Sanger sequencing, also known as the chain-termination method.
   b. *Development time:* 2.
   c. *Development cost:* 2.
   d. *Analysis time:* 4.
   e. *Analysis cost:* 4.
   f. *Specificity:* 4.
   g. *Repeatability:* 4.
   h. *Degraded DNA analysis:* 3.
   i. *Mixture detection-expected:* 3.
   j. *Mixture detection-unexpected:* 3.
   k. *Commercial availability:* Yes.
   l. *Discussion:* Sanger DNA sequencing based on the chain-termination method has been used widely throughout academic and medical research institutions since the 1980s. Despite the fact that NGS technologies are now emerging, Sanger sequencing still remains the most popular sequencing method, especially for species identification and taxonomic studies and remains the most well-suited choice for routine authentication of ingredients, from all organisms including plants, animals, fungi, and bacteria. DNA sequencing, though Sanger sequencing technology, is highly specific, accurate, and reproducible. The protocols

Latin Name	Family	Common Name	On Label	Percentage (%)
Allium cepa	Alliaceae	Green onion	Yes	0.10
Allium sativum	Alliaceae	Garlic	Yes	0.10
Alternaria alternata	Pleosporaceae	Leaf spot fungus	No	0.17
Amaranthus sp.	Amaranthaceae	Amaranth	Yes	0.00
Apium graveolens	Apiaceae	Celery	Yes	0.17
Asparagus officinalis	Asparagaceae	Asparagus	Yes	0.12
Beta vulgaris	Chenopodiaceae	Beet	Yes	0.30
Brassica oleracea	Brassicaceae	Broccoli, kale, red cabbage, Brussels sprouts, cauliflower	Yes	9.79
Capsicum annuum	Solanaceae	Bell pepper	Yes	0.10
Cicer arietinum	Fabaceae	Garbanzo bean	Yes	0.22
Cladosporium sp.	Davidiellaceae	Mold	No	0.05
Cucumis sativus	Cucurbitaceae	Cucumber	Yes	1.95
Daucus carota	Apiaceae	Carrot	Yes	1.41
Fagopyrum esculentum	Polygonaceae	Buckwheat	Yes	0.00
Fragaria sp.	Rosaceae	Strawberry	Yes	1.04
Fusarium sp.	Nectriaceae	Soil fungus	No	0.17
Gibellulopsis nigrescens	Plectosphaerellaceae	Black fungus	No	0.05
Glycine max	Fabaceae	Soy	No	16.15
Helianthus annuus	Asteraceae	Sunflower seed	Yes	0.07
Lactobacillus bulgaricus	Lactobacillaceae	Lactobacillus bacteria	Yes	0.00
Linum usitassimum	Linaceae	Flax seed	Yes	0.05
Lycopersicon esculentum	Solanaceae	Tomato	Yes	0.00

Medicago sativa	Fabaceae	Alfalfa	Yes	11.08
Petroselenium crispum	Apiaceae	Parsley	Yes	5.59
Phaseolus vulgaris	Fabaceae	Kidney bean	Yes	0.05
Phoma sp.	Incertae sedis	Soil fungus	No	0.20
Phomopsis longicdla	Valsaceae	Mold	No	0.05
Pleospora sp.	Pleosporaceae	Plant fungus	No	0.05
Prunus avium	Rosaceae	Cherry	Yes	0.30
Rubus fruticosus	Rosaceae	Blackberry	Yes	0.35
Rubus idaeus	Rosaceae	Raspberry	Yes	0.32
Saccharomyces cerevisiae	Saccharomycetaceae	Brewer's yeast	Yes	21.37
Salvia hispanica	Lamiaceae	Chia seed	Yes	0.05
Sesamum indicum	Pedaliaceae	Sesame seed	Yes	0.00
Spinacea oleracea	Amaranthaceae	Spinach	Yes	28.47
Vaccinium corymbosum	Ericaceae	Blueberry	Yes	0.10
Vigna angularis	Fabaceae	Adzuki bean	Yes	0.00

FIGURE 5.3 Next-generation sequencing test results using 454 pyrosequencing from a finished product. The table in this figure indicates the Latin name, Family, and Common name of all species identified in a finished dietary supplement capsule. Additionally, the fourth column indicates whether or not the species was listed on the label. In the final column, the percentage of sequences originating from each species is indicated.

are straightforward and amenable to automation. The upfront development time and cost, as well as sample analysis time and cost, is relatively low, especially when using universal primers. Degraded DNA can be a problem if large fragments are being amplified; however, methods can be modified to amplify multiple shorter overlapping fragments to counteract degradation. Data analysis is usually simple and DNA standards are not needed with each test sample. Taxonomic discrimination is dependent on the markers targeted, but a well-designed sequencing assay can be highly discriminatory. When universal primers are used, unexpected adulterants can easily be detected. Additionally, mixtures and hybrids can also be detected by observing overlapping bases in the DNA chromatogram. Therefore, based on these criteria authentication of species based on DNA sequencing (using Sanger sequencing technology) is arguably the most appropriate routine method for commercial applications, primarily raw material ingredient authentication and detection of adulterants. There are numerous published papers on using DNA sequencing to detect adulteration in raw materials (see [19]) and finished products, from dietary supplements to teas [50–53]. See next sections for more information on sample preparation and validation of DNA sequencing methods.

DNA SEQUENCING SAMPLE PREPARATION AND METHOD VALIDATION

DNA Extraction

As with any DNA-based method, the first step in the process is DNA extraction. There are numerous published methods, for instance, using cetyl trimethylammonium bromide or using silica columns, as well as a variety of commercially available kits that use silica columns or glass-coated magnetic beads. However, the suitability of the method or kit depends on the starting material, and the nature of the DNA extraction procedure can dramatically affect the success of the amplification reactions. For instance, many medicinal plants contain an abundance of secondary metabolites (polysaccharides, tannins, essential oils, phenolics, alkaloids, and waxes), which if not properly removed can interfere with the PCR and produce a negative result. Additionally, many of the materials used commercially have not been specifically dried or stored in a way that preserves the DNA (such as rapid desiccation using silica gel). Therefore, the DNA of commercial materials is often degraded and minor adjustments to standard protocols may be necessary.

Gene Selection and Amplification

Choosing the desired gene region to sequence is the most critical step in using DNA sequences for plant identification; the gene must be variable enough to distinguish between the target and nontarget species. Although for animals, a single gene or

DNA barcoding region, cytochrome oxidase I is useful for authentication; because of the heterogeneous rates of evolution between plant groups, different gene regions are necessary in different groups to resolve relationships. For example, in a recent validation study funded by the National Center for Complementary and Alternative Medicine, testing eight different potential regions across five commonly used medicinal herbs found that a unique combination of two loci was best suited for each plant group (Reynaud, unpublished). For instance, in some groups, particular loci would not amplify at all or had little or no variation, while in other groups, they were easily amplified and contained significant variation. Because the base mutations that are used for species identification occur randomly in the genome, it is difficult to predict which regions are going to contain the features necessary for species resolution. Therefore, testing and validating multiple candidate regions is critical for DNA sequence authentication. In some cases, a single gene region may be sufficient for identification. However, multiple genes from different parts of the genome (i.e., chloroplast or nuclear) are necessary for independent verification of the identity and to ensure that different genes have not undergone different evolutionary histories, such as hybridization [54]. Fortunately, intraspecific hybrids can easily be identified visually using DNA sequencing methods as long as nuclear genes are analyzed as they provide evidence from both parents, as opposed to chloroplast genes, which are typically maternally inherited. Figure 5.4a illustrates what a DNA chromatogram looks like when a hybrid, or mixture of very closely related species with equal concentrations, is present. Figure 5.4b shows an example of a DNA chromatogram from a sample that is contaminated by a more distantly related species; the low level of underlying *noise* is due to the presence of a secondary DNA sequence that is of a different length from the dominant signal.

Once the appropriate gene regions are selected, amplification of the genes is performed using a thermal-cycler, or PCR machine. In addition to the design of the primers used to select the specific regions, a number of other factors can affect the success of amplification including the enzymes used in the reaction, as well as

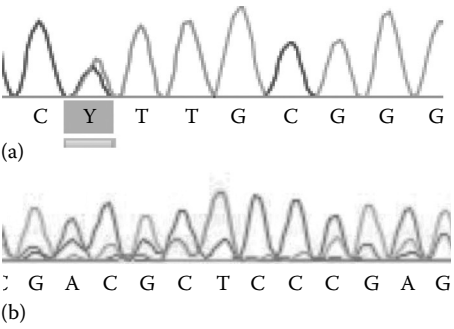

FIGURE 5.4 DNA chromatograms showing evidence of hybridization and contamination. (a) Hybridization or contamination by a close relative is evident by two overlapping bases (as indicated by *Y*, where both *C* and *T* are present) at specific bases in a sequence. (b) Contamination by a more distant relative is evident by multiple overlapping bases along all or most of the sequence.

the temperature and cycling times used in the PCR machine. All of these factors must be tested in a proper validation study to examine their affect on the sensitivity and specificity of the methods.

DNA Sequencing and Reference Material Comparison

Once PCR products are amplified, the DNA is sequenced using Sanger sequencing. Nowadays, capillary electrophoresis machines are the most widely used for sequencing the DNA. The resultant DNA chromatograms can be obtained electronically from the sequencing machine and used for further analysis and identification. Comparison of the DNA sequences to authentic reference materials is a critical step in performing authentication of a test sample. As with any method, the reliability and accuracy of a method depends on the quality of the reference materials, both of the target and nontarget species. One major benefit to using DNA sequences for reference materials is that huge databases of reference sequences can be built to which test samples can be compared, as opposed to alternative methods whereby reference materials must be run alongside the test sample, which limits the scope of references that can be used. Currently, AuthenTechnologies has initiated development of the Herbal Reference Barcode (HERB™) Database of validated DNA reference sequences obtained from herbarium vouchers for plant authentication and detection of adulterants, in partnership with major herbaria around the United States including its partner at the Center for Herbal Identity at the University of California, Berkeley. What separates DNA from chemical or other analytical methods is that it is the *only* method that can analyze herbarium vouchers; because DNA is consistent throughout all organs of a plant, we are able to obtain DNA reference sequences from tiny leaf fragments from the vouchers. This enables us to obtain reference materials not only from commonly traded herbs, but also from rare ones and other close relatives not typically available for purchase through other reference material suppliers. Most importantly, by having the ability to base DNA identification methods on herbarium material that can unambiguously be authenticated by a botanist, using the important taxonomic features, DNA-based methods are extremely scientifically robust.

Once sequences from authentic reference materials are obtained for the appropriate gene regions, the DNA sequence of test samples is compared in a number of ways to make an identity confirmation. Typically, the first step is to produce an aligned DNA matrix, which allows us to compare homologous bases in the DNA sequence. Although there are a number of computer algorithms that can help with producing alignments, it is imperative to check each alignment visually, as this is critical for making an accurate identification. For materials in which there is little sequence divergence within the target species, and the nontarget species are easily distinguished, taxonomic identification can be made by visually inspecting the DNA sequences and identifying specific bases that uniquely discriminate the target species. For those species with more variation that are difficult to assess visually, test materials can be authenticated by building a phylogenetic tree, or branching diagram, using one of a number of different algorithms. Because identification cannot be made if a test sample falls outside the range of variation of the reference

materials, it is imperative to validate methods using a wide range of samples from within the target and closely related species, as described subsequently.

Method Validation

It is critical when using DNA sequencing for species authentication that the gene regions are validated for specificity, such as by following AOAC's Guidelines for Botanical Identification Method Validation [55]. These validation procedures are based on testing a wide sampling of the target and most problematic nontarget species (which in the case of genetics are the most closely related ones) to ensure that the genes differentiate between them a desired percentage of the time (i.e., probability of identification [POI] exceeds 95%). Additionally, the range of potential variation that can be expected in the target species must be characterized, in order to not falsely reject authentic materials. Gene specificity must be validated before other methods of genetic analysis are used, such as site-specific PCR or even NGS methods. Fortunately, because DNA sequences are not affected by environmental conditions or plant part, the range of variation within species is often negligible. Therefore, once gene regions are carefully selected, the reliability of these methods to correctly identify species is exceptional. In fact, for all of the validation studies completed at AuthenTechnologies to date, POI is 100% or nearly so for all genes selected after careful pre-validation studies have been performed to examine the rates of variation within genera (unpublished data). Figure 5.2 illustrates a validation study of black cohosh (*Actaea racemosa*) and its close relatives based on nuclear ribosomal DNA (nrDNA) data, in which approximately 20 herbarium vouchers of *Actaea racemosa* were sampled from across its geographic range, as well as multiple exemplars from its most closely related species. In this figure, all samples of black cohosh are nearly indistinguishable genetically, indicated by the cluster of *A. racemosa* specimens with no resolution between them. However, the group of *A. racemosa* is well supported (bootstrap support value of 99) and distinct from all other closely related species of *Actaea* analyzed. Therefore, this figure demonstrates the utility of DNA sequence data to unequivocally identify black cohosh and exclude all other close relatives.

CONCLUSIONS

Accurate authentication of natural products and the ingredients used to formulate them is the foundation for any quality control process. However, because of the rampant rates of adulteration and the fact that many materials used by manufacturers are powdered, processed, or blended, the already difficult task of identification can become impossible unless alternative and/or cutting-edge technologies are considered and integrated into QC processes, or used to replace methods that are not suitable. In this chapter, we shed light on the *real* rates of adulteration in the marketplace using DNA and on the issue of substitution by close relatives, which often go undetected by alternative methods. We briefly reviewed the importance of using DNA as the foundation for developing and validating *any* methods of identification and understanding patterns in chemical and morphological variation, because DNA is the *basis* for species taxonomy. Finally, in the *DNA Toolkit*, we reviewed several of

the most applicable genetic methods for species authentication and detection of adulterants. As with any category of methods, genetic methods vary in their overall reliability and careful selection of the most appropriate method for the intended purpose is necessary. Whether it is authenticating incoming lots of raw materials, verifying label claims on finished products, developing authenticated reference materials, or identifying the most closely related species to include in validation studies of other methods, DNA-based methods offer a wide array of extremely valuable methods that should be considered as part of any authentication or quality control process.

REFERENCES

1. W.J. Kress, K.J. Wurdack, E.A. Zimmer, L.A. Weigt, and D.H. Janzen. Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America* 102, 2005: 8369–8374.
2. N.J. Sucher and M.C. Carles. Genome-based approaches to the authentication of medicinal plants. *Planta Medica* 74, 2008: 603–623.
3. H.F. Yancy et al. A protocol for validation of DNA-barcoding for the species identification of fish for FDA regulatory compliance. *Laboratory Information Bulletin* 24, 2008: 1–25.
4. S.M. Handy et al. A single-laboratory validated method for the generation of DNA barcodes for the identification of fish for regulatory compliance. *Journal of AOAC International* 94, 2011: 1–10.
5. M.T. Cimino. Successful isolation and PCR amplification of DNA from National Institute of Standards and Technology Herbal Dietary Supplement Standard Reference Material powders and extracts. *Planta Medica* 76, 2010: 495–497.
6. J. Ma, S.L. Chen, M.E. Thibault, and J. Ma. Enhancing quality control of botanical medicine in the 21st century from the perspective of industry: The use of chemical profiling and DNA barcoding to ensure accurate identity. *HerbalGram* 97, 2012: 58–68.
7. D.T. Harbaugh and B.G. Baldwin. Phylogeny and biogeography of the sandalwoods (*Santalum*, Santalaceae): Repeated dispersals throughout the Pacific. *American Journal of Botany* 94, 2007: 1028–1040.
8. A. Bennici. The convergent evolution in plants. *Rivista di Biologia* 96, 2003: 485–489.
9. E.D. Brodie. Convergent evolution: pick your poison carefully. *Current Biology* 20, 2010: R152–R154.
10. M. Wink. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* 64, 2003: 3–19.
11. D. Tautz, P. Arctander, A. Minelli, R.H. Thomas, and A.P. Vogler. A plea for DNA taxonomy. *Trends in Ecology & Evolution* 18, 2003: 70–74.
12. B. Sidjimova, S. Berkov, S. Popov, and L. Evstatieva. Galanthamine distribution in Bulgarian *Galanthus* species. *Pharmazie* 58, 2003: 936–937.
13. S. Berkov, B. Sidjimova, L. Evstatieva, and S. Popov. Intraspecific variability in the alkaloid metabolism of *Galanthus elwesii*. *Phytochemistry* 65, 2004: 579–586.
14. N. Rønsted, G.D. Weiblen, W. Clement, N. Zerega, and V. Savolainen. Reconstructing the phylogeny of figs (*Ficus*, Moraceae) to unravel the origin of fig-wasp mutualisms. *Symbiosis* 45, 2008: 45–56.
15. N. Rønsted et. al. Can phylogeny predict chemical diversity and potential medicinal activity of plants? A case study of Amaryllidaceae. *BMC Evolutionary Biology* 12, 2012: 182.
16. M.M. Larsen, A. Adersen, A.P. Davis, M.D. Lledo, A.K. Jäger, and N. Rønsted. Using a phylogenetic approach to selection of target plants in drug discovery of acetylcholinesterase inhibiting alkaloids in Amaryllidaceae tribe Galantheae. *Biochemical Systematics and Ecology* 38, 2010: 1026–1034.

17. C.H. Saslis-Lagoudakis et al. The Use of Phylogeny to Interpret Cross-Cultural Patterns in Plant Use and Guide Medicinal Plant Discovery: An Example from *Pterocarpus* (Leguminosae). *PLoS One* 6(7), 2011: e22275. doi:10.1371/journal .pone.0022275.
18. I. Schmitt and F.K. Barker. Phylogenetic methods in natural product research. *Natural Product Reports* 26, 2009: 1585–1602.
19. G. Heubl. New aspects of DNA-based authentication of Chinese medicinal plants by molecular biological techniques. *Planta Medica* 76, 2010: 1963–1974.
20. M. Yamazaki, A. Sato, K. Saito, and I Murakoshi. Molecular phylogeny based on RFLP and its relation with alkaloid patterns in *Lupinus* plants. *Biological & Pharmaceutical Bulletin* 16, 1993: 1182–1184.
21. N. Trifi-Farah and M. Marrakchi. *Hedysarum* phylogeny mediated by RFLP analysis of nuclear ribosomal DNA. *Genetic Resources and Crop Evolution* 48, 2001: 339–345.
22. N. Mori, T. Moriguchi, and C. Nakamura. RFLP analysis of nuclear DNA for study of phylogeny and domestication of tetraploid wheat. *Genes & Genetic Systems* 72, 1997: 153–161.
23. N.J. Gawel, R.L. Jarret, and A.P. Whittemore. Restriction fragment length polymorphism (RFLP)-based phylogenetic analysis of *Musa. Theoretical and Applied Genetics* 84, 1992: 286–290.
24. M. Yamasaki, A. Sato, K. Shimomura, K. Saito, and I. Murakoshi. Genetic relationships among *Glycyrrhiza* plants determined by RAPD and RFLP analyses. *Biological & Pharmaceutical Bulletin* 17, 1994: 1529–1531.
25. J.P. Loh, R. Kiew, A. Kee, L.H. Gan, and Y.Y. Gan. Amplified fragment length polymorphism (AFLP) provides molecular markers for the identification of *Caladium bicolor* cultivars. *Annals of Botany (London)* 84, 1999: 155–161.
26. K.T. Chen et al. Identification of *Atractylodes* plants in Chinese herbs and formulations by random amplified polymorphic DNA. *Acta Pharmacologica Sinica* 22, 2001: 493–497.
27. M. Zhang, H.R. Huang, S.M. Liao, and J.Y. Gao. Cluster analysis of *Dendrobium* by RAPD and design of specific primer for *Dendrobium candidum*. *Zhongguo Zhong Yao Za Zhi* 26, 2001: 442–447.
28. W.Y. Ha, P.C. Shaw, J. Liu, F.C. Yau, and J. Wang. Authentication of *Panax ginseng* and *Panax quinquefolius* using amplified fragment length polymorphism (AFLP) and directed amplification of minisatellite region DNA (DAMD). *Journal of Agricultural and Food Chemistry* 50, 2002: 1871–1875.
29. J.J. Qi, X.E. Li, J. Song, A.E. Eneji, and X. Ma. Genetic Relationships among *Rehmannia glutinosa* cultivars and varieties. *Planta Medica* 74, 2008: 1846–1852.
30. T. Feng, S. Liu, and X.-J. He. Molecular authentication of the traditional Chinese medicinal plant *Angelica sinensis* based on internal transcribed spacer of nrDNA. *Electronic Journal of Biotechnology* 13, 2010: 1–10.
31. C.Z. Wang, P. Li, G.-Q. Jin, and C.-S. Yuan. Identification of *Fritillaria pallidiflora* using diagnostic PCR and PCR-RFLP based on nuclear ribosomal DNA internal transcribed spacer sequences. *Planta Medica* 71, 2005: 384–386.
32. C. Howard, P.D. Bremner, M.R. Fowler, B. Isodo, N.W. Scott, and A. Slater. Molecular identification of *Hypericum perforatum* by PCR amplification of the ITS and 5.8S rDNA region. *Planta Medica* 75, 2009: 864–869.
33. N. Techen, I.A. Khan, Z. Pan, and B.E. Scheffler. The use of polymerase chain reaction (PCR) for the identification of *Ephedra* DNA in dietary supplements. *Planta Medica* 72, 2006: 241–247.
34. N. Techen, Z. Pan, B.E. Scheffler, and I.A. Khan. Detection of *Illicium anisatum* as adulterant of *Illicium verum*. *Planta Medica* 75, 2009: 392–395.

35. V. Joshi, N. Techen, B.E. Scheffler, and I.A. Khan. Identification and differentiation between *Hoodia gordonii* (Masson) Sweet ex Decne., *Opuntia ficus indica* (L.) P. Miller, and related *Hoodia* species using microscopy and PCR. *Journal of Herbs, Spices & Medicinal Plants* 15, 2009: 253–264.
36. M. Hellebrand, M. Nagy, and J.T. Mörsel. Determination of DNA traces in rapeseed oil. *European Food Research and Technology* 206, 1998: 237–242.
37. N. Gryson, F. Ronsse, K. Messens, M. De Loose, T. Verleyen, and K. Dewettinck. Detection of DNA during the refining of soybean oil. *Journal of the American Oil Chemists' Society* 79, 2002: 171–174.
38. M. Busconi, C. Foroni, M. Corradi, C. Bongiorni, F. Cattapan, and C. Fogher. DNA extraction from olive oil and its use in the identification of the production cultivar. *Food Chemistry* 83, 2003: 127–134.
39. C. Breton, D. Claux, I. Metton, G. Skorski, and A. Bervilleä. Comparative study of methods for DNA preparation from olive oil samples to identify cultivar SSR alleles in commercial oil samples: Possible forensic applications. *Journal of Agricultural and Food Chemistry* 52, 2004: 531–537.
40. S. Kumar, T. Kahlon, and S. Chaudhary. A rapid screening for adulterants in olive oil using DNA barcodes. *Food Chemistry* 127, 2001: 1335–1341.
41. M.-A.L. Clarke, J.J. Dooley, S.D. Garrett, and H.M. Brown. An investigation into the use of PCR-RFLP profiling for the identification of fruit species in fruit juices. FSA Final Report Q01111. CCFRA Project 98200, 2008.
42. J. Novak, S. Grausgruber-Groger, and B. Lukas. DNA-based authentication of plant extracts. *Food Research International* 40, 2007: 388–392.
43. P.Y. Tsoi, H.S. Wu, M.S. Wong, S.L. Chen, W.F. Fong, P.G. Xiao, and M.S. Yang. Genotyping and species identification of *Fritillaria* by DNA chip technology. *Acta Pharmaceutica Sinica* 4, 2003: 185–190.
44. T. Li, J. Wang, and Z. Lu. Accurate identification of closely related *Dendrobium* species with multiple species-specific gDNA probes. *Journal of Biochemical and Biophysical Methods* 62, 2005: 111–123.
45. Y.B. Zhang, J. Wang, Z.T. Wang, P.P. But, and P.C. Shaw. DNA microarray for identification of the herb of *Dendrobium* species from Chinese medicinal formulations. *Planta Medica* 69, 2003: 1172–1174.
46. W.Y. Lin, L.R. Chen, and T.Y. Lin. Rapid authentication of *Bupleurum* species using an array of immobilized sequence-specific oligonucleotide probes. *Planta Medica* 74, 2008: 464–469.
47. P. Chavan, K. Joshi, and B. Patwardhan. Review DNA microarrays in herbal drug research. *Evidence-Based Complementary and Alternative Medicine* 3(4), 2006: 447–457. doi:10.1093/ecam/nel075.
48. A.R. Ottesen et al. Baseline survey of the anatomical microbial ecology of an important food plant: *Solanum lycopersicum* (tomato). *BMC Microbiology* 13, 2013: 114. http://www.biomedcentral.com/1471-2180/13/114.
49. M.L. Coghlan, J. Haile, J. Houston, D.C. Murray, N.E. White, P. Moolhuijzen, M.I. Bellgard, and M. Bunce. Deep sequencing of plant and animal DNA contained within traditional Chinese medicines reveals legality issues and health safety concerns. *PLoS Genetics* 8(4), 2012: e1002657. doi:10.1371/journal.pgen.1002657.
50. S.G. Newmaster, M. Grguric, D. Shanmughanandhan, S. Ramalingam, and S. Ragupathy. DNA barcoding detects contamination and substitution in North American herbal products. *BMC Medicine* 11, 2013: 222.
51. D.P. Little. The use of DNA barcode techniques to identify the constituents of herbal dietary supplements. *Planta Medica* 78, 2012: IL11.

52. D.A. Baker. DNA barcode identification of black cohosh herbal dietary supplements. *Journal of AOAC International* 95, 2012: 1023–1034.
53. S. Chen et al. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS One* 5, 2010: e8613.
54. D.T. Harbaugh, H. Oppenheimer, K. Wood, and W.L. Wagner. Taxonomic revision of the Hawaiian red-flowered sandalwoods (*Santalum*) and discovery of an ancient hybrid lineage. *Systematic Botany* 35, 2010: 827–838.
55. AOAC International; Guideline Working Group. AOAC guidelines for validation of botanical identification methods. *Journal of AOAC International:* 95, 2012: 268–272.