Chapter 2
DNA-Based Authentication of TCM-Plants: Current Progress and Future Perspectives

G. Heubl

2.1 Introduction

Traditional Chinese Medicine (TCM), with its long history, is deeply rooted in the Chinese culture and represents one of the oldest forms of medical therapy in the world. TCM has been used for thousands of years in China for health maintenance, disease prevention, and used to a lesser extent for the application of a variety of clinical therapies. Today, the global use of TCM is rapidly increasing and over 130 countries in the world are using Chinese Herbal Medicine (Hsiao 2007). The majority of drugs harvested and processed in China are of plant origin. A highly developed industry delivers medical plant preparations to the local Chinese customers and exports plant-based drugs to Asia and developing markets in Europe, Canada, and the USA.

Increasing interest by multinational pharmaceutical companies in herbal-based medicine is contributing to the significant economic growth of the global TCM market. It is estimated that 70–80 % of people worldwide rely chiefly on traditional herbal remedies as their primary form of health care (Farnsworth and Soejarto 1991; Pei 2001). Herbal medicine is becoming more main-stream and fashionale in richer countries. Therefore the TCM market has been growing at 10–20 % annually in Europe and North America over recent years (Hamilton 2003; Ten Kate and Laird 1999). In addition, there are many related botanical products sold as health foods, food supplements, herbal teas, and for various other purposes related to health and personal care. In terms of the number of species, the use of plants as medicine represents by far the most important application area of natural resources. Plants provide the predominant ingredients of medicine in most medical traditional
systems of healing and have been the source of inspiration for several new drug search endeavors of major pharmaceutical companies (Schippmann et al. 2002).

### 2.2 Biodiversity and Medicinal Plant Resources in China

China is one of the world’s megabiodiverse countries with 32,308 species of seed plants belonging to 363 families and 3,427 genera (Ding 2002). The country spans a huge geographical area, is characterized by enormous variations in geographical, climatological, and topographical features and covers five climatic zones (cold-temperate, temperate, warm-temperate, subtropical, and tropical). The geology and geography is very complex, comprising the highest mountain ranges on Earth (the Himalayas), vaste plateaus such as Qinghai-Xizang (Tibetan) or Pamir, and arid basins such as Tarim, which contains the largest desert in China (Taklamakan). Main rivers of Asia such as Mekong, Brahmaputra, Yangtze, and Yellow River originate in Qinghai-Xizang Plateau (Lopez-Pujol and Zhao 2004). All features contribute to the enormous diversity of biomes (from rainforests to deserts) found in China, as well as to the enormous species diversity including many medicinal plants (Lopez-Pujol et al. 2006).

There is no reliable data for the total number of medicinal plants on earth and available information for countries and regions vary greatly. Estimates for the number of plant species used medicinally include 35,000–70,000 worldwide (Farnsworth and Soejarto 1991; Schippmann et al. 2002), 7,295–11,146 in China (Pei-Gen 2007; Hamilton 2004; Huang et al. 2002a, b). According to statistics there are 12,807 kinds of remedies in Chinese medicine. Of these, 11,146 (80 %) are plant based, 1,581 are animal based, and 80 are mineral based (Gao et al. 2002). Regarding the diversity of Chinese medicinal plants, only 500–600 species are commonly used in Traditional Chinese Medicine and prescribed by Chinese medical practitioners, 1,430 in Mongolian Medicine, and 1,106–3,600 in Tibetan Medicine (Pei 2001, 2002b). In the 2005 edition of the Chinese Pharmacopoeia 1,146 monographs and 538 herbal drugs have been included. The improved version of the People’s Republic of China Pharmacopoeia 2010 contains 1,174 monographs (in total 4,567 monographs) and enumerates more than 4,600 varieties of species. In Europe and USA 75–150 herbal drugs are mostly used by TCM practitioners (Table 2.1).

The Chinese medicinal industry represents a significant portion of the pharmaceutical industry in China. There are 1,200 Chinese medicinal, industrial enterprises that manufacture approximately 8,000 Chinese herbal medicine products. The total annual sales of functional food, TCM preparations, medicinal plant extracts, and other processed materials exceed US $ 40 billion (Liu et al. 2011). Today, it is a big challenge to protect the medicinal plant resources from overexploitation and habitat destruction (Srivastava et al. 1996). For example, 70–80 % of natural materials were collected from the wilderness to meet the annual demand of approximate one million tons (Balunas and Kinghorn 2005). The annual sales of these natural resources have
increased to more than 100 times in relation to the levels of 1980. Many of these natural resources are derived from species that are threatened or have become rare or endangered by large-scale exploitation (Nalawade et al. 2003; Cole et al. 2007). With the rapid increase of consumer demand for crude drugs and natural health products, many medicinal plant species are threatened with extinction due to overexploitation and habitat destruction. To ensure the sustainable use of medicinal plant resources, a conservation framework consisting of conservation strategies, cultivation practices, and various technologies has been developed. Conservation strategies include establishing in situ and ex situ conservation centers, setting up government policies and regulations, establishing methods for resource surveying and trade monitoring as well as establishing and enforcing Good Agricultural Practices (Liu et al. 2011).

About 140 new drugs have originated directly or indirectly from Chinese medicinal plants by means of modern scientific methods, confirming that these plants are an important resource (Liu and Yaniv 2005; Lee et al. 2005). The first compound derived from Chinese herbal remedies to enter the Western pharmacopoeia was Ephedrine, an amphetamine-like stimulant isolated from the Chinese medicinal herb “Ma-Huang” (*Ephedra sinica*). The next significant pharmaceutical breakthrough derived from Chinese medicine was the isolation of Artemisinin from “Qing-Hao” (*Artemisia annua*). Researchers found that Artemisinin was beneficial for fever and killed even chloroquine-resistant strains of *Plasmodium*, the parasite that causes malaria. Recent work suggests that Artemisinin may also have anticancer properties, a hypothesis which was also established for *Coix lacryma-jobi* var. *ma-yuen* the Job’s Tear. Modern pharmacological studies have demonstrated that the Coix “seed” (a bony utricle of the single female flower enclosing the fruit) and its preparation-KLT (Kanglaite Injection) possess extensive pharmacological

### Table 2.1 Diversity and inventory of the Chinese flora with data on medicinal plants

<table>
<thead>
<tr>
<th>Resource</th>
<th>Total number of species</th>
<th>Medicinal plant species number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algae</td>
<td>5,000*/8,997*/12,500*</td>
<td>115*</td>
</tr>
<tr>
<td>Mosses</td>
<td>2,200*a,c,d,e,f</td>
<td>43*a</td>
</tr>
<tr>
<td>Ferns</td>
<td>2,300*/2,600*a,b,d,e,f</td>
<td>456*a</td>
</tr>
<tr>
<td>Seed plants</td>
<td>25,000*/30,000*b,c,d</td>
<td>10,188*a</td>
</tr>
<tr>
<td>Gymnosperms</td>
<td>192*/200*/250*/270*d</td>
<td></td>
</tr>
<tr>
<td>Angiosperms</td>
<td>25,000*/30,000<em>c,d/31,500</em></td>
<td></td>
</tr>
<tr>
<td>Total plants</td>
<td>34,692–47,570</td>
<td>10,802</td>
</tr>
<tr>
<td>Lichens</td>
<td>2,000*c</td>
<td>52*c</td>
</tr>
<tr>
<td>Fungi</td>
<td>8,000*c,d</td>
<td>292*c</td>
</tr>
</tbody>
</table>

*Gao et al. (2002)*
*Lopez-Pujol and Zhao (2004)*
*Xue (1997)*
*SEPA (1998)*
*Li (2003)*
*NEPA (1994)*
*Flora of China Project Missouri Botanical Garden (http://www.mobot.org/press/Assets/FP/flora_china.asp). Estimated numbers of other medicinal resources in China: Medicinal animals 1,581; medicinal mine 80 (Huang et al. 2002a, b)*
activities (Li 2001; Lu et al. 2008). Active components have been isolated and defined in many other Chinese herbs, for example, anthraquinone glycosides in rhubarb (*Rheum officinale*), gingerols in ginger (*Zingiber officinale*), ginkgolides in *Ginkgo* (*Ginkgo biloba*) or berberine, an antibacterial component from Chinese goldthread (*Coptis chinensis*) (Wang et al. 2007a).

### 2.3 Adulterants, Substitutes, and Confused Species

Although many Chinese medicinal plants have been used successfully in China for a long period of time they have never been subject to the stringent quality standards and regulations for pharmaceutical products in Europe or USA. To make traditional and innovative plant products acceptable for modern medicine, it is necessary to have reliable botanical, phytochemical and molecular identification tools for the identification of medicinal plant species (Liang et al. 2004).

A severe problem of the global TCM market is that many erroneous substitutes and adulterants of Chinese medicinal plants are traded due to their lower costs or due to the misidentification of species with similar morphological features.

In recent times the use and demand of herbal preparations has been growing in Western countries where medicinal plants have gained popularity and attention among physicians and patients. Among the alternative medicines TCM is becoming one of the most widely used therapies throughout the world (Normile 2003). One of the reasons for the increasing interest in herbal medicines is the belief that, being natural and traditionally used, they are hence safe and do not possess the potential for negative side effects. This coupled with lower costs compared to conventional medications is the major attraction of these treatments. Nevertheless, their natural origin is not a guarantee of safety. In literature, many reports point out the risks associated with the use of herbal products (Chan 1997; Ernst 2004; Ng et al. 2009).

The quality of a herbal drug can be negatively affected by the use of inherent toxic herbs, by fraudulent action due to substitution or adulteration, by contamination, by misidentification, by confusion of species, and by inappropriate labeling. There are several incidents of Chinese herbs which document that adulterants or substitutes caused serious intoxications and even deaths (But 1994; Graham-Brown 1992; Ng et al. 2009; Chan 2003; Mazzanti et al. 2008; Zhao et al. 2007a; Chan and Critchley 1996; Gertner et al. 1995; But et al. 1996; Chen et al. 2002; McIntyre 1998; Yang and Chen 1998).

A case of encephalopathy and neuropathy was reported following ingestion of a decoction supposedly prepared from “Long Dan Cao” (*Gentiana rigescens* radix). Investigation showed that the toxicity was in fact due to adulteration of the herb with the roots of *Podophyllum emodi* or *P. hexandrum* which contain the neurotoxin podophyllotoxin in high concentration (But 1994; Ng et al. 2009). Several cases of renal damage attributed to “Fang-Ji” (*Stephania tetrandra*) in a weight-loss preparation were actually caused by “Guang Fang Ji” (*Aristolochia fangchi*) that contains the highly toxic aristolochic acid, a known nephrotoxin which causes renal failure.
and urothelial carcinoma. The confusion in the latter case has obviously arisen from the similar Chinese names (Fugh-Berman 2000). Several other poisoning cases are documented and resulted from erroneous substitution of herbs.

One further example of species confusion, which had serious consequences, involves the common name “Bai Mao Teng.” This name has been used for *Solanum lyratum* and *Aristolochia mollissima*. Apparently, these herbs belong to two different families but look similar. While *S. lyratum* is not harmful, *A. mollissima* contains the toxic aristolochic acid that can cause kidney failure and cancer of the urinary tract (Zhao et al. 2006).

Another incident refers to commercial “Mu Tong,” a Chinese diuretic drug, which is associated with five species: *Clematis armandii* and *Clematis montana* (“Chuan Mu Tong”), *Akebia quinata* (“Wu Ye Mutang”), *A. trifoliata* (“San Ye Mu Tong”), *Aristolochia manshuriensis* (“Guan Mu Tong”), and *A. moupinensis/A. kaempferi* (“Huai Mu Tong”). This has led to a serious confusion with the consequence that substitution with *Aristolochia manshuriensis* caused renal failure (Zhu 2002; Cheung et al. 2006; Debelle et al. 2008).

Very often the high price of some herbs is an incentive for criminals to make quick profit by manufacturing counterfeit products. The more precious and rare the crude drug is, the more likely the counterfeit products will find their way into the market. A classic example is *Cordyceps sinensis* (recently renamed as *Ophiocordyceps sinensis*), a fungus parasitizing the larva of some species of insects and the dead caterpillar (Dong and Yao 2010). This scarce drug is found only in fairly inaccessible alpine regions on the Tibetan Plateau and is one of the best-known traditional Chinese medicinal products, with great benefits to human health and with a huge economic value. There are many substitutes of so-called *Cordyceps* that are traded worldwide, such as *Cordyceps militaris* (the most commonly used substitute), *C. martialis*, *C. hawkesii*, *C. liangshanensis*, *C. barnesii*, *C. cicadicola*, *C. gracilis*, *C. ramose*, *C. ophioglossoides*, and *C. gunnii*. In addition, there are counterfeit plant products of the fungus and mimics such as *Stachys geobombycis*, *Stachys sieboldii*, and *Lycopus lucidus* that have been found on the market. Consequently there is a serious problem for authentication and quality control of *Cordyceps* (Li et al. 2006).

A review published by Zhao et al. (2006) enumerates further examples and summarizes the facts and reasons for confusion in the current Hong Kong medicinal herb market.

### 2.4 Authentication of Herbal Material on the DNA Level

The authentication of Chinese medicinal plants, depending on the correct identification of species, is an essential prerequisite to ensure safety, herbal drug quality, and therapeutic efficacy (Zhao et al. 2006). Identification of herbal materials, which commonly consist of dried or processed parts, is generally difficult because many useful diagnostic characters are lost during the drying process. This is particularly
true when one herb has more than one common name, or where one common name is used for more than one herb (Zhao and Li 2004). In practice, the identification of medicinal plants relies mainly on morphological, anatomical, and phytochemical characters. Many pharmacopoeias refer to macroscopic and microscopic evaluation (morphology, histology) and chemical profiling (TLC-, HPLC-, GC-fingerprinting) for quality control and standardization of raw and processed herbs (Chan 2003; Siow et al. 2005; Wagner et al. 2011). However, chemical variability within the plant material often hinders the confirmation of its botanical identity as the chemical composition is affected by growth and storage conditions as well as by the harvesting process. Otherwise microscopic examination of drugs requires botanical expertise for the unequivocal authentication, as related species often possess similar features.

With the improvements in molecular biotechnology and plant genetics in the past decades, genetic tools are considered to provide more reliability for authentication of herbal materials at the DNA level (Kumar et al. 2009). Thus in the meanwhile, various DNA-based molecular marker techniques are applied in many fields and their application is remarkably increasing for species characterization in medicinal plants (Shaw et al. 2002, 2009; Joshi et al. 2004; Zhang et al. 2007; Sucher and Carles 2008). This is especially useful in case of those species that are frequently substituted or adulterated with other species or in case of varieties that are morphologically and/or phytochemically almost indistinguishable.

Benefiting in the first place from PCR techniques, DNA markers have become a powerful tool for identification and authentication of plant, animal, fungal, and bacterial species (Yip et al. 2007; Kaplan et al. 2004; Pereira et al. 2008; Hao et al. 2010). Contrary to chemical fingerprinting which is strongly influenced by the age of the sample, physiological conditions, environmental factors, cultivation area, harvesting period, drying, and storage conditions, DNA is an extremely stable macromolecule that is not affected by external factors and therefore can be recovered from fresh, dried, and even processed biological material. Additionally the marker molecules are not tissue specific and thus can be detected at all stages of organism development. Moreover, only a small amount of a sample is sufficient for analysis.

This review provides an overview of DNA-based technologies and most commonly used assays with an emphasis on those that are based on DNA hybridization, restriction enzymes, random PCR amplifications, species-specific PCR primers, and DNA sequencing. A critical evaluation of all methods is presented focusing on their discriminatory power, sensitivity, reproducibility, user-friendliness, and costs (see Table 2.3).

The previously described incidents highlight the importance of a correct botanical classification and the need of an adequate knowledge of morphological characteristics of herbal drugs in order to perform a proper identification. As more and more people worldwide use Chinese herbs, authentication becomes an increasing problem because adulterated and substituted Chinese medicinal materials are widely common in the market (Mills and Bone 2005).

Therefore authentication of Chinese medicinal materials is the key for safety, appropriate use and maximum therapeutic potency, minimization of trading fraud,
and last but not least for the increase of consumer confidence in Chinese medicine. Identification at species level is required for quality assurance, which includes both identifying the crude plant product and evaluating its pharmaceutical quality (Wagner et al. 2011). Thus, authentication is a fundamental step for the successful, reliable clinical application and for accurate experimental studies of TCM plants. This will ensure the safe and effective use of Chinese medicinal herbs throughout the world.

2.4.1 Types of DNA Markers Used in Plant Genome Analysis

There are various types of DNA-based molecular techniques that are used to evaluate DNA polymorphism in order to authenticate plant taxa (Sucher and Carles 2008; Shaw et al. 2009; Yip et al. 2007; Kaplan et al. 2004; Pereira et al. 2008; Heubl 2010). These are hybridization-based methods, polymerase chain reaction (PCR)-based methods, and sequencing-based methods. In recent times the use of multilocus sequence analysis (MLSA), which is commonly used in phylogenetic studies, has proven its discriminatory power. Additionally DNA microarrays that contain thousands of probes are a promising new development for sensitive and high-throughput taxon identification (Trau et al. 2002; Schena et al. 1998).

2.4.1.1 RAPD (Randomly Amplified Polymorphic DNA)

The RAPD technology utilizes short synthetic oligonucleotides (10 bp long) of random sequences as primers to generate a high number of anonymous DNA fragments via PCR reaction. The large number of amplification products is generally separated on agarose gels and stained with ethidium bromide or SyBRgreen.

Using an appropriate annealing temperature in the PCR cycle, oligonucleotide primers bind to several priming sites on the complementary sequences in the template genomic DNA. If these priming sites are within an amplifiable distance discrete DNA fragments are generated. Nucleotide variation between different sets of template DNAs will result in the presence or absence of bands because of changes in the priming sites. Polymorphism of amplified fragments is caused by (1) base substitutions or deletions in the priming sites, (2) insertions that render priming sites too distant to support amplification, or (3) insertions or deletions that change the size of the amplified fragment (Weising et al. 2005).

Because of the simplicity (no prior sequence information is necessary), low costs, efficiency in developing a large number of DNA markers in a short time, and requirement for less sophisticated equipment, RAPDs have found a wide range of applications. Although the RAPD method is easy to perform, the issue of reproducibility has been an important concern. In fact, the RAPD reaction is far more sensitive than conventional PCR because of the length of a single and arbitrary primer, which is used to amplify anonymous regions of a given genome. Special
care is needed for keeping out contaminant DNA (from infections and parasites) in the material to avoid misleading patterns. RAPDs are inherited as dominant-recessive characters which mean that homozygotes and heterozygotes cannot be distinguished.


2.4.1.2 AP-PCR (Arbitrary Polymerase Chain Reaction)

AP-PCR (or Arbitrarily Chosen Primers ACP-PCR) is a special variation of RAPD which is using single primers approximately 10–50 bp in length (Welsh and McClelland 1990). In AP-PCR the amplification follows three steps. In the first two cycles annealing is under nonstringent conditions. Higher primer concentrations are used in the first cycle. Often primers of variable length are used and products are mostly analyzed on polyacrylamide gels. AP-PCR has been applied to various groups for identification of species and analysis of genetic variation (Munthali et al. 1992; Kersten et al. 2007). Similar to RAPDs, reproducibility can be a problem for fingerprints generated by a single primer, because small changes in annealing conditions can affect banding pattern.

2.4.1.3 DAF (DNA Amplification Fingerprinting)

DNA Amplification Fingerprinting (DAF) is a variant of the RAPD technique and was developed by Caetano-Anollés et al. (1991a). For PCR amplification, very short oligonucleotide primers of arbitrary sequence (mostly five nucleotides) are used to amplify short fragments of genomic DNA resulting in a very complex banding pattern (Caetano-Anollés et al. 1991a, b). DAF uses low stringency amplification conditions so that primers can anneal arbitrarily at multiple sites on each template DNA strand and initiate DNA synthesis. The method needs careful optimization of parameters and only two temperature cycles are required. DAF products are routinely separated by polyacrylamide gels and detected by silver staining (Chawla 2002). Another approach called ASAP (arbitrary signatures from
amplification profiles) was developed to enhance the level of informativeness of DAF reactions by using primers which contain both a 5′ mini-hairpin sequence and a short “core” arbitrary 3′ sequence. These arbitrary mini-hairpin primers increase detection of polymorphic DNA and direct the controlled amplification of small template molecules, thereby generating “sequence signatures” from PCR-amplified fragments (Caetano-Anollés and Gresshoff 1996).

2.4.1.4 ISSR (Inter Simple Sequence Repeat)

In higher plants, Inter Simple Sequence Repeat or ISSR markers are frequently applied because they are known to be abundant, very reproducible, highly polymorphic and easy to use (Zietkiewicz et al. 1994; Bornet et al. 2002). ISSR, also known as anchored simple sequence repeat (ASSR), has been used in genetic fingerprinting, gene tagging, phylogenetic analysis, species and cultivar identification, and assessment of hybridization (Kurane et al. 2009).

The ISSR technique is nearly identical to RAPD except that ISSR primers are designed from microsatellite regions and are longer (approximately 14 bp or more) than RAPD primers. Microsatellites are very short stretches of DNA that are “hypervariable,” expressed as different variants within populations and among different species. ISSR uses the presence of Simple Sequence Repeats (SSRs) which are characterized by mono-, di-, or trinucleotide repeats (e.g., AA... or AG... CAG...) that have 4–10 repeating units side-by-side. These SSRs are ubiquitous, abundant, and highly polymorphic. The primers used can be 5′ or 3′ anchored by 1–3 selective nucleotides to prevent internal priming and to amplify only a subset of the targeted inter-repeat regions. ISSR markers access variations in the numerous microsatellite regions dispersed throughout the genome. The PCR products are mostly separated on agarose gels and stained with ethidium bromide. Alternatively amplified DNA fragments can also be screened and detected using capillary electrophoresis which significantly increases the amount of information compared to the traditional agarose gel electrophoresis. Since ISSRs are dominant markers, the amplified fragments are scored as diallelic. Presence (0/1) of loci can be used for genetic similarity or cluster analysis. Changes in the amplified products can arise through structural changes in the region (insertions or deletions) or the loss of primer binding sites.

ISSR markers overcome the weakness of low reproducibility of RAPDs, the high costs of AFLPs, the complexity of SSRs and thus this technique is less time consuming, more cost-efficient, use of radioactivity is not required, no prior sequence information is necessary, and it shows high polymorphism.

ISSRs have been used for screening genetic diversity and authentication of Dendrobium (Shen et al. 2006), Cistanche (Shi et al. 2009), Fritillaria (Li et al. 2009a), Salvia (Song et al. 2010), Rehmannia (Wang et al. 2005), Vitex (Hu et al. 2007), Cannabis (Kojoma et al. 2002), Rhodiola (Xia et al. 2007), Cymbidium (Wang et al. 2009), Ammopiptanthus (Ge et al. 2005), Swertia (Joshi and Dhawan 2007), Glycyrrhiza (Yao et al. 2008), and Houttuynia (Wu et al. 2005).
2.4.1.5 AFLP (Amplified Fragment Length Polymorphism)

Amplified fragment length polymorphism (AFLP) originally developed by Zabeau and Vos (1993) is a powerful tool for DNA fingerprinting of organismal genomes and it combines the use of RFLP and PCR techniques. This multilocus approach needs no prior sequence information, it is highly reproducible, with the ability to screen a large number of loci (ca. 50–100 fragments per reaction) for polymorphisms. It is a very useful technique for DNA fingerprinting, especially when very little information on the genome of the plant under study is available (Mueller and Wolfenbarger 1999; Blears et al. 1998).

The procedure of this technique (Vos et al. 1995) is a multistep process (1) Digestion of total genomic DNA with two restriction enzymes. One restriction enzyme is a frequent cutter (four-base recognition site, e.g., MseI), the second restriction enzyme is a rare cutter (six-base recognition site, e.g., EcoRI). (2) Adapters specific to the restriction sites are ligated to the fragment ends which serve as binding sites for selective primers in PCR amplification. (3) A first PCR (preselective amplification) is performed. The PCR primers consist of a core sequence (part of the adapter), a restriction enzyme specific sequence, and 1–5 selective nucleotides (the higher the number of selective nucleotides, the lower the number of bands obtained per profile). (4) Preselective amplification products undergo another PCR run, and again a subset of those fragments is selected. Usually, for the second selective amplification, two extra nucleotides are added to the primers. (5) Amplified fragments, labeled with fluorescent or radioactive tags, are separated on acrylamide gels or with automated genetic analyzers and dominant markers are scored as the presence/absence of loci (Weising et al. 2005).

The advantages of AFLPs lie in their high genomic abundance, considerable reproducibility, in their generation of many informative bands per reaction, in their wide range of applications, and in the fact that no sequence data are required for primer construction. Compared with the widely used RFLP, AFLP is faster, less labor intensive, and provides more information.

Disadvantages include the need for purified, high molecular weight DNA, the dominance of alleles, and the possible nonhomology of comigrating fragments belonging to different loci. Because of the highly informative fingerprinting profiles, which are usually obtained as results, AFLPs can be applied in studies involving genetic identity, parentage, identification of clones and cultivars, and in phylogenetic studies of closely related species.

AFLP analyses have been used in *Panax* (Choi et al. 2008; Ha et al. 2002), *Actaea* (Zerega et al. 2002), *Plectranthus* (Passinho-Soares et al. 2006), *Caladium* (Loh et al. 1999), *Cannabis* (Datwyler and Weiblen 2006), and *Rehmannia* (Qi et al. 2008).
This method, also termed RAHM (random amplified hybridization microsatellite) or RAMS (randomly amplified microsatellites), combines arbitrarily primed PCR (RAPD) with microsatellite hybridization to produce polymorphic genetic fingerprints (Weising et al. 2005; Weising and Kahl 1998). No prior sequence information is needed. Genomic DNA is first amplified with a single arbitrary 10-mer primer (as in RAPDs) or microsatellite-complementary 15- to 16-mer PCR primer. After electrophoretic separation and staining of the PCR products, the gel is either dried or blotted onto a nylon membrane and subsequently hybridized to a radiolabeled mono-, di-, tri-, or tetranucleotide repeat probe such as [GT]8, [GA]8, or [CAA]5. Subsequent autoradiography reveals reproducible, probe-dependent fingerprints that are completely different from the ethidium bromide staining patterns and that are polymorphic at an intraspecific level. An advantage of the RAMPO technique is the low complexity of banding patterns, which is considerably facilitating the detection of species-specific bands. RAMPO bands appear to be less sensitive to misinterpretation than RAPD bands, because not only the size but also the hybridization signal intensity of two bands (i.e., the presence and copy number of a certain microsatellite) are criteria for homology. The method is mainly used for identification and discrimination of genotypes within and among populations, cultivates, and germplasm e.g., in *Ficus* (Chatti et al. 2007) and *Phoenix dactylifera* (Soumaya et al. 2008). RAMPOs share their fate with other marker technologies that are partly or totally based on blot hybridization. These methods are barely used anymore, because more convenient marker systems are available for most purposes.

**2.4.1.7 RFLP (Restriction Fragment Length Polymorphism)**

RFLP analysis was one of the first techniques to be widely used for detecting variations at the DNA level. The principle of this method is based on the comparison of banding patterns from DNA sequences digested with specific restriction enzymes (e.g., HaeIII, EcoRI, BamHI). Restriction enzymes are endonucleases produced by bacteria with the function to cut specific DNA sequence motifs of invading DNA molecules. Each enzyme has a specific, typically palindromic recognition sequence. Consequently it recognizes and cuts DNA in a predictable way, resulting in a reproducible set of DNA fragments of different lengths. If two organisms (strains, individuals, or species) differ in the distance between sites of cleavage of a particular restriction endonuclease, then also the length of the fragments produced differs. Consequently RFLP is a result of (a) point mutation creating or destroying a restriction site and (b) insertion/deletions altering the size of a given restriction fragment. These differences in fragment lengths can be detected by gel electrophoresis, hybridization, and visualization. To detect specific fragments, the DNA restriction fragment profile of the agarose gel is transferred to
a nitrocellulose or nylon membrane. Afterwards a single-stranded DNA probe is conveniently labeled, using any standard method (e.g., a radioisotope or digoxigenin) and hybridized with the target DNA, which is stuck to the membrane. Polymorphisms are detected by the presence or absence of bands. RFLP is a robust methodology: the markers are relatively polymorphic, codominantly inherited, and highly reproducible. The method also provides opportunity to simultaneously screen numerous samples. The technique is time consuming, costly, labor intensive and requires a large quantity of good quality or undegraded DNA (Weising et al. 2005).

RFLP combined with DNA hybridization has mainly been used for phylogenetic studies in the past e.g., in *Lupinus* (Yamazaki et al. 1993), *Hedysarum* (Trifi-Farah and Marrakchi 2001), *Triticum* (Mori et al. 1997), *Musa* (Gawel et al. 1992), and for detection of *Dendrobium* (Li et al. 2005) and *Fritillaria* (Tsoi et al. 2003).

### 2.4.1.8 Microsatellites or SSR (Simple Sequence Repeats)

Microsatellites also known as simple sequence repeats (SSRs), short tandem repeats (STRs), or simple sequence length polymorphisms (SSLPs) are the smallest class of simple repetitive DNA sequences (Litt and Luty 1989; Gupta et al. 1996). Based on tandem repeats of short (2–6 bp) DNA sequences, these markers are highly polymorphic due to variation in the number of repeat units and dispersed throughout most eukaryotic genomes. The large number of alleles and the high levels of variability among closely related organisms made PCR-amplified microsatellites the marker system of choice for a wide variety of applications like population genetic studies, genome mapping, and marker-assisted breeding (Valdes et al. 1993; Akkaya et al. 1995; Schuler et al. 1996). It is meanwhile proven that the predominant mutation mechanism in microsatellite tracts is slipped-strand mispairing (Levinson and Gutman 1987). The repeat length at specific SSR loci is easily assayed by PCR using primers specific to conserved regions flanking the repeat. PCR fragments are usually separated on polyacrylamide gels or capillary sequencers in combination with fluorescent detection systems. The reason for the wide usage of nuclear microsatellites is their high abundance, enormous extent of allelic diversity, and suitability for automatization. Meanwhile many SSR primers deduced from flanking sequences of known microsatellites are deposited in DNA databases.

In addition to nuclear microsatellites, chloroplast microsatellites are also particularly effective markers for analysis of the genetic diversity (Clark et al. 2000) and phylogeography of plant populations (Chen et al. 2011), studying mating systems, gene flow via both pollen and seeds, detection of hybridization, and introgression (Agarwal 2008). One limitation of the approach is the need of sequence data for primer construction. Primer sequences flanking chloroplast microsatellites are usually inferred from fully or partially sequenced chloroplast genomes. In general, these primer pairs produce polymorphic PCR fragments from the species of origin and their close relatives, but transportability to more distant taxa is limited. Attempts to design universal primers to amplify chloroplast microsatellites have
resulted in a set of consensus chloroplast microsatellite primers (ccmp1–ccmp10) that aims at amplifying cpSSR regions in the chloroplast genome of dicotyledonous angiosperms (Weising and Gardner 1999). Most of the primer pairs derived from A or T mononucleotide repeats \( (n = 10) \) identified in the tobacco chloroplast genome, were functional as genetic markers in the Actinidiaceae, Brassicaceae, and Solanaceae (Chung and Staub 2003). Universal primers for the amplification of chloroplast microsatellites in grasses (Poaceae) have also been developed (Provan et al. 2004).

A major limitation of SSRs are the time and high development costs required to isolate and characterize each locus when preexisting DNA sequence is not available. Typically, this process requires the construction and screening of a genomic library of size-selected DNA fragments with SSR-specific probes, followed by DNA sequencing of isolated positive clones, PCR primer synthesis, and testing. With the availability of large numbers of ESTs and other DNA sequence data, development of EST-based SSR markers is less time consuming and expensive. The methodology and applications of nuclear microsatellite markers in plants and other organisms has been subjected to numerous reviews, including Goldstein and Schlötterer (1999), Zane et al. (2002), Li et al. (2002), Varshney et al. (2005), and Squirrell et al. (2003).

Microsatellites have been applied in Panax (Kim et al. 2007a; Joe et al. 2009), Acanthopanax (Kim and Chung 2007), Dendrobium (Fan et al. 2009), Cymbopogon (Kumar et al. 2007), Bupleurum (Chun et al. 2009), and Schisandra (Boqian et al. 2009).

2.4.1.9 SAMPL (Selective Amplification of Microsatellite Polymorphic Loci)

The SAMPL technique was introduced by Morgante and Vogel (1994). It combines the high and controllable multiplexing rate of the AFLP technique with the high levels of microsatellite polymorphism by using AFLP-type primers together with compound microsatellite primers (Weising et al. 2005).

SAMPLs differ from AFLPs by using primers with compound microsatellite motifs in combination with oligonucleotides complementary to the end-ligated adapters for the selective amplification step (Paglia et al. 1998). In brief, genomic DNA is digested with restriction enzymes (commonly EcoRI and MseI), and the resulting fragments are ligated to adapters. Afterwards a preamplification reaction for all ligated DNA fragments is carried out with primers annealing to the adapters. These preamplified products are then used as templates for a selective SAMPL-polymerase chain reaction (PCR) reaction that uses the adapter-primer (EcoRI oligo-1) in combination with an end-labeled microsatellite-based 15-mer oligonucleotide (Karp and Edwards 1997) to amplify a group of fragments from those fragments that were restricted, ligated, and preamplified. A disadvantage of multilocus SSR profiling is the capture of only some of the polymorphism associated with microsatellites due to the prevalence of dominant markers and difficulty in identifying allelic fragments in complex DNA fingerprints. This
multiplexing genome profiling technique has not been used adequately in plant genomics, although a few reports have already documented its potential for detecting polymorphisms (Molina and Kahl 2002). This method was used for analysis of genetic diversity in *Cicer* (Winter et al. 1999), *Lactuca* (Witsenboer et al. 1997), and *Tribulus* (Sarwat et al. 2008).

### 2.4.1.10 DAMD (Directed Amplification of Minisatellite-Region DNA)

DAMD is a DNA fingerprinting method based on amplification of the regions rich in minisatellites at relatively high stringencies by using previously found VNTR core sequences as primers (Heath et al. 1993; Somers and Demmon 2002). Minisatellites also known as variable number of tandem repeats (VNTR) or hyper-variable repeats (HVR) are similar to microsatellites (SSR) except that the tandem repeat DNA sequences are longer and generally consist of 10–60-bp motifs. Extreme variations in the tandem repeat copy number of minisatellite loci are responsible for the polymorphism observed. By using the VNTR core sequences as primers, the directed amplification of minisatellite-region DNA (DAMD) with PCR is capable of producing RAPD-like results for the identification of species (Silva et al. 2001). They are also used to generate highly variable probes for DNA fingerprinting. This method is more reproducible than RAPD due to the longer primers used.

Recently, DAMD-PCR has been applied successfully for genotyping of wheat cultivars and rice species (Zhou et al. 1997). The method has been used for authentication of *Panax* (Ha et al. 2002), *Capsicum* (Ince et al. 2009), *Salvia* (Karaca et al. 2008), and *Morus* (Bhattacharya et al. 2005).

### 2.4.1.11 SNP (Single Nucleotide Polymorphism)

Single nucleotide polymorphisms (SNPs) are widely observed between individuals, ecotypes, and species, serving as efficient molecular markers particularly in genetic analysis and breeding programs, also including ecological and evolutionary studies. SNPs are single-base pair positions in the genomes of two (or more) individuals, at which different sequence alternatives (alleles) exist. Polymorphisms result from point mutations (either transition or transversion events) causing single base-pair differences between DNA sequences. According to most recent estimates, one SNP occurs every 100–300 bp (or every 1,000 bp) in any genome (Kwok 2001). SNPs are codominant, single-locus, biallelic markers and they are the most abundant molecular markers known so far. The major SNP genotyping techniques fall into at least six groups (1) direct sequencing, (2) restriction enzyme digestion (cleaved amplified polymorphic sequences = CAPS), (3) allele-specific PCR, (4) allele-specific primer extension, (5) allele-specific oligonucleotide hybridization, and (6) allele-specific oligonucleotide ligation (Weising et al. 2005).
Several molecular markers such as restriction fragment length polymorphism (RFLP), cleaved amplified polymorphic sequence (CAPS), amplification refractory mutation system (ARMS), and single-strand confirmation polymorphism (SSCP) are based on SNPs.

Many biotechnology companies are marketing DNA microarrays that can test a sample DNA for thousands of SNP sequences.

SNPs have been applied in authentication of *Perilla* varieties (Luo et al. 2006a), *Dendrobium officinale* (Ding et al. 2008), *Panax* cultivars (Wang et al. 2010a), *Boehmeria* varieties (Li et al. 2010b).

### 2.4.1.12 ARMS (Amplification Refractory Mutation System)

ARMS, also known as allele-specific polymerase chain reaction (AS-PCR) is a simple, timesaving, and effective method for detecting any mutations involving single base changes (SNPs) or small deletions. It has become a standard technique that allows the discrimination of alleles (Newton et al. 1989). The basis of ARMS is that oligonucleotides with a mismatched 3’-residue will not function as primers in the PCR. ARMS allows amplification of test DNA only when the target allele is contained within the sample and it does not amplify the nontarget allele. Following an ARMS reaction, the presence or absence of a PCR product is diagnostic for the presence or absence of the target allele. A main advantage of ARMS is that the amplification step and the authentication step are combined, in a way that the presence or absence of a PCR product is diagnostic for the presence or absence of the target allele. The method provides a quick screening assay that does not require any form of labeling as the amplified products are visualized simply by agarose gel electrophoresis and ethidium bromide staining. Multiplex ARMS or MARMS are a similar approach but there are several primer combinations to be optimized simultaneously, which increases the complexity of the procedure. The ARMS technique has been applied in authentication of *Alisma* (Li et al. 2007), *Panax* (Zhu et al. 2004; Diao et al. 2009), *Rheum* (Yang et al. 2004), *Dendrobium* (Ding et al. 2008; Qian et al. 2008), and *Curcuma* (Sasaki et al. 2002).

### 2.4.1.13 CAPS or PCR-RFLP (Cleaved Amplified Polymorphic Sequence)

CAPS, originally named PCR-RFLP, is a combination of PCR of target DNA and subsequent digestion with a restriction enzyme (Maeda et al. 1990; Lum et al. 2005). CAPS markers are generated in two steps. In the first step of a standard CAPS experiment, a defined sequence is amplified using specific 20–25 bp primers. In the second step, the PCR-product is digested with a restriction enzyme usually with a four-base recognition specificity. The digested fragments are separated on agarose gels and stained with ethidium bromide. To identify suitable combinations of amplicons and restriction enzymes, a wide range of PCR primer pairs and restriction enzymes need to be screened during the initial phase of a CAPS
project, using a small set of templates. Combinations that reveal informative polymorphisms are then applied to the full set of organisms under investigation. However, the ability of CAPS to detect DNA polymorphism is not as high as SSRs or AFLPs because nucleotide changes affecting restriction sites are essential for the detection of DNA polymorphism by CAPS. Furthermore, the development of CAPS markers is only possible where mutations disrupt or create a restriction enzyme recognition site. Advantages of CAPS include the involvement of PCR requiring only low quantities of template DNA, the codominance of alleles, and the high reproducibility. The results can be easily scored and interpreted. Compared to RFLPs, CAPS analysis does not include the laborious and technically demanding steps of Southern blot hybridization and radioactive detection procedures. However, in comparison with RFLP analysis, CAPS polymorphisms are more difficult to find because of the limited size of the amplified fragments. Furthermore, sequence data are needed to design PCR primers.

PCR-RFLP has been used for authentication of Alisma (Li et al. 2007), Angelica (Watanabe et al. 1998), Sinopodophyllum and Dysosma (Gong et al. 2006), Ephedra (Guo et al. 2006), Fritillaria (Wang et al. 2005, 2007b), Artemisia (Lee et al. 2009), Panax (Diao et al. 2009; Do et al. 2001; Lu et al. 2010b; Um et al. 2001), Actinia (Zhao et al. 2007b), Atractylodes (Mizukami et al. 2000), Glehnia (Mizukami et al. 1993a), Astragalus (Lu et al. 2009), Dendrobium (Zhang et al. 2005a), Duboisia (Mizukami et al. 1993b), and Codonopsis (Fu et al. 1999).

2.4.1.14 SCAR (Sequence Characterized Amplified Region)

In 1993, Paran and Michelmore introduced a new type of RAPD-derived molecular marker, which circumvented several of the drawbacks inherent to RAPDs. A SCAR marker can be used to rapidly amplify a diagnostic nucleic acid from herbal materials using a pair of specific oligonucleotide primers designed from polymorphic RAPD (Semagn et al. 2006; McDermott et al. 1994) or ISSR (Albani et al. 2004) fragments.

Polymorphic fragments from RAPDs or ISSR are selected among amplified fingerprints. After cloning and sequencing for the selected polymorphic regions, pairs of internal primers are designed to amplify a unique and specific sequence designed as a SCAR marker. SCARs are advantageous over RAPD markers as they detect only a single locus, their amplification is less sensitive to reaction conditions, and they can potentially be converted into codominant markers (Paran et al. 1991). Prior sequence information (i.e., sequencing the polymorphic fragments) is required for designing the primers flanking the polymorphic region. As PCR inhibitory effects of ingredients can lead to false negative results, amplification of a control fragment using the same DNA template should be performed to ensure that the quality of sample DNA is suitable for PCR. The concept of generating locus-specific SCARs from anonymous PCR fragments is not restricted to RAPDs but was applied to other multilocus marker techniques such as ISSR or AFLP (Shan et al. 1999; Xu et al. 2001).
The SCAR technique has been used for authentication of *Panax* (Wang et al. 2001; Choi et al. 2008) and for discrimination of species of *Artemisia* (Lee et al. 2006a; Theerakulpisut et al. 2008), *Phyllanthus* (Dnyaneshwar et al. 2006; Devaiah and Venkatasubramanian 2008a), *Sinocalycanthus* (Ye et al. 2006), *Embelia* (Devaiah and Venkatasubramanian 2008b), and *Lycium* (Sze et al. 2008).

### 2.4.1.15 SSCP (Single-Strand Confirmation Polymorphism)

SSCP is a powerful mutation detection system. The principle of this technique is that under neutral condition, the single-stranded DNA (ssDNA) folds into a tertiary structure. Differences in DNA sequences (often a single base pair) alter the single stranded DNA in the tertiary conformation (by differential folding), which in turn affect the mobility of the ssDNA in a gel. Based on their mobility differences, SNPs can be detected (Orita et al. 1989).

F-SSCP is an adapted version of SSCP analysis involving amplification of the target sequence using fluorescent primers (Makino et al. 1992). The method is not frequently applied for authentication e.g., in *Boesenbergia* (Techaprasan et al. 2008).

### 2.4.1.16 MSAP (Methylation-Sensitive Amplified Polymorphism)

This tool is a modification of the AFLP technique and was developed for monitoring the state of genomic DNA methylation. Genomic DNA is double-digested with one of the methylation-sensitive enzymes *Hpa*II or *Msp*I and then with the methylation-insensitive *Eco*RI. The resulting fragments are ligated with the corresponding double-stranded adapters, a first preselective amplification is carried out followed by a selective amplification step.

MSAP was first developed to determine DNA methylation events in dimorphic fungi (Reyna-Lopez et al. 1997) and later adapted for the detection of cytosine methylation in the rice genome (Xiong et al. 1999), pepper (Portis et al. 2004), apple (Xu et al. 2000), and Siberian ginseng (Chakrabarty et al. 2003).

### 2.4.1.17 LAMP (Loop-Mediated Isothermal Amplification)

Loop-mediated isothermal amplification (LAMP), which amplifies target nucleic acids with high specificity, efficiency, and rapidity under isothermal conditions was developed by Notomi et al. (2000). This method relies on auto-cycling strand displacement DNA synthesis performed by a DNA polymerase with high strand displacement activity. A specially designed set of two inner and two outer primers is used. An inner primer containing sequences of the sense and antisense strands of the target DNA initiates LAMP. The following strand displacement DNA synthesis primed by an outer primer releases a single-stranded DNA. This serves as template
for DNA synthesis primed by the second inner and outer primers that hybridize to the other end of the target, which produces a stem–loop DNA structure. In subsequent LAMP cycles, one inner primer hybridizes to the loop on the product and initiates displacement DNA synthesis, yielding the original stem–loop DNA and a new stem–loop DNA with a stem twice as long. The cycling reaction continues with accumulation of $10^9$ copies of target in less than an hour. The final products are stem–loop DNAs with several inverted repeats of the target and cauliflower-like structures with multiple loops formed by annealing between alternately inverted repeats of the target in the same strand. Because LAMP recognizes the target by six distinct sequences initially and by four distinct sequences afterwards, it is expected to amplify the target sequence with high selectivity (Nagamine et al. 2001, 2002).

This technique was applied to detect *Panax ginseng* (Sasaki et al. 2008), the botanical source of Ginseng (Ginseng Radix), and to distinguish this species from *Panax japonicus*. It was also used for the detection of *Lophophora williamsii* (Sasaki et al. 2009) and *Curcuma longa* (Sasaki and Nagumo 2007).

### 2.4.1.18 SDA (Subtracted Diversity Array)

PCR-based plant identification techniques are often limited by their low throughput, whereas hybridization-based microarray technology represents a rapid and high-throughput tool for genotype identification. Using an innovative technique, a “Subtracted Diversity Array” (SDA) was constructed from a pooled genomic DNA library of 49 angiosperm species, from which pooled non-angiosperm genomic DNA was subtracted (Jayasinghe et al. 2007). The subtraction was carried out using the Clontech PCR-Select cDNA Subtraction Kit. This new SDA method was shown to be superior to conventional molecular identification methods in terms of accuracy, sensitivity, and efficiency, as well as capacity for high-throughput and broad application. The SDA technique was validated for potential genotyping use. The results indicated a successful subtraction of non-angiosperm DNA. This study demonstrates the potential of establishing a highly informative, reliable, and high-throughput microarray-based technique for the novel application of sequence-independent genotyping of major angiosperm clades.

Niu et al. (2011) showed that SDAs technique is suitable to differentiate two ginseng species, *Panax ginseng* and *Panax quinquefolius*, that are frequently mixed for adulteration. Further, SDA was sensitive enough to detect a deliberate adulteration of 10% *P. quinquefolius* in *P. ginseng*. Thirty-nine species-specific features including 30 *P. ginseng* specific and nine *P. quinquefolius* specific were obtained. This resulted in a feature polymorphism rate of 10% from the 376 features used for fingerprinting the two ginseng species. The functional characterization of 14 *Panax* species-specific features by sequencing revealed one putative ATP synthase, six putative uncharacterized proteins, and two retroelements to be different in these two species.
2.4.1.19 MLPA (Multiplexed Ligase-Dependent Probe Amplification)

An assay well suited to medicinal plant species identification is the Multiplexed Ligase-dependent Probe Amplification (MLPA) assay (Barthelson 2009; Shen and Wu 2009). MLPA is a semi-quantitative PCR-based technique initially developed by Schouten et al. (2002). It uses the sensitivity of the polymerase chain reaction, but increases the specificity by including a key ligation step for those MLPA probes that hybridize to a DNA sequence. Several key features distinguish this technique from other PCR-based techniques. First, the amplification is ligation dependent as amplification of non-ligated oligos does not take place. Since ligation occurs at high temperatures, specificity is further ensured. Second, the amplification is highly multiplexed, allowing the detection of up to 50 targets in one single tube assay. This further allowed mutation testing at the exonic and sub-exonic level for a single gene or multiple genes in one assay, thus becoming a cost-effective medium-throughput test. Third, although this is not a real-time PCR because the primers are in excess of templates and the amplification is in linear range, the amount of amplicons generated at the end is in proportion to the templates (ligation products). Last and more importantly, a common PCR primer is used for the amplification of all target sequences, which is a key feature to ensure the relative quantification of each target with respect to a control sample. Due to its low costs, excellent sensitivity, reliability, and ease of development and implementation, the MLPA technique has become a very popular research and diagnostic tool.

2.4.1.20 Real-Time PCR

Real-time quantitative PCR allows the sensitive, specific, and reproducible quantitation of nucleic acids. Since its introduction, real-time quantitative PCR has revolutionized the field of molecular diagnostics and has become the main technical platform for nucleic acid detection in research and development, as well as in routine diagnostics (Klein 2002; Arya et al. 2005). The intention of real-time PCR is the detection of a specific DNA sequence in a sample by measuring the accumulation of amplified products during the PCR using fluorescent technology. Consequently it allows the researcher to better determine the amount of starting DNA in the sample before the amplification process. Present day real-time methods generally involve fluorescence labeling to show the amount of DNA present at each cycle of PCR. In a few cases real-time PCR has recently been used in identification of Chinese medicinal plants (Matsuyama and Nishi 2011; Slanc et al. 2006).

DNA sequence analysis of rDNA internal transcribed spacer (ITS) and fluorescence melting curve analysis using real-time PCR were applied for authentication of the traditional Chinese medicinal plant *Cimicifuga foetida* from four substitutes: *C. heracleifolia*, *C. dahurica*, *C. acerina*, and *C. simplex*. According to the melting temperature—which is a function of the GC/AT ratio, length, and nucleotide sequences of the amplified product—*C. foetida* was differentiated from all other
species (Ying et al. 2009). Based on real-time PCR technology there are further studies for authentication in the genera *Euphorbia* (Xue et al. 2008a), *Gentiana* (Xue et al. 2008b), and *Drynaria* (Xue and Xue 2008) species from adulterants.

The ongoing development of quantitative DNA-based methods using real-time PCR could enable a quantitative analysis of species composition in mixed plant materials and products in the future (Table 2.2).

### 2.5 DNA Sequencing Analysis

DNA sequencing is the process of determining the precise order of the nucleotides or bases (A,T,G,C) in a particular DNA molecule. The most common approach used for DNA sequencing is the dideoxy or Sanger method which was developed in the mid 1970s. It mimics the basic process used to copy DNA in a cell during chromosomal replication, except that the procedure is done in a tube or microtiter plate using a minimal set of components. Normally the length of the “sequence read” can vary from about 50 to more than 1,000 bases. If the region to be sequenced exceeds the length of a typical sequencing read internal primers have to be used to generate overlapping in order to reconstruct the complete sequence of a longer DNA region. Most large-scale DNA sequencing facilities use fluorescent dyes (rather than a radioactivity isotope) to label and detect the four bases, and capillary electrophoresis to separate DNA molecules on the basis of size. Because the end of each terminated molecule contains a dye-labeled base, the sequence of the strand complementary to the template can be determined. Modern Capillary Sequencer can run plates of 96 samples in a couple of hours and can produce read lengths of more than 700 bases. Using software provided by the manufacturers the signal of the dyes is determined for each position so that the proper base can be identified. The order of the bases is displayed in a chromatogram or trace file.

In the last decade, instrumentation for DNA sequencing has improved dramatically in terms of increasing read length and accuracy, high throughput, and decreasing costs. Meanwhile there are several next-generation DNA sequencing platforms, such as Roche’s (454) pyrosequencing system, Illumina’s Solexa Genome Sequencer or Applied Biosystem’s SOLiD Genome Sequencer.

Currently DNA sequencing is applied in various fields as in analysis of phylogenetic relationship, population genetics, systematics, and evolution (Baldwin et al. 1995). DNA polymorphisms are revealed by determining the nucleotide sequence in a defined region of the genome and aligning the sequence with homologous regions of related organisms (Alvarez and Wendel 2003).

By choosing appropriate regions of the nuclear, plastidal, or mitochondrial genome this approach provides a highly reproducible analysis at various taxonomic ranks to differentiate TCM plants from its substitutes or adulterants.

In order to ensure a correct species identification based on DNA sequence data it is necessary to have herbarium specimen for verification or a reliable database that guarantees that the reference specimen was correctly identified by a taxonomic
<table>
<thead>
<tr>
<th>Technique</th>
<th>Importance for authentication</th>
<th>Reproducibility</th>
<th>Quantity of DNA required</th>
<th>Level of polymorphism</th>
<th>Locus specificity</th>
<th>Technical demand</th>
<th>Sequence information required</th>
<th>Automation</th>
<th>Running costs</th>
<th>Development costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAPD</td>
<td>+</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>No</td>
<td>Low</td>
<td>No</td>
<td>Yes</td>
<td>Low/medium</td>
<td>Low/medium</td>
</tr>
<tr>
<td>ISSR</td>
<td>++</td>
<td>Medium/high</td>
<td>Low</td>
<td>Medium</td>
<td>No</td>
<td>Low/medium</td>
<td>No</td>
<td>Yes</td>
<td>Low/medium</td>
<td>Low/medium</td>
</tr>
<tr>
<td>AFLP</td>
<td>++</td>
<td>High/medium</td>
<td>Medium</td>
<td>Medium</td>
<td>No</td>
<td>Low</td>
<td>No</td>
<td>Yes</td>
<td>Medium</td>
<td>Low</td>
</tr>
<tr>
<td>RFLP</td>
<td>+</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
<td>Yes</td>
<td>High</td>
<td>Yes</td>
<td>No</td>
<td>High</td>
<td>Medium/High</td>
</tr>
<tr>
<td>ARMS</td>
<td>++</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Yes</td>
<td>Low</td>
<td>Yes</td>
<td>Yes</td>
<td>Low/medium</td>
<td>Medium</td>
</tr>
<tr>
<td>SSR</td>
<td>+</td>
<td>High/medium</td>
<td>Low</td>
<td>High/medium</td>
<td>No</td>
<td>Low/medium</td>
<td>Yes/no</td>
<td>Yes</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>CAPS</td>
<td>++</td>
<td>High</td>
<td>Low</td>
<td>Low/medium</td>
<td>Yes</td>
<td>Low/high</td>
<td>Yes</td>
<td>Yes</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>SCAR</td>
<td>++</td>
<td>High</td>
<td>Low</td>
<td>Low/medium</td>
<td>Yes</td>
<td>Low/medium</td>
<td>Yes</td>
<td>Yes</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>RAMPO</td>
<td>+</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>Yes</td>
<td>Medium/High</td>
<td>No</td>
<td>Yes</td>
<td>Low/medium</td>
<td>Medium</td>
</tr>
<tr>
<td>CAPS</td>
<td>++</td>
<td>High</td>
<td>Low</td>
<td>Low/medium</td>
<td>Yes</td>
<td>Low/medium</td>
<td>No</td>
<td>Difficult</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>SSCP</td>
<td>+</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Yes</td>
<td>Medium/High</td>
<td>No</td>
<td>Yes</td>
<td>Low/medium</td>
<td>Medium</td>
</tr>
<tr>
<td>LAMP</td>
<td>+</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Yes</td>
<td>Medium</td>
<td>Yes</td>
<td>Yes</td>
<td>Low/medium</td>
<td>Medium</td>
</tr>
<tr>
<td>SNP</td>
<td>+</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Yes</td>
<td>Medium</td>
<td>Yes</td>
<td>Yes</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Sequencing DNA-barcoding</td>
<td>+++</td>
<td>High</td>
<td>Low</td>
<td>Low/medium</td>
<td>Yes</td>
<td>Medium</td>
<td>Yes</td>
<td>Yes</td>
<td>Medium</td>
<td>Medium</td>
</tr>
</tbody>
</table>

The table is based on the most frequent molecular tools and does not represent all DNA-based molecular techniques. Strength and weakness of DNA methods are also summarized by Weising et al. (2005), Yip et al. (2007), Pereira et al. (2008), Hao et al. (2010), Mondini et al. (2009), and Agarwal (2008).
expert. Additionally the sequence should be obtained in independent studies including related taxa. A common way to assign a particular sequence to a taxon is to perform a BLAST search (Basic Local Alignment Search Tool) in the databases of NCBI (GenBank), BOLD, MMDBD. However, care must be taken when assigning the questioned sequence to the species with the highest similarity, because several gaps and false sequences are known to be present in these databases (Heubl 2010).

There are many studies concerning the application of DNA sequence-based markers to differentiate medicinal taxa used in TCM from its substitutes or adulterants.

Sequencing analyses based on nuclear ITS have been applied to Panax (Ngan et al. 1999; Kim et al. 2007b), Asarum (Kelly 1998; Liu et al. 2005; Yamauchi et al. 2007), Astragalus (Dong et al. 2003; Yip and Kwan 2006), Dendrobium (Xu et al. 2006; Ding et al. 2002; Lau et al. 2001; Zhang et al. 2003), Fritillaria (Wang et al. 2005), Leonurus (Yang et al. 2006), Perilla (Luo et al. 2006b), Phyllanthus (Lee et al. 2006b), Rehmannia (Albach et al. 2007), Salvia (Wang et al. 2005), Swertia (Xue et al. 2006a), Plantago (Sahin et al. 2007), Bupleurum (Yang et al. 2007), and Euphorbia (Xue et al. 2006b).

Another frequently used marker is the nuclear 5S rDNA intergenic spacer used for authentication of Adenophora (Zhao et al. 2003a), Aconitum (Carles et al. 2005), Angelica (Zhao et al. 2003b), Astragalus (Dong et al. 2003; Ma et al. 2000), Curcuma (Xia et al. 2005), Epimedium (Sun et al. 2004), Fritillaria (Cai et al. 1999), Crocus (Ma et al. 2001), Ligularia (Zhang et al. 2005b), Pueraria (Lee and Wen 2004), Rheum (Yang et al. 2001) and Ephedra (Long et al. 2004), rpl16-rpl14 spacer in Scutellaria (Xue et al. 2006b), atpF-atpA in Angelica (Hosokawa et al. 2006), trnD-trnT in Dyosma (Gong et al. 2006), trnK in Actinidia (Zhao et al. 2007c), Atractylodes (Mizukami et al. 2000) and Curcuma (Sasaki et al. 2002), matK in Agastache (Luo et al. 2002), Panax (Komatsu et al. 2001), rbcL in Dryopteris (Zhao et al. 2007c), Cnidium (Kondo et al. 1996), or Pinellia (lin et al. 2006).

Recently sequencing analysis has been applied to proof the possibility of DNA-based authentication of plant extracts. Internal transcribed spacer (ITS) was successfully amplified from different extracts types from Echinacea species and Matricaria chamomilla (Novak et al. 2007).
2.6 DNA Barcoding

DNA barcoding, a term first created by Hebert et al. (2003a) is a novel molecular and bioinformatical tool designed to provide rapid, accurate, automatable, and cost-effective identification of species. Contrary to other molecular methods, it can be used on a large scale and with high reliability. For DNA barcoding, the unique nucleotide sequence patterns of small DNA fragments (400–800 bp) are used as specific reference collections to identify specimens and to discover cryptic taxa (Vijayan and Tsou 2010).

DNA barcoding uses a short genetic marker from a standard locus (alternatively from nuclear, mitochondrial, or plastidal DNA) of an organism. An ideal and successful DNA barcode marker should be suitable for a wide range of taxa (breadth of taxonomic application), routinely retrievable with a universal primer pair, be short enough to be accessible to bidirectional sequencing, and provide a unique sequence for maximal discrimination among species which means high variation between species but conserved within the species, so that the intra-specific variation will be insignificant (CBOL 2009). Additionally these markers should be flanked by evolutionary conserved regions so that universal primers can be used and they should be free of insertions or deletions to be easily alignable (Shneyer 2009) (Fig. 2.1).

2.6.1 Barcoding and Herbal Monographs

DNA authentication of medicinal species used in TCM is of major importance because there are many cases in the Chinese Pharmacopoeia where several species are listed under one common name or listed as synonyms, subspecies, or varieties in herbal monographs.

A prime example is the genus *Dendrobium* (Orchidaceae) which is represented in the Flora of China by 78 species (14 endemic). Overlapping distribution ranges coupled with high morphological variation makes proper species identification difficult. Based on processing methods, Herba Dendrobii is classified into fresh *Dendrobium*, “Fengdou Shihu,” and “Huangcao Shihu,” the latter being the predominant form of Herba Dendrobii on Chinese herbal medicine market.

additionally processed. A survey of the medicinal *Dendrobium* species revealed that in total 27 other *Dendrobium* species are also used clinically as substitutes for “Huangcao Shihu” at various places in China, a situation which may cause inconsistent therapeutic effects or even endanger safety of consumers (Li et al. 1986, 1991; Ma et al. 1995). The stems of *Dendrobium* species, have very similar morphological and anatomical characteristics, and the traditional authentication

<table>
<thead>
<tr>
<th>DNA segment tested for suitability</th>
<th>Proposed/recommended DNA markers</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>rpoC1, rpoB, matK, trnH-psbA, nrITS, trnL-F</td>
<td>rpoC1 + rpoB + matK or rpoC1 + matK + trnH-psbA</td>
<td>Chase et al. (2007)</td>
</tr>
<tr>
<td>TrnL (UAA) intron</td>
<td>trnL (UAA) intron</td>
<td>Taberlet et al. (2007)</td>
</tr>
<tr>
<td>matK, trnH-psbA, psbK-psbI, atpF-atpH</td>
<td>matK or matK + trnH–psbA + psbK-psbI</td>
<td>Lahaye et al. (2008a)</td>
</tr>
<tr>
<td>accD, matK, trnH-psbA, rbcL, rpoB, rpoC1, ycf5, ndhJ</td>
<td>matK or matK + trnH–psbA</td>
<td>Lahaye et al. (2008b)</td>
</tr>
<tr>
<td>trnH-psbA, rbcL, rpoC, CO1, rpoB, matK</td>
<td>trnH-psbA + rbcL or rpoB or rpoC</td>
<td>Erickson et al. (2008)</td>
</tr>
<tr>
<td>psbA-trnH, matK, rbcL, rpoC1, ycf5, ITS2, and ITS</td>
<td>ITS2 + psbA-trnH</td>
<td>Chen et al. (2010)</td>
</tr>
</tbody>
</table>

Proposed markers are based on universality, success of PCR amplification, interspecific sequence variability, and species discriminatory power/identification capability. Several proposed combinations of loci have been summarized by Pennisi (2007), Shneyer (2009), and Vijayan and Tsou (2010). The combination *rbcL* + *trnL-F* (intron and IGS) has been proposed as a two-locus barcode for ferns (De Groot et al. 2011). The best performing single loci for mosses were the *rbcL* and *rpoC1* coding regions (Liu et al. 2010b)
of different “Huangcao Shihu” samples is therefore far from reliable. In addition, the chemical constituents of many *Dendrobium* species are still unknown, as proper chemical analysis methods have not been developed. However, the determination of the botanical origins of different “Huangcao Shihu” samples and their quality control through morphological and chemical studies is fraught with difficulty. Consequently DNA barcoding may offer an alternative method for the identification of the used *Dendrobium* species for herbal medicines (Xu et al. 2006; Takamiya et al. 2011).

A further remarkable example where multiple species are used in TCM is the genus *Fritillaria* (Liliaceae) which includes 24 species (15 endemic) in China. Bulbs of various *Fritillaria* species (Bulbus Fritillariae) are among the most popular herbal medicines in China and have been used as antitussive and expectorant herbs. According to the Chinese Pharmacopoeia, herbal “Beimu” is derived from the bulbs of nine *Fritillaria* species. These include *Fritillaria thunbergii* Miq. (“Zhebeimu”), *F. cirrhosa* D. Don. (Chuanbeimu), *F. unibracteata* Hiao et Hsia, *F. przewalskii* Maxim ex Batal, *F. delavayi* Franch, *F. ussuriensis* Maxim. (Pingbeimu), *F. walujewii, F. pallidiflora* Schrenk (“Yibeimu”), and *F. hupehensis* Hsiao et K.C (“Hubeibeimu”).
Furthermore other *Fritillaria* species (e.g., *Fritillaria anhuiensis* S.C. Chen et S.F. Yin, *Fritillaria puqiensis* G.D. Yu et G.Y. Chen) are often used in different regions in China as substitutes. These distinctions can be very confusing for consumers. As there are no specific microscopic characteristics and the pattern and concentration of chemical components are often unstable, for quality control and standardization of *Fritillaria* derived herbal medicine, exact species identification using DNA barcoding is the only reliable method (Wang et al. 2007b).

Another example where the Chinese Pharmacopoeia allows some interchangeable use of herbals is the genus *Epimedium* (Berberidaceae) which is one of the most popular herbal drug (“Yinyanhuo”) comprising 41 species (40 endemic) in China. Herba Epimedii is prepared from aerial parts of five *Epimedium* species listed in the Chinese Pharmacopoeia: *E. brevicornum* Maxim., *E. koreanum* Nakai, *E. sagittatum* (Sieb. et Zucc.) Maxim., *E. pubescens* Maxim., and *E. wushanense* T. S. Ying. In common with many traditional medicinal plants, a major problem with *Epimedium* is the absence of a rigorous method to authenticate species. Taxonomists have variably reported numbers ranging from 20 to 50 species (Sun et al. 2005). Traditional herbalists do not differentiate among *Epimedium* species, but rather use a mixture of species together as *Herba Epimedii*. These species differ significantly in concentrations of major and minor constituents (Wu et al. 2003). The extensive variation in morphological traits make the species difficult to classify taxonomically and there is some confusion due to lack of scientific research.

Considering the examples mentioned above there are numerous monographs in the Chinese Pharmacopoeia where multiple species (further examples *Dioscorea* and *Uncaria* each with five species listed) are a legitimate source for the preparation of herbal medicines. As these species have a different composition of bioactive compounds and are quantitatively and qualitatively not equivalent, the therapeutic efficacy of a TCM medication can vary to a high extent. For species-rich genera lacking morphological characters or taxa of controversial botany, DNA barcoding as a new tool is a prerequisite to clearly distinguish the authentic herb from the substitute/adulterant. If the barcode DNA-analysis of all frequently used Chinese drugs becomes available in the near future, the application of this method and further DNA fingerprinting techniques should be included in monographs to optimize the quality and safety proof of the drugs (Fig. 2.2).

### 2.6.2 Standard Markers for Barcoding

In certain animal groups the mitochondrial cytochrome *c* oxidase subunit 1 (CO1, *cox1*) gene sequence (ca. 650 bp in length) is currently used as a universal DNA barcode (Hebert et al. 2003b; Waugh 2007). It has been introduced as standard marker as it can be easily sequenced and provides greater than 97 % species-level specificity in different animal phyla, e.g., birds (Hebert et al. 2004), mammals (Luo et al. 2011), fishes (Ward et al. 2009), amphibians (Vences et al. 2005), and various groups of arthropods (Virgilio et al. 2010). Based on this mitochondrial
DNA barcoding is progressing rapidly in many groups of animals, which can be reviewed online via the Canadian Barcode of Life (http://www.bolnet.ca) and the Consortium for the Barcode of Life (CBOL; http://www.barcoding.si.edu). Other alternative mitochondrial regions such as cytochrome b, 12S, 18S (SSU), 28S (LSU) rRNA, and the nuclear ITS2 region have been proposed (but not established) for animal identification (Wong et al. 2001; Chen et al. 2010; Ferri et al. 2009; Yao et al. 2010). However, the cytochrome c oxidase marker, which is widely used in animal barcoding, seems promising only for algae (Robba et al. 2006; Lane et al. 2007) but it is not suitable for higher plants because of a much slower substitution rate in cox1 which is two- to threefold lower than in chloroplast genes and 10- to 20-fold lower than in nuclear genes (Wolfe et al. 1987; Drouin et al. 2008). Furthermore the mitochondrial genome in plants is characterized by the occurrence of large structural rearrangements and often nonfunctional copies of mitochondrial genes (pseudogenes) which cause erroneous results. Consequently in plants appropriate barcoding regions have been analyzed, tested, and selected for discriminating

Fig. 2.2 DNA sequence BLAST search against the databases of GenBank (NCBI) and BOLD using the barcoding markers ITS and rbcL + matK sequences of Artemisia annua as a query. NCBI (http://www.ncbi.nlm.nih.gov) offers the BLAST search tool to perform fast searching with rigorous statistical methods for judging the significance of matches. The Barcode of Life Data Systems (BOLD; http://www.boldsystems.org/views/login.php) and its Identification System (IDS) for rbcL and/or matK is an identification tool for plant species. The ID engine uses all sequences uploaded to BOLD to locate the closest match. It returns a species-level identification when possible.
among the 260,000–422,000 species of seed plants (Govaerts 2001; Thorne 2002),
but until now no agreement exists which marker is the most promising, though most
researchers agree that more than one region is necessary. Consequently a multilocus
approach based on the plastid genome with a focus on coding and noncoding
regions (introns or spacers) is currently the most effective strategy for species
identification in plants (Chase et al. 2005, 2007; Kress et al. 2005; Newmaster
et al. 2006, 2008; Cowan et al. 2006; Ford et al. 2009; Kress and Erickson 2007;
Lahaye et al. 2008a, b; Fazekas et al. 2008).

In order to promote the use of DNA barcoding for all eukaryotic organisms, the
Consortium for the Barcode of Life (CBOL, http://barcoding.si.edu) was established
in 2004 at the National Museum of Natural History in Washington, which currently
includes more than 120 organizations from 45 nations.

In 2005 the Plant Working Group (PWG CBOL) reported that five very
promising regions were chosen for further barcoding studies. These regions were
\( \text{matK} \) (encodes a maturase and is located within the \( \text{trnK} \) gene), \( \text{rpoB} \) (RNA
polymerase subunit), \( \text{rpoC1} \) (RNA polymerase subunit), \( \text{accD} \) (subunit of acetyl-
CoA carboxylase), and \( \text{ccsA} \) (previously known as \( \text{ycf5} \); the gene encoding a protein
involved in cytochrome \( \text{c} \) biosynthesis).

In a subsequent work in 2007, additional candidate plastid DNA regions were
tested in different groups (gymnosperms, angiosperms, ferns, equisetens, and
mosses).

Based on assessments of recoverability, sequence quality, and levels of species
discrimination, a core two-locus combination of \( \text{rbcL} \) and \( \text{matK} \) as the plant barcode
was recommended. This combination was shown to successfully discriminate
among 907 samples from 550 species at the species level with a probability of
72 % (CBOL Plant Working Group 2009). The group admits that the two-locus
barcode is far away from perfection due to the limited identification rate, and thus
further research for other appropriate candidates is necessary. Additional
combinations of noncoding and coding plastid regions have been tested for
barcoding purposes (Fazekas et al. 2008; Pennisi 2007; Ford et al. 2009).

Other combinations involving three plastid regions have also been proposed by
various working groups that include the Royal Botanical Gardens, Kew, UK http://
www.rbgkew.org.uk/barcoding (Chase et al. 2007).

In view of the available markers as a core barcode to identify land plants a
combination of two loci from plastidal DNA, \( \text{rbcL} + \text{matK} \) was adopted as “the
plant barcode” by the Executive Committee of the Consortium for the Barcoding of
Life (CBOL Plant Working Group 2009) after much deliberation (e.g., Kress et al.
2005; Hollingsworth et al. 2009). They agreed that \( \text{rbcL} \) and \( \text{matK} \) are approved and
the most suitable barcode regions for land plants. Based on these markers, the
BOLD Identification System (BOLD-IDS) provides a species identification tool
that accepts DNA sequences with a minimum sequence length of 500 bp from these
two barcode region and returns a taxonomic assignment at the species level (a list of
the nearest matches) when possible. The group admits that the two-locus barcode is
far away from perfection due to the low identification rate and that the search is not
over.
Additionally the plastidal *psbA-trnH* intergenic spacer and internal transcribed spacer region (ITS1-5.8S rRNA-ITS2) of nuclear ribosomal DNA have been tested as supplementary barcodes. Kress et al. (2005) suggested that these two noncoding regions might have potential as universal plant barcodes. Generally the ITS, which is part of the ribosomal operon, is organized in large blocks (tandem arrays of nearly 50–100 copies) in the chromosome of the nuclear organizer region. It is one of the most commonly sequenced DNA regions (ca 700 bp) used in plant phylogenetic studies at the generic and species level. The advantage of these spacers lies in their high variability. They are adjacent to the conserved 5.8S rRNA region and flanked by conserved 18S and 26S rRNA genes, which facilitate the development of primers. Meanwhile more than 50,000 plant ITS sequences have been deposited in GenBank (Hajibabaei et al. 2007). As nuclear ITS is arranged in multiple copies, paralogs may occur due to hybridization and subsequent incomplete concerted evolution. As a consequence in some groups the presence of several functional copies was detected (Alvarez and Wendel 2003). Not only the sequence is an effective barcode but also the secondary structure of the ITS2 region could provide useful information for species identification and has potential as a molecular morphological characteristic (Yao et al. 2010).

The *trnH-psbA* spacer, though short in length (ca. 450 base pairs), is one of the most variable plastid regions in angiosperms and can be easily amplified across a broad range of land plants (Shaw et al. 2007). In general intergenic spacers and introns tend to be more variable than genes and are therefore considered as better identifiers.

Recently, Chen et al. (2010) compared seven candidate DNA barcodes from Chinese medicinal plant species. This study including a total of 4,800 species from 753 distinct genera revealed that the Internal Transcribed Spacer 2 (ITS2) of nuclear rDNA and the *psbA-trnH* locus are promising universal barcodes for plant identification. A test of the discrimination ability of ITS2 showed that the rate of successful identification at species level was 92.7 % for ITS2 and 72.8 % for *psbA-trnH*.

Shaw et al. (2007) compared chloroplast genomes of Solanaceae, Fabaceae, and Poaceae and showed that no less than nine intergenic spacers, which were practically not used in molecular phylogenetic studies (*rpl32-trnL, trnQ-rps16, trnV-ndhC, ndhF-rpl32, psbD-trnT, psbl-petA, rps16-trnK, atpI-atpH, petL-psiE*), were far more variable than the most variable ones of those spacers used earlier. Probably these regions have potential as DNA barcoding markers, if not universal, then suitable for studies on higher taxonomic level.

Moreover, further DNA barcode markers have been used to identify medicinal plants from different plant families and genera (Chen et al. 2010; Song et al. 2009; Luo et al. 2010; Pang et al. 2010; Gao et al. 2010a; Sun et al. 2010; He et al. 2010) including also medicinal Pteridophytes (Ma et al. 2010) (Table 2.3).

In 2008 several DNA barcoding projects have been initiated in China when the country became a central node of the International Barcode of Life project (iBOL). The Barcoding Chinese Plants Project is closely associated with the Germplasm Bank of Wild Species (GBOWS, Kunming Institute) which was founded to
safeguard and barcode about 6,000 species of vascular plants of the flora of China. Based on a nationwide seed collecting program a large data set, involving 6,286 individuals representing 1,757 species in 141 genera of 75 families of seed plants, was assembled by the China Plant BOL Group. Comprehensive analysis of the dataset was made to assess the universality, sequence quality, and discriminatory power of the chosen barcoding markers (Li et al. 2011a).

Meanwhile an integrated DNA barcode database (MMDBD) is under construction (Lou et al. 2010). This platform contains ca. 1,300 species of Chinese medicinal plants listed in the Pharmacopoeia of the People’s Republic of China and it provides information on storage, retrieval, comparison, and analysis of DNA sequences, for distinguishing medicinal materials from their common substitutes and adulterants (see http://www.cuhk.edu.hk/icm/mmdbd.htm).

In combination with phylogenetic analyses, in recent years barcoding activities have been intensified in many genera including Chinese medicinal plants. Barcoding projects based on nuclear and/or plastidal markers has been applied in different plant families e.g., Rosaceae (Pang et al. 2011), Caprifoliaceae (Liu et al. 2010a), Asteraceae (Gao et al. 2010b), Loranthaceae (Li et al. 2009b), Euphorbiaceae (Pang et al. 2010), Fabaceae (Gao et al. 2010b), Polygonaceae (Song et al. 2009), Rutaceae (Luo et al. 2010), Myristicaceae (Newmaster et al. 2008), Lemnaceae (Wang et al. 2010b), Lamiaceae (Han et al. 2009), as well as in the genera *Panax* (Zuo et al. 2011), *Dendrobium* (Asahina et al. 2010), *Lonicera* (Sun et al. 2010), *Paris* (Zhu et al. 2010), *Aconitum* (He et al. 2010), *Taxillus* (Li et al. 2010c), *Amomum* (Zhen-Yan and Ling 2010), *Astragalus* (Guo et al. 2010), *Paeonia* (Zhang et al. 2009), in Cycadales (Sass et al. 2007), and also in medicinal Pteridophytes (Ma et al. 2010).

Molecular authentication of herbal medicinal materials has increased enormously in the last decade and its advantages are undisputed. This new tool has been included for the first time in the *Pharmacopoeia of the People’s Republic of China* and its online Supplementary Note 2 as a standard method for the identification of some traditional Chinese medicines (Li et al. 2011b).

Once fully developed, DNA barcoding has the potential to completely revolutionize our knowledge of plant diversity. In the future scientists will be able to quickly identify known species and retrieve information about them. Compiling a public library of sequences linked to vouchered specimens and their scientific binomial names will make barcoding a practical and efficient tool for the identification of species, including all Chinese medicinal plants (Fig. 2.3).

### 2.7 DNA Microarrays (DNA Chip Technology)

The DNA chip technology developed by Fodor et al. (1991) enables the production of a “biochip” designed to identify fluorescent-labeled DNA or RNA fragments through their hybridization to oligonucleotide probes. DNA microarrays are a high-throughput technology for simultaneous analysis of multiple genes in many taxa or
Fig. 2.3 (continued)
samples (Fodor et al. 1993; Gershon 2002). To apply this technique for identification and authentication of herbal material, it is necessary to identify a distinct DNA sequence that is unique to each species (Preeti et al. 2006a, b). The DNA sequence information is then used to synthesize a corresponding probe on a chip. These probes are capable of detecting complementary target DNA sequences if present in the test sample being analyzed. These immobilized DNA fragments are arranged in

Fig. 2.3 Steps in the identification process using the platform MMDBD (Medicinal Materials DNA Barcode Database—http://137.189.42.34/mherbsdb/index.php). Search and result page with sequence similarity BLAST search using the \textit{rbcL} sequence of \textit{Artemisia annua} for query (steps 1–3). Information page for medicinal materials including herb name, species name, family name, medical part, pharmacopoeia information, status in CITES, adulterant, DNA sequence (with voucher), and key reference (step 4)
a regular pattern on a microarray by fixation on glass slides, silicon, or nylon membranes (Gebauer 2004).

DNA extracted from the target sample and labeled with a specific fluorescent molecule is then hybridized to the microarray DNA. A positive hybridization result is detected by the intensity of the fluorescence, which reflects the stability of the hybridization between the oligonucleotide probe and the target sequence, and it is visualized with fluorescence scanning or imaging equipment. Each hybridization surface may contain a very large number of unique oligonucleotide probes (up to 400,000), permitting several thousand individual nucleotide positions to be characterized at the same time. The DNA microarray field is a combination of several technologies, including automated DNA sequencing, DNA amplification by PCR, oligonucleotide synthesis, nucleic acid labeling chemistries, and bioinformatics.

Recently this technique has been applied for the identification of various species of *Fritillaria* (Tsoi et al. 2003), *Dendrobium* (Li et al. 2005; Zhang et al. 2003), and *Bupleurum* (Lin et al. 2008). Previously, the nucleotide sequences of the nuclear 18S rRNA gene of 13 *Panax* taxa were determined. On the basis of the nucleotide differences, a DNA microarray (PNX array) was developed for the identification of various *Panax* drugs (Zhu et al. 2008). A silicon-based DNA microarray was designed and fabricated for the identification of toxic traditional Chinese medicinal plants. Species-specific oligonucleotide probes were derived from the 5S ribosomal RNA gene of *Aconitum carmichaeli*, *A. kusnezoffii*, *Alocasia macrorrhiza*, *Croton tiglium*, *Datura inoxia*, *D. metel*, *D. tatula*, *Dysosma pleiantha*, *Dy. versipellis*, *Euphorbia kansui*, *Hyoscyamus niger*, *Pinellia cordata*, *P. pedatisecta*, *P. ternata*, *Rhododendron molle*, *Strychnos nux-vomica*, *Typhonium divaricatum*, and *T. giganteum* (Carles et al. 2005).

The analyses demonstrated that DNA microarray-based technology can provide a rapid, high throughput tool for correct botanical identification, for authentication of crude plant materials, standardization, and for quality control being used for hundreds of samples simultaneously (Debouck and Goodfellow 1999).

This application of DNA microarrays will not only benefit the herbal drug industry but can also facilitate the identification of herbal products by regulatory authorities (Fig. 2.4).

### 2.8 Limitations of Genetic Markers in Herbal Drug Technology

Molecular authentication methods in comparison to macroscopic, microscopic, and phytochemical analyses have several advantages, which make them suitable for the identification of plants used in TCM. The DNA-based techniques are not affected by environmental factors, independent from the physical form of the plant material, and require only a low amount of material. Although DNA analysis is currently considered to be cutting-edge technology, it has certain limitations:
• The applicability of a DNA-based method depends generally on the quality and quantity of the DNA. When plant drugs are processed under extreme conditions DNA degradation (fragmentation) is observed, making PCR amplification impossible. Drug processes involving mechanical stress, high temperature, pH variations, enzymatic activities, and fermentations affect the primary structure of DNA and cause often hydrolysis, oxidation, and deamination of the DNA. Plant-derived polyphenolics are also capable of oxidative DNA damage particularly in the presence of transition metal ions (Bauer et al. 2003; Malik et al. 2003; Novak et al. 2007).

• The isolation of high quality genomic DNA being essential for molecular applications becomes difficult for many medicinally used plants due to the presence of secondary compounds which serve mainly for chemical defense against herbivores. High concentrations of these secondary compounds (polysaccharides, tannins, essential oils, polyphenolics, alkaloids, etc.) often influence DNA extraction, PCR reaction, or restriction digestion. In tissues of medicinal plants, secondary compounds generally get accumulated and the problem becomes increasingly severe as the material gets older. Polysaccharide contaminations are particularly problematic as they can inhibit the activity of many commonly used enzymes, such as polymerases, ligases, and restriction endonucleases. Polyphenol contaminations of DNA make it resistant to restriction enzymes and interact irreversibly with proteins and nucleic acids. Choosing

![Workflow overview with major steps of an automated and high throughput DNA microarray platform for species identification](image-url)
the most suitable DNA extraction procedure may help to eliminate the PCR inhibitors (Friar 2005; Pirttilä et al. 2001; Fleischmann and Heubl 2010).

- Sometimes plant materials are contaminated with symbiotic or pathogenic microbes. Especially endophytic fungi often occur as symbionts living within the tissues of their angiosperm hosts. DNA isolation techniques for obtaining genomic plant DNA do not discriminate between plant and fungal DNA and PCR primers with broad applicability also amplify DNA contaminants. Finger-printing using RAPD analyses are particularly vulnerable to these fungal contaminants because of the short length of their primers. Although the amount of endofungal DNA is presumed small as compared to the host genomic DNA, researchers need to consider its potential presence. Problems with endophytic fungi can be eliminated with plant-specific primer design for nuclear markers (e.g., ITS) (Saar et al. 2001).
- The orthology of characters is one of the fundamental and implicit assumptions in the use of DNA sequence data to reconstruct phylogeny or to establish barcodes for species. However, some studies revealed the presence of some degree of intra-individual variations among the copies of ITS1 and ITS2 sequences of the nuclear ribosomal cistron (Feliner and Rosselló 2007). Various reasons such as recent hybridization, lineage sorting, recombination among copies, high mutation rate, and pseudogene formation (nonfunctional paralogous) are considered to be the reasons for such variations (Song et al. 2008). Additionally nonfunctional copies of plastidial or mitochondrial pseudogenes, which have been sometimes transferred to the nucleus, have been detected in various eukaryotic organisms, too. For DNA barcoding as a practical molecular method to identify species only orthologous DNA sequences can be used. Consequently cloning of PCR products is sometimes inevitable (Bailey et al. 2003; Buckler et al. 1997).

2.9 Perspectives for Authentication Using DNA Barcoding

The principle of DNA barcoding is that a standardized, fast-evolving short segment of the genome from chloroplast, mitochondrial, or nuclear DNA can be used to quickly and easily identify any organism. For the process of identification and authentication, different aspects of classical taxonomy, laboratory practice, and data management must be considered.

The first step in authentication of medicinal plants is sampling of the correct taxon and determination using identification keys as provided in the Flora of China. The designated species should match the description of the “type” specimen in all relevant diagnostic features. Confirmation of the determined taxon by a plant taxonomist/systematist or expert in the field of pharmacognosy is favorable. The researcher should also give detailed information for all specimens that are used for barcoding, such as latin taxon name, collector, collection date and number, locality, geospatial coordinates (GPS data), as well as information on infraspecific variants
(variety, chemotype, hybrid) in case of cultivated plant material. These data combined with high resolution images should be deposited in public databases (e.g., BOLD, MMDBD). Special emphasis should be placed on unambiguous association of DNA barcodes with individual specimens and corresponding data records. Thus the specimens that are used for DNA barcoding need to be preserved individually, each with a unique identifying number, or sample ID. Consequently deposition of voucher specimens in local herbaria or natural history collections is important for comparison and verification (see sampling instructions BOLD/CCDB).

For barcoding studies it is also essential to reconstruct molecular phylogenies (using software e.g., PAUP, MrBayes, etc.) including a maximum of species of a taxonomic group. There are many genera of medicinal plants in China which exhibit an extraordinary diversity as Astragalus (401 species), Aconitum (211 species), Artemisia (186 species), Berberis (215 species), Clematis (147 species), Polygonum (113 species), or Salix (275 species) and phylogenetic information is limited or missing. Consequently in these genera misidentification is very common. In this respect the knowledge of closely related taxa and sister species relationships, the primary candidates that show low interspecific divergence and often share haplotypes, is of main interest. It is conspicuous that DNA barcoding studies generally neglect or underestimate intraspecific genetic variability and variation in the barcode region is not considered (Moritz and Cicero 2004). Thus phylogenetic reconstructions using comprehensively sampled groups are fundamental to explore intraspecific variation and interspecific divergence. For this purpose adequate sampling of several individuals across the geographic range of a species is required. There is also need to examine groups in more detail with frequent hybridization, recent radiations, introgressions, or incomplete lineage sorting (Li et al. 2011b; China Plant BOL Group).

Concerning laboratory practice and improvements, a cost-effective long-term storage for plant tissue samples (e.g., fresh tissue storage in liquid nitrogen or silica gel desiccation) has to be tested and used for DNA preservation. Furthermore DNA extraction from (old) herbarium specimens, increase of sequencing success, particularly from samples containing long mononucleotide repeats, is required. With application of new polymerases with higher fidelity and processivity or using mixtures that include repair enzymes amplification of degraded DNA can be overcome (Mitchell et al. 2005; Hajibabaei et al. 2005). Additionally development and optimization of DNA barcoding primers for a certain taxonomic group or highly degraded DNAs will greatly improve PCR success and multiplex PCR reactions that can routinely amplify barcode markers will significantly reduce laboratory costs. To add more reliability in the identification of species and to complement the barcoding regions more suitable, nuclear barcode loci (low-copy nuclear regions as ADH, waxy, leafy) must be developed. Future barcoding activities should be focused on the discriminating power of standard markers (rbcL, matK, ITS, trnH-psbA) in species-rich angiosperm families as Asteraceae, Fabaceae, Rosaceae, Ranunculaceae, Lamiaceae, and Apiaceae that contain many important Chinese medicinal plants. Next-generation sequencing technologies will
provide exciting new opportunities for barcoding of multiple samples or even entire plastid genomes that can be used as the ‘next-generation plant barcode’ (Chen et al. 2010; Sucher et al. 2012; Hollingsworth et al. 2011).

In the field of data management, sophisticated tools are required to automatically query segments of individual sequences, to check for different affinities or incongruence within markers, to decode the peaks in over-laid traces of sequencing chromatograms, to check for editing errors or pseudogenes, and to detect microinversions. Furthermore there is demand for improved bioinformatic search routines because sequence identification methods that use local pairwise alignments (e.g., BLAST) are unable to accurately differentiate between highly similar sequences and are not designed to cope with hierarchic phylogenetic relationships or within taxon variability (Little 2011).

Comprehensive databases containing voucher specimens, macro and microscopic data, chemical profiling, and DNA barcoding information would clearly be beneficial for the authentication of medicinal plants and for providing consumers with a safe product. The molecular detection technologies therefore undoubtedly contribute to the research and development of herbal drugs.

2.10 Guidelines for Authentication of Chinese Medicinal Plants

- For ensuring the safety, efficacy, and quality of traditional Chinese herbal medicines classical botanical methodologies have to be applied including collection and conservation of the botanical material (herbarization of plants and silica gel desiccation of leaves for DNA extraction). It is important to supply representative portions of the plant for correct identification, particularly flowering parts, fruits, and seeds. Large specimens should be dissected and mounted to show the main distinguishing characteristics.

- Additionally proper documentation is required: who collected the plant, date of collection, collection site (preferably by GPS coordinates), a unique collection number or code, and habitat information (surrounding vegetation, soil, exposition, and landscape). Supplementary photographs of the growing plant in its habitat should be provided. It is advisable to collect several sets of the same specimen so that, following correct identification, one can retain a specimen for later reference. The plant should be determined by an expert (taxonomist) and given a legitimate Latin binominal name (including genus, species, subspecies/variety, and author) according to the guidelines and rules of the International Code of Botanical Nomenclature (ICBN) (McNeill et al. 2006).

- Online taxonomic sources such as the Flora of China Checklist, The Plant List (TPL), The International Plant Names Index (IPNI), World Checklist of selected plant families or Medicinal Plant Names Index (MPNI) offer access to accepted names. If available, the local and/or English common names and relevant information, such as the cultivar name, ecotype, chemotype should also be provided.
• From the collected and properly dried plant material (one or several representative individuals) a voucher specimen should be prepared by mounting the plant(s) on a herbarium sheet in a format suitable for conserving the morphological key characters (e.g., flowers, fruits) of the specimen. This voucher which serves as a permanent record and reference of an individual plant species in time and space must be labeled with scientific name and detailed collection data (as mentioned above) and deposited in a registered public herbarium, museum, or repository of a certified research institute. It is appropriate to prepare two or more duplicate voucher specimens (“back ups”) that can be sent to taxonomic experts (plant taxonomists) anywhere in the world for confirmation of the species identification. Detailed information on procedures to properly collect, press, and prepare voucher specimens are available (Hildreth et al. 2007; Smillie and Khan 2010; Eisenmann et al. 2012).

• Seeds or other propagation materials should be specified, and all necessary information relating to the identity, quality, and performance of their products, as well as their breeding history should be provided. For the identification of plants grown from seed material it is essential to prepare a voucher specimen. Due to the need for herbal drugs of consistent quality and reliable supply, methods for commercial field cultivation and post-harvest processing should be developed to guarantee high standards with regard to the required bioactive constituents. To achieve genetically improved or new cultivars breeding programs should be developed.

• For chemical standardization it is useful to collect bulk material from the population of the reference sample to reduce potential collecting errors. These plants should be harvested during the appropriate season or time period to ensure the best possible quality of source materials. Collection practices should ensure the long-term survival of wild populations and their associated habitats. Medicinal plants/herbal drugs from species that are listed as endangered (CITES, Convention on International Trade in Endangered Species of Wild Fauna and Flora) must not be collected unless the relevant competent authority has given its authorization.

• Using DNA information for species identification, DNA sequences have to be generated to detect differences in nucleotide positions that are a prerequisite for sequence similarity search in databases. This function allows the user to conduct homology searches between the sequence of interest and the data in the public sequence databases (GenBank; MMDBD; BOLD). To make identification reliable, it is necessary to have complete information of all species of a plant genus. This can only be achieved by conducting phylogenetic reconstructions (using a combination of universal DNA barcoding markers of the nuclear and chloroplast genome) including a comprehensive sampling of species from genera of frequently used TCM plants. Only this molecular phylogenetic approach provides evidence for the identification of genetically closely related species or cryptic taxa. All analyses should include voucher numbers of samples, primer sequences, and PCR conditions for generating barcode sequences. All sequences
should be deposited in DNA databases (GenBank; MMDBD; BOLD) for future data retrieval, similarity search, and species identification.

- The focus of DNA barcoding analyses should be on these medicinal plant species which are worldwide and the most frequently used official Chinese plant drugs. Special attention should be paid to those medicinal plants that are frequently substituted or adulterated with other species that are morphologically and/or phytochemically indistinguishable. Based on available sequence information (barcodes), other diagnostic molecular identification tools (PCR-RFLP, ARMS, SCAR, etc.) with high discrimination power should be developed for easy, cheap, and reliable identification of species, varieties, and cultivars.
- To examine correlations between DNA barcoding and chemical profiling, chromatographic techniques (e.g., TLC, HPLC, GC as well as IR, MS, and NMR spectrometry) should be applied in an integrated strategy using the same particular plant source.
- In view of the amount of DNA barcodes deposited meanwhile in databases, it is highly recommended that molecular authentication should be implemented as a standard method for identification of Chinese herbal medicinal materials in future editions of the Pharmacopoeia of the People’s Republic of China.
- Beyond that interdisciplinary workshops/conferences are necessary to exchange ideas, to highlight perspectives of future research, to stimulate new work in the field of TCM, and to enhance the dialogue between scientists from plant taxonomy, molecular biology, agriculture, phytochemistry, pharmacognosy, and medicine. In addition a global information network on TCM research is required which integrates available databases. A modernization of Traditional Chinese Medicine has become necessary and urgent to increase the worldwide acceptance.

2.11 Future Developments

New innovative automated assays and specific tools for DNA analysis are emerging and will contribute to the next generation of technologies. These are minisequencing (Pastinen et al. 1996; Cai et al. 2000), nanoscale DNA sequencing (Pastinen et al. 1996), Microsphere-based suspension arrays (Lowe 2000), and next-Generation Sequencing (Mardis 2008; Lerner and Fleischer 2010). Further extremely promising developments are the nanopore technology for identification of DNA bases with very high confidence and the arrayed primer extension reaction (APEX) which is an enzymatic genotyping method to analyze hundreds to thousands of variations in the genome simultaneously in a single multiplexed reaction (Fortina et al. 2005; Pirrung et al. 2000). Another upcoming method for large-scale multiplex analysis of nucleic acid sequences is the oligonucleotide ligation assay (OLA), which can be applied for the detection of known single nucleotide polymorphisms (SNPs) and allelic discrimination in highly polymorphic genes (Kurg et al. 2000). These novel approaches to DNA sequencing promise complete genomic analysis and have a
high multiplexing capacity and great potentials for genotyping and future taxon identification (Grossmann et al. 1994).

New high-throughput sequencing (HTS) technologies enable application of new molecular approaches. The recent introduction of massive parallel sequencing technology producing millions of DNA sequence reads (in total 0.5–60 giga base pairs) in a single run has revolutionized genomic research in biology and medicine. These so-called next-generation sequencing platforms, such as Roche/454, Illumina/Solexa, Helicos, and ABI/Solid system can sequence DNA faster and at much lower costs in comparison to the conventional 96-capillary system of Sanger sequencing.

Next-generation sequencing techniques (NGS) can be used to address new and long-standing questions previously inhibited by technological and financial limitations (Kircher and Kelsko 2010). Although no “third-generation” platform has been made commercially available yet, several companies have prototype technologies in active development.

References


Korea Herbal Pharmacopoeia (2002) Ministry of Health Family Welfare of South Korea, Seoul, Korea


Pharmacopoeia of the People’s Republic of China (2005) State pharmacopoeia committee. Chemical Industry Publisher, Beijing, China


Sze SCW, Song JX, Wong RNS, Feng YB, Ng TB, Tong Y, Zhang KYB (2008) Application of SCAR (sequence characterized amplified region) analysis to authenticate Lycium barbarum (wolfberry) and its adulterants. Biotech Applied Biochem 51:15–21


Xu M, Huaracha E, Korban SS (2001) Development of sequence-characterized amplified regions (SCARs) from amplified fragment length polymorphism (AFLP) markers tightly linked to the Vf gene in apple. Genome 44:63–70


Zhao ZZ, Li TYS (2004) Hong Kong commonly confused Chinese medicines. Chinese Medicine Merchants Association, Hong Kong
Zhao KJ, Dong TT, Cui XM, Tu PF, Tsim KW (2003a) Genetic distinction of radix adenophorae from its adulterants by the DNA sequence of 5S-rRNA spacer domains. Am J Chin Med 31:919–926
Evidence and Rational Based Research on Chinese Drugs
Wagner, H.; Ulrich-Merzenich, G. (Eds.)
2013, XII, 525 p. 91 illus., 30 illus. in color., Hardcover
ISBN: 978-3-7091-0441-5