Review

DNA markers in the authentication of Traditional Chinese Medicine

Ying Liu¹, Hannah Tan², Yuanling Qi³, Xinmin Ruan ⁴* and Naiyou Liu⁴,⁵

¹College of Natural and Agricultural Sciences, University of California, Riverside, CA 92521, USA.
²Department of Surgery, University of California, Los Angeles, CA 90095, USA.
³Department of Oncology, Weifang People's Hospital, Shandong 261041, China.
⁴Heart Center, Guangdong Provincial Hospital of Traditional Chinese Medicine, Guangdong University of Traditional Chinese Medicine, Guangzhou 510105, China.
⁵Department of Surgery, Cedars-Sinai Medical Center, David Geffen School of Medicine, University of California, Los Angeles, CA 90048, USA.

Accepted 10 December, 2010

Traditional Chinese medicine (TCM), with its multi-target effects, is gaining more and more attention all over the world due to its specific theory and long historical clinical practice. But the absence of an objective and accurate inspection system is frequently cited as one of the major hurdles for its modernization and globalization. Benefiting from modern molecular biology and polymerase chain reaction (PCR) techniques, DNA markers have become one of the most reliable methods for identification and authentication of Chinese medicinal materials. This paper reviewed the recent progress of DNA markers, including PCR-based markers, hybridization-based markers, sequenced-based markers, and DNA microarray in the authentication of TCM.

Key words: DNA marker, Traditional Chinese medicine (TCM), polymerase chain reaction (PCR).

INTRODUCTION

Traditional Chinese medicine (TCM) is an integral part of Chinese culture. It has been used in China for more than 20 centuries in the prevention and healing of human diseases. Due to its long history of clinical use, reliable therapeutic efficacy, unique treatments, and systemic theories, TCM is getting increasing global attention; furthermore, an increasing variety of natural bioactive compounds have been developed from TCM by many big, worldwide pharmaceutical companies to meet the contemporary trend of "back to nature" (Shaw et al., 2002). However, ensuring the therapeutic effects of TCM is a challenging task because of the nature of multi-component mixtures. Substitutes and adulterants of TCM materials may interfere with their therapeutic effects, even resulting in poisoning. Thus, one major hurdle that might impair TCM's potential future as "medicine of choice" is the lack of standardization.

Traditionally, identification of TCM involves its morphology, histological characteristics, and/or content determination, which needs a lot of time and labor. In recent years, many significant analytical advancements and tools were made in relation to this particular problem. One of the most reliable methods of identification of TCM is DNA analysis. DNA markers for TCM identification are less affected by age, physiological conditions, environmental factors, as well as harvest, storage and processing methods (Jiang et al., 2010). Further, only a small amount of DNA is needed, and the physical form of the sample does not restrict detection. This is extremely useful when the TCMS are expensive or short in supply.
In this article, we review versatile DNA-based molecular markers that can be employed to analyze DNA for quality assurance, control, and authentication of TCM.

PCR based markers

PCR technique permits the amplification of any sequence of up to 40 kb of DNA, even in samples containing only minute quantities of DNA. Thus, the PCR technique is extremely useful in studying the genetic similarity or dissimilarity of TCM. Following amplification, the polymerase chain reaction (PCR) products are fractionated on agarose, polyacrylamide, or another type of gel matrix before it is detected by ethidium bromide (EtBr), autoradiography, or fluorescence by using a fluorescence labeled primer. To date, a variety of PCR-based methods have been established for the authentication of TCM.

Polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP)

RFLP was the first DNA profiling technique inexpensive enough to see widespread application. However, it is slow and needs a lot of labor. Also, it requires a large amount of sample DNA. Therefore, PCR-RFLP analysis, which consumes a minute amount of DNA, is much more suitable for TCM (Meyer et al., 1995). By using PCR-RFLP, a defined DNA fragment is first amplified by PCR, then it is digested by a certain restriction endonuclease to generate a restriction polymorphic profile unique to the species concerned (Shaw et al., 2002). Theoretically, the region for PCR amplification should be highly conserved among species so that it can be easily amplified. The analysis of sequence variation in the regions such as ribosomal DNA (rDNA) and a large subunit of ribulose-1,5-bisphosphate carboxylase (rbcL) has become an effective method for identification of medicinal herbs. The PCR-RFLP of rDNA or rbc1 has been successfully carried out on Glehnia (Mizukami et al., 1993), Epimedium (Nakai et al., 1996), Atractylodes (Mizukami et al., 1996; Cheng et al., 1997), Panax (Ngan et al., 1999), Codonopsis (Fu et al., 1999), Fritillaria (Zhang et al., 2005), Fritillaria (Wang et al., 2005; 2007), Sinopodophyllum and Dipsosoma (Gong et al., 2006), Ganoderma (Zhou et al., 2008), Akebia (Kitaoka et al., 2009) and Lonicera (Peng et al., 2010).

Random amplification of polymorphic DNA (RAPD)

RAPD is a PCR based analysis, but it only amplifies segments of DNA which are essentially unknown. This technique uses a single arbitrarily chosen primer of 10 nucleotides long as both the forward and reverse primers in a PCR reaction (Williams et al., 1990). Typically the scientist performing RAPD creates several arbitrary primers, and then proceeds with the PCR using a large template of genomic DNA to amplify fragments. By resolving the resulting patterns, a semi-unique profile can be obtained from a RAPD reaction. Other similar approaches include Arbitrarily-Primed PCR (AP-PCR) (Welsh and McClelland, 1990), and DNA amplified fragments (DAF) (Caetano-Anollés et al., 1991 a, b). The only difference among these approaches is that AP-PCR uses primers approximately 20 nucleotides long, while DAF uses a primer which is 5-8 nucleotides long. In the past 20 years, RAPD and related methods have been extensively applied to study the genetic similarity or dissimilarity of TCM incuding Glycyr rhiza (Yamazaki et al., 1994), Panax (Shaw and But, 1995), Cannabis (Gillan et al., 1995; Jagadish et al., 1996; Shirot a et al., 1998), Clematis (Zhang et al., 1996), Elephantopteris (Cao et al., 1996), Tatarisci (Cao et al., 1996), Sanges (Wang and Zhou, 1996; 1997), Coptis (Cheng et al., 1997), Indigofera (Zhang et al., 1997), Asari (Huang et al., 1998), Liriope (Wu et al., 1998), Anoectochilus (Cheng et al., 1998), Perilla (Ito et al., 1998), Codonopsis (Fu et al., 1999), Trichosanthes (Wang et al., 1999), Scutellaria (Hosokawa et al., 2000), Dipsosoma (Fu et al., 2000), Atractylodes (Chen et al., 2001; Guo et al., 2001), Magnolia (Guo et al., 2001), Paonia (Zhou et al., 2002), Rehmannia (Cheng et al., 2002), Cortex Magnoliaceae Officinalis (Liu et al., 2004), Morinda (Ding et al., 2006), Pogostemon (Pan et al., 2006), Selaginella (Li et al., 2007), and Lycium (Zhang et al., 2001; Sze et al., 2008).

Amplified fragment length polymorphism (AFLP)

The AFLP technique is based on the selective PCR amplification of restriction fragments from a total digest of genomic DNA (Vos et al., 1995). Developed in the early 1990s, AFLP consists of three steps: (1) Digestion of total genomic DNA by restriction enzymes, followed by ligation of adaptors to the sticky ends of the restriction fragments. (2) Selective amplification of a subset of the restriction fragments by using primers complementary to the adaptor sequence and restriction site sequence. (3) Visualization of the amplified fragments on denaturing polyacrylamide gels, either through either autoradiography or fluorescence methodologies (Zabeau and Vos, 1993). The AFLP technology is highly sensitive and reproducible, capable of detecting various polymorphisms in different genomic regions simultaneously. Thus, AFLP has become successfully used for the identification of genetic variation in strains or closely related species of TCM, such as Panax quinquefolius and Panax ginseng (Shaw et al., 1998), Fritillaria (Cai et al., 1999), Panax (Hong et al., 2005), Radix (Chen et al., 2005), and Magnolia (He et al., 2009) etc.
Simple sequence repeats (SSRs) and inter-simple sequence repeat (ISSR)

SSRs, also referred to as microsatellites, or sometimes short tandem repeats (STRs), are short sequences of nucleotides (2 to 6 units in length) that are repeated in tandem (Grist et al., 1993). SSRs can be amplified for identification by PCR, using the unique sequences of flanking SSRs as primers. Developed by Zietkeiwitcz et al. (1994), ISSR is a genomic region between SSRs. The complementary sequences of two neighboring SSRs are used as PCR primers, thus the variable region between them is amplified (Zietkeiwitcz et al., 1994). Both SSBs and ISSR have proven to be versatile molecular markers for assessing genetic relatedness in TCM populations of the same species such as Primula (Nan et al., 2003), Vitex rotundifolia (Hu et al., 2008), Cistanche (Shi et al., 2009), and Vitex trifolia var. simplicifolia (Liu et al., 2010b).

Direct amplification of length polymorphisms (DALP)

DALP technique uses an arbitrarily primed PCR to produce genomic fingerprints and to enable direct sequencing of DNA polymorphisms in virtually any species (Desmarais et al., 1998). Compared to arbitrarily primed fingerprinting, it detects a larger number of polymorphic loci and simplifies the procedures for recovery of polymorphic DNA bands. Dr. Shaw's group has successfully adopted DALP technique to authenticate P. ginseng and P. quinquefolius from different farms (Ha et al., 2002).

Other PCR-based methods

Many other alternative PCR-based techniques have been developed. Single-strand conformation polymorphism (SSCP), or single-strand chain polymorphism detects the changed migration rate of DNA molecules of identical length due to sequence-dependent, differential intramolecular folding of ssDNA by non-denaturing gel electrophoresis (Orita et al., 1989). PCR-SSCP technique has been used to differentiate Cannabis sativa and Humulus lupulus (Kohiyoura et al., 2000), and to identify Mycobacterium (Jiang et al., 2009). Denaturing gradient gel electrophoresis (DGGE) is a molecularfingerprinting method that separates polymerase chain reaction (PCR)-generated DNA products by applying them to an electrophoresis gel that contains a denaturing agent (Fischer and Lerman, 1983). DGGE-RAPD analysis, combining DGGE and RAPD, was proved to be highly reproducible and gives a higher level of polymorphism and consequently more markers compared to RAPD only (Dweikat et al., 1993; Bahieldin et al., 2006). This will be a useful tool for TCM authentication.

HYBRIDIZATION-BASED MARKERS

By using hybridization-based markers, DNA profiles are visualized by hybridizing the restriction enzyme-digested DNA, to a labeled probe, which is a DNA fragment of known origin or sequence. Restriction fragment length polymorphism (RFLP) is a typical molecular marker based on the differential hybridization of probe DNA to DNA fragments cleaved with restriction enzymes. RFLP is most widely used in genome mapping, marker-aided breeding, systematic, and evolution studies (Kochert, 1994; Shaw et al., 2002). SSRs or microsatellites, which are another representative of hybridization-based marker technology are generally less than 6 bp in length and are repeated from a few to thousands of times. The regions flanking the microsatellites can be amplified, providing co-dominant sequence-tagged sites (STS) or a repetitive sequence to act as a probe. The methodology has been derived from RFLP, and specific fragments are visualized by hybridization with a labeled microsatellite probe (Litt and Luty, 1989). SSRs have been used to generate DNA fingerprints in TCM populations of the same species such as P. ginseng and P. quinquefolius (Leung and Ho, 1998), Primula (Nan et al., 2003), Vitex rotundifolia (Hu et al., 2008), Cistanche (Shi et al., 2009), and V. trifolia var. simplicifolia (Liu et al., 2010b).

SEQUENCING-BASED MARKERS

DNA sequencing is a definitive means for the identification of TCM. The reducing cost of DNA sequencing has led to the availability of large sequence data sets derived from the discovery of whole genome sequencing and large scale expressed sequence tag (EST), which is a valuable source of new genetic markers. A representative region for sequencing is the ITS from ribosomal RNA genes (known as ribosomal DNA or rDNA). Eukaryotic rDNAs are found as parts of repeat units that are arranged in tandem arrays, located at the chromosomal sites known as nucleolar organizing regions (NORs). Each repeat unit consists of a transcribed region (having genes for 18S, 5.8S and 26S rRNAs and ITS). The length and sequences of ITS region of rDNA repeats are believed to be fast evolving and therefore may vary, making this region a useful sequence for phylogenetic studies (Baldwin et al., 1995; Mitchell and Wagstaff, 1997). Generally, sequence homologies within species are high while those between species or families are low, indicating the ITS region can be used as a marker from interspecific level to family level (Shaw et al., 2002). Researchers have sequenced ITS region for the authentication of TCM, such as Codonopsis (Fu et al., 1999), Panax (Ngan et al., 1999), Dendrobium (Lau et al., 2001; Ding et al., 2002; Xu et al., 2006), Alpinia (Zhao et al., 2001), Hericium (Lu et al., 2002), Gentiana (Ji et al., 2003; Liu et al., 2010a), Artemisia (Kim et al., 2004),
Ephedra (Long et al., 2004; Kakiuchi et al., 2006), Fritillaria (Wang et al., 2005), Euphorbia (Xue et al., 2007a), Croton (Xue et al., 2007b), Pogostemon (Zhang et al., 2007), Nelumbo (Lin et al., 2007), Psammosilenae (Liu et al., 2008), Fallopia (Zheng et al., 2008), Cimicifuga (Xue et al., 2009), and Lonicera (Peng et al., 2010).

Researchers also have sequenced regions of DNA (18S, 5.8S, 26S and 5s rDNAs etc.) for verifying TCM, including Cordyceps (Ito and Hirano, 1997), Fritillaria (Cai et al., 1999; Cai et al., 2001; Li et al., 2003), Astragalus (Ma et al., 2000), Crocus (Ma et al., 2001), Notoginseng (Cao et al., 2001), Alpinia (Zhao et al., 2002), Adenophoreae (Zhao et al., 2003a), Angelica (Zhao et al., 2003b), Panax (Zhang et al., 2006), and Curcuma (Cao et al., 2010). Now, with a further decrease in the cost of sequencing, we expect that DNA sequencing will become a more common means for the authentication of TCM.

**DNA MICROARRAY**

DNA microarray was developed in response to the need for a high-throughput, efficient, and comprehensive strategy that can simultaneously measure the expression of all the genes, or a large defined subset, encoded by a genome. It consists of an arrayed series of thousands of microscopic spots of DNA oligonucleotides, known as probes, which are used to hybridize DNA targets labeled by fluorophore-, silver-, or chemiluminescence (Schena et al., 1995, 1996). DNA microarray based technology can be useful as an efficient, accurate tool for the identification and authentication of TCM (Chavan et al., 2006). Although DNA microarray technology has huge potential for authentication and quality control of TCM, it is still in the exploratory stage and needs validation by other biological experiments. Standardized, sensitive, reproducible microarray platforms, databases, and visualization methods for expression profiles are needed.

**CONCLUSION**

DNA technology provides an independent and reliable approach for the differentiation of Chinese medicinal individual, species and population, homogeneity analysis, and detection of adulterants. At present, the authentication of TCM is moving towards a more comprehensive and overall direction. More and more system biology methods, including metabolomics, genomics and proteomics, are introduced to this area. One of the immediate tasks for researchers in this field is to build a reference library of TCM with genetic information (such as DNA sequences and fingerprints), in order to promote the further development of more scientific, effective, and practical DNA methods for the authentication of TCM.

**ABBREVIATIONS**

AFLP, Amplified fragment length polymorphism; AP-PCR, arbitrarily-primed PCR; DAF, DNA amplified fragments; DALP, direct amplification of length polymorphisms; DGGE, denaturing gradient gel electrophoresis; ISSR, inter-simple sequence repeat; ITS, internal transcribed spacers; PCR, polymerase chain reaction; PCR-RFLP, polymerase chain reaction-restriction fragment length polymorphism; RAPD, random amplification of polymorphic DNA; RFLP, restriction fragment length polymorphism; SSCP, single-strand conformation polymorphism; SSRs, simple sequence repeats; TCM, traditional Chinese medicine.

**REFERENCES**


Anoeotrichium formosanum and Anoeotrichium kushonensis species with RAPD markers. Planta Med. 64:46-49.


Liu et al.         125


Ma XQ, Duan JA, Zhu DY, Dong TT, Tsim KW (2000). Species


Zhao KJ, Dong TT, Tu PF, Song ZH, Lo CK, Tsim KW (2003b).