

An antibody is a very powerful authentication tool. It can distinguish isozymes as well as other proteins. A single difference in the amino acids may change the epitope of the protein which is recognised by the antibody. Accordingly, antibodies can differentiate polymorphic proteins and species-specific proteins from different organisms. For examples: antibodies were developed against *Prorocentrum lima* (Ehrenberg) Dodge and *P. rostratum* Stein (Lopez-Rodas and Costas, 1999); monoclonal antibodies were raised against a conjugate of ginsenoside Rf and bovine serum albumin; the anti-Rf monoclonal antibody was used for quantification of Rf in crude ginseng fractions and in body fluids (Nah *et al.*, 2000); an antibody specific to lumbricus (the crude drug dried earthworm) used in Chinese medicine was shown not to cross-react with other crude drugs (Bai *et al.*, 1997) and a selected antibody enzyme immunoassay was subsequently developed to measure the contents of the lumbricus component in herbal products.

Immunological analyses are highly reproducible and the antibodies may be developed into a diagnostic kit for rapid testing. However, tremendous effort is required for screening a suitable antigen and for generating the antibody.

1.2. Molecular Marker Technologies

Benefiting from molecular cloning and PCR techniques, DNA markers have now become a popular means for identification and authentication of plant and animal species. DNA-based markers are less affected by age, physiological condition of samples and environmental factors. They are not tissue-specific and thus can be detected at any phase of organism development. Only a small amount of sample is sufficient for analysis and the physical form of the sample does not restrict detection. These non-stringent requirements are particularly relevant for Chinese medicinal materials that are expensive or in limited supply. The power of discrimination of DNA-based markers is so high that very closely related varieties can be differentiated. In the following, we provide a general account of representative molecular marker technologies, which will be elaborated in the following chapters. Although RNA-based markers have also been considered, for simplicity, herein we shall focus primarily on DNA-based

Table 2. Comparison of the commonly used fingerprinting methods

Molecular Markers	Degree of Polymorphism per run		Reproducibility
	<u>Interspecific</u>	<u>Intraspecific</u>	
1. RAPD	++	+	+
2. SSR	++++	+++	++
3. AFLP	++++	++++	++
4. DALP	+++	++	++
5. PCR-RFLP	++	+	+++

'++++' means the highest, '+' means the lowest

markers. A comparison of the common DNA fingerprinting methods is presented in Table 2.

A diverse array of DNA-based marker technologies has been established to explore various DNA polymorphism and they can be classified into three broad types, namely polymerase chain reaction (PCR)-based markers, hybridisation-based markers and sequencing-based markers.

1.2.1. PCR-Based Markers

PCR enzymatically multiplies a defined region of the template DNA (Fig. 1). The specific multiplication is attributed to the presence of primers, which are single-stranded polynucleotides that recognise and bind to the complementary DNA sequence on template DNA. The amplification process starts with denaturation of the double-stranded template DNA to single-stranded DNA under a high temperature, usually between 90–95°C, followed by the specific annealing of the primer(s) to the single-stranded template DNA at a lower temperature. The annealing primers are then extended by a thermostable DNA polymerase. Repeating the denaturation-annealing-extension cycle leads to an exponential accumulation of the DNA fragment of the defined sequence. The amplified products are then fractionated on agarose, polyacrylamide or other gel matrix and detected by ethidium bromide (EtBr) or silver staining, autoradiography (using an isotope-labelled

primer), or fluorescence (using a fluorescence-labelled primer). The distance between the priming sites is usually from 100 bp to a few kilobases (kb) (and hence the size of the amplified fragments), although the recently developed ‘long distance PCR’ allows amplification of up to 40 kb or beyond. PCR amplification of any region of a DNA sample is possible, providing their flanking sequences are known, as these are needed for designing primers. Owing to PCR’s sensitivity and ability to amplify DNA from a small amount of materials *in vitro*, a variety of PCR-derived methods have been established.

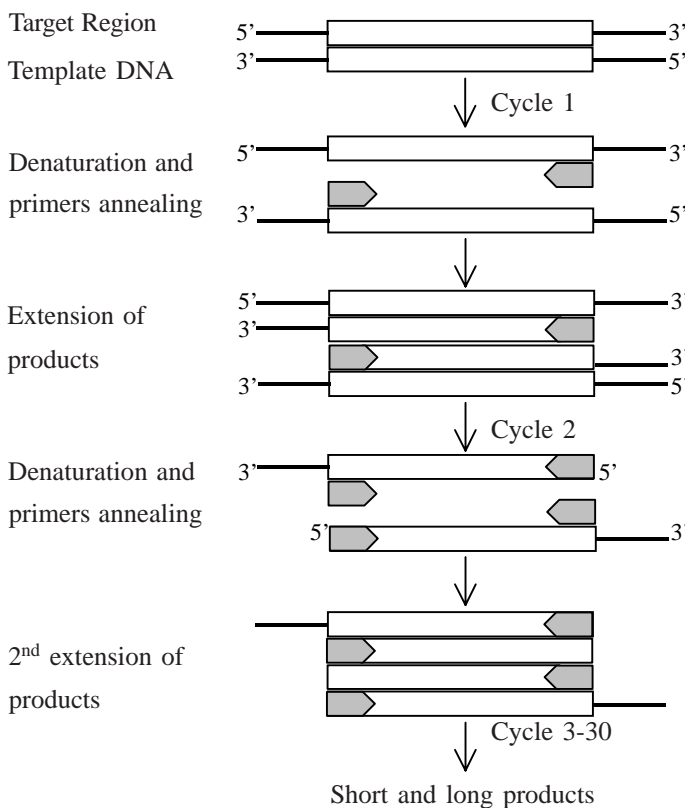


Figure 1. Schematic diagram of PCR. The primer used may be specific target sequence (e.g. in PCR-RFLP) or arbitrarily chosen (e.g. in RAPD and AP-PCR).

1.2.1.1. PCR by arbitrarily chosen primers (ACP-PCR)

In 1990, two teams simultaneously reported the development of PCR-based, novel, genetic screening techniques: random amplified polymorphic DNA (RAPD) and arbitrarily-primed PCR (AP-PCR) (Williams *et al.*, 1990; Welsh and McClelland, 1990). To amplify the template DNA, RAPD uses a single primer of 10 nucleotides long and AP-PCR uses primers approximately 20 nucleotides long. Another similar approach, called DNA amplified fragments (DAF) which uses a primer 5–8 nucleotides long, was described by Caetano-Anolles *et al.* (1991a; 1991b). These techniques use a single arbitrarily chosen oligonucleotide as both the forward and reverse primer in a PCR reaction. A given primer is able to hybridise to a number of sites along a genomic DNA, but not all of these hybridisations lead to the generation of detectable PCR products. To get a successful amplification, the distance between two adjacent hybridisation sites should be within 2 kb, i.e. the size limit of a PCR fragment under normal conditions. These fragments are generated from different locations of the genome and hence multiple loci may be examined simultaneously. Several companies have now marketed various RAPD kits, making this a convenient fingerprinting method.

ACP-PCR has been extensively applied to study the genetic similarity or dissimilarity of Chinese medicinal materials including *Glycyrrhiza* (Yamazaki *et al.*, 1994), *Cannabis* (Gillan *et al.*, 1995; Jagadish *et al.*, 1996; Shirota *et al.*, 1998), *Clematis* (Zhang *et al.*, 1996), *Coptis* (Cheng *et al.*, 1997a), *Indigofera* (Zhang *et al.*, 1997), *Liriope* (Wu *et al.*, 1998c), *Anoectochilus* (Cheng *et al.*, 1998), *Perilla* (Ito *et al.*, 1998), *Scutellaria* (Hosokawa *et al.*, 2000) and snakes (Wang and Zhou, 1996; 1997; Wang *et al.*, 1996). PCR by arbitrarily chosen primers has also been used to differentiate between cultivated and wild garlic cultivars (Al-Zharim *et al.*, 1997) and differentiate *P. quinquefolius* in different populations (Bai *et al.*, 1997). Our group is one of the first to apply RAPD and AP-PCR to authenticate dried Chinese medicinal materials and has successfully distinguished herb samples derived from *Panax*, *Taraxacum*, *Elephantopus*, *Acorus*, *Cremastra*, *Dysosma*, *Epimedium*, *Codonopsis* and various commercial ginseng products (Cheung *et al.*, 1994; Shaw and But, 1995; Cao *et al.*, 1996a; 1996b; Ngan, 1997; Ngan *et al.*, 1997; Fu *et al.*, 1999;

2000). Recently, we have extended the techniques to identify *Codonopsis pilosula* from different localities (Zhang *et al.*, 1999).

1.2.1.2. Simple Sequence Repeat (SSR)

SSR markers or microsatellites are tandem repeats interspersed throughout the genome and can be amplified using primers that flank these regions (Grist *et al.*, 1993). These markers are also termed simple sequence length polymorphism (SSLP) or sequence-tagged microsatellite site (STMS). They can be used as probes in a hybridisation-based approach, e.g. low C_0t DNA. In a mixture of denatured DNA, SSR re-associates quickly owing to the low complexity of the nucleotide composition. SSLP can be used to identify different strains and races of organisms and to assess genetic relatedness in populations. For example, species-specific fingerprints were identified in bacteria, fungi, mussels, ginseng and *Amaranthus* (Leung, 1999). Primers for SSR analysis can also be constructed by searching the GenBank for SSR loci of related species or by screening genomic libraries. Searching databases for existing information is a quick way to develop a probe for DNA fingerprinting and no prior knowledge of the sequence is needed. In addition, primers can be synthesised based on a repeat sequence of $(CA)_n$, for example, primers $(CA)_8RG$ or $(AGC)_6TY$ have a degenerate 3'-anchor (Godwin *et al.*, 1997). SSR has been successfully used to construct detailed genetic maps of several organisms and to study genetic variation within populations of the same species, such as grapes, honeybees and tropical trees (Brown *et al.*, 1996).

1.2.1.3. Amplified Fragment Length Polymorphism (AFLP)

AFLP is essentially a combination of RFLP and PCR techniques (Vos *et al.*, 1995). Genomic DNA is first digested by appropriate restriction enzymes. A subset of resultant fragments representing many loci is then ligated to synthetic adaptors and amplified with specified primers which are complementary to a selective sequence on the adaptors. Subsequent separation of the resultant fragments is performed on a highly resolving

sequencing gel and visualised using autoradiography. Where radiolabelled nucleotides are not used in the PCR step, fluorescence or silver staining techniques can be used to visualise the amplification products (Chalhoub *et al.*, 1997). AFLP analysis is similar to RAPD assay in that no prior knowledge of the sequence is required; however, AFLP detects a greater number of loci than RAPD does (Russell *et al.*, 1997). The complexity of the AFLP profiles is dictated by the primers and restriction enzymes chosen, as well as the composition of the genomic DNA. In the past five years, AFLP has been used for various organisms including bacteria (Lin and Kuo, 1996; Janssen *et al.*, 1997), fungi (Muller *et al.*, 1996; Leissner *et al.*, 1997), plants (Lu *et al.*, 1996; He and Prakash, 1997), insects (Reineke *et al.*, 1998) and humans (Latorra and Schanfield, 1996) in a broad range of studies, such as the characterisation of species (Russell *et al.*, 1997; Picardeau *et al.*, 1997), molecular evolution and biological diversity (Keim *et al.*, 1997), comparison of the differential expression of genes (Habu *et al.*, 1997), gene mapping (Otsen *et al.*, 1996), chromosome landing (Cnops *et al.*, 1996), management of endangered plants (Travis *et al.*, 1996) and chimerism analysis of allogeneic transplantation (Schreiner *et al.*, 1996).

Our laboratory has adopted AFLP in the study of Chinese herbal materials. We have successfully discriminated between *P. quinquefolius* and *P. ginseng* as well as between various cultivars of *P. quinquefolius* from different farms (Shaw *et al.*, 1998). Details of this application are described in Chapter 7.

1.2.1.4. Direct Amplification of Length Polymorphism (DALP)

This method resembles arbitrarily-primed fingerprinting but detects a larger number of polymorphic loci and greatly simplifies the procedures for recovery of polymorphic DNA bands (Desmarais *et al.*, 1998). It uses a selective forward primer containing a 5' core sequence (e.g. the M13-40 universal sequencing primer) plus additional bases at the 3' end, and a common reverse primer. The PCR products generated can be sequenced directly with the same forward and reverse primers. DALP has been used to detect the polymorphism from different classes of organisms (virus, vertebrate

and invertebrate, etc.); different species of mussels (*Mytilus galloprovincialis* and *M. edulis*) and gulls (*Larus cachinnans* and *L. audouinii*) and intra- and inter-subspecies in mice (*Mus spretus*, *M. musculus domesticus* and *M. m. musculus*) (Desmarais *et al.*, 1998). We have adopted this method to differentiate *P. ginseng* and *P. quinquefolius* from different farms and details of this application are described in Chapter 6.

1.2.1.5. Polymerase Chain Reaction-Restriction Fragment Length Polymorphism (PCR-RFLP)

Traditional RFLP analysis is not suitable for Chinese medicinal materials as it demands a large amount of intact DNA. Therefore, PCR-RFLP, a method that consumes a minute amount of DNA, is more appropriate. In this method, a defined DNA fragment is first amplified by PCR and then digested with a selected restriction endonuclease to generate a restriction polymorphic profile unique to the species concerned. It is preferable that the region for PCR-RFLP analysis is flanked by a sequence that is conserved across species so that this region can be readily amplified from many species using 'universal' primers. Two suitable candidates for such regions are ribosomal DNA (rDNA) and a large subunit of the ribulose-1,5-bisphosphate carboxylase L (rbcL) gene. The PCR-RFLP of rDNA has been carried out on *Glehnia* and *Atractylodes* (Mizukami *et al.*, 1993; 1996; Cheng *et al.*, 1997b). Various *Epimedium* species have also been differentiated by the PCR-RFLP of the rbcL region (Nakai *et al.*, 1996). We have also successfully used this approach to differentiate various *Panax* species (Ngan *et al.*, 1999) and *Codonopsis* from their adulterants (Fu *et al.*, 1999) and obtained a patent (Wang *et al.*, 1999a) for the application of this technique.

1.2.1.6. Other PCR-based techniques

Many other PCR-based variants have been developed. Single strand conformation polymorphism (SSCP) analysis detects the changed migration rate of DNA molecules due to sequence-dependent, differential intramolecular folding of ssDNA under non-denaturing gel electrophoresis

conditions (Orita *et al.*, 1989). PCR-SSCP has been used to differentiate *Cannabis sativa* and *Humulus lupulus* (Kohjyouma *et al.*, 2000). Denaturing gradient gel electrophoresis (DGGE) is the most widely applied method for mutation analysis of genes. It is based on the concept that the electrophoretic mobility of a DNA molecule depends upon its denatured state, which is in turn dictated by its primary structure. Two molecules that differ by as little as a single nucleotide will exhibit a slight but detectable variation in migration rate when they are separated through a gel of increasing denaturant concentration. RAPD-DGGE analysis, combining RAPD and DGGE (Dweikat *et al.*, 1993), provides an alternative tool to pedigree assessment. However, the use of these other PCR-based techniques for TCM authentication has not been explored.

1.2.2. Hybridisation-Based Markers

Hybridisation-based marker technologies use cDNA, cloned DNA elements, or synthetic oligonucleotides as probes, which are labelled with radioisotopes or with conjugated enzymes that catalyse a coloured reaction, to hybridise DNA. The DNA is either cleaved with restriction enzymes or amplified by PCR, separated by gel electrophoresis, and transferred to a solid support matrix. RFLP is representative of this type of technology and is most widely applied in genome mapping, marker-aided breeding, systematics and evolution studies (Kochert, 1994; Heslop-Harrison and Schwarzacher, 1996; Paterson, 1996). Restriction polymorphism occurs when mutations remove an existing restriction site or create a new restriction site. These alterations are detected by using a hybridisation probe. The choice of the DNA probe/restriction enzyme is crucial in RFLP analysis.

DNA probes for detecting repeated DNA sequences can be divided into two types: (1) tandem repeats which are present as clusters along chromosomes and (2) dispersed repeats which are distributed over all the chromosomes. Shorter repeating sequences, that are generally less than 6 bp in length and are repeated from a few to many thousands of times, are abundant in eukaryotes. They are designated 'microsatellites' (Litt and Luty, 1989) or SSR. The regions flanking the microsatellite loci can be amplified

by PCR, thus providing co-dominant sequence-tagged sites (STS) or a repetitive sequence to act as a probe to generate DNA fingerprints such as in the differentiation of *P. ginseng* and *P. quinquefolius* (Leung and Ho, 1998). The probe, low C_0t DNA, is produced by shearing the isolated DNA from a reference DNA and the repetitive sequence is renatured and labelled. It is then hybridised to the membrane that has been fixed with restriction enzyme-digested DNA from the examined subject, which has been separated by electrophoresis. As a result, a specific DNA fingerprint is displayed on the autoradiogram. The abundance and ubiquitous distribution of microsatellites make them very valuable in linkage mapping, in identification of quantitative traits loci and in forensic cases. Additionally, a remarkable feature of microsatellites, not shared by minisatellites, is that primers developed in one species can be used in related taxa (Coote and Bruford, 1996).

1.2.3. Sequencing-Based Markers

DNA sequencing is a definitive means for identifying TCM. Further, variations due to transversion, transition, insertion or deletion can be assessed directly and information on a defined locus can be obtained. A representative sample is the internal transcribed spacers (ITS) from ribosomal DNA (rDNA). The ITS region of 18s–26s rDNA has proved to be a useful sequence for phylogenetic studies in many angiosperm families (Baldwin *et al.*, 1995; Mitchell and Wagstaff, 1997). The level of ITS sequence variation suitable for phylogenetic analysis is found at various taxonomic levels within families, depending on the lineage. Our laboratory has sequenced the ITS regions of *Codonopsis*, *Panax* and *Dendrobium* species, and their adulterants (Fu *et al.*, 1999; Ngan *et al.*, 1999; Lau *et al.*, 2001). In general, sequence homologies within species were found to be high and those between species or families low, indicating that the ITS region can be used as a marker from family to interspecific level.

A number of researchers have also sequenced other regions of DNA and now use them as diagnostic tools for authentication purposes. The spacer region of 5s rDNA was sequenced to verify that the crude drug ‘angelica root’ was derived from *Angelica acutiloba* rather than from *Bupleurum*

falcatum, *Cnidium officinale* or *Glehnia littoralis*. A sequence of 300 bp amplified from the conserved sequences of the 5s rRNA gene was used to confirm the origin of *Angelica acutiloba*. This sequence is identical within the species and among the wild varieties but varies substantially with other medicinal species belonging to Umbelliferae (Mizukami, 1995). The spacer sequence of 5s rDNA has also been used to authenticate *Fritillaria* (Cai *et al.*, 1999) and *Astragalus* (Ma *et al.*, 2000); 18s rDNA for verifying several ginseng drugs (Fushimi *et al.*, 1996); *rbcL* for discriminating between 'banxia' and 'tiananxing' as well as 'banxia' related natural medicines (Kondo *et al.*, 1998); the intergenic spacer between the *trnL* 3' exon and *trnF* in chloroplast DNA for intraspecific variation in *Cannabis sativa*; and the *trnK* sequence for identification of *Atractylodes*-derived crude drugs (Mizukami *et al.*, 2000). For animal-derived drugs, the 12s ribosomal RNA and cytochrome *b* genes have been used to identify two species of *Hippocampus* (Wu *et al.*, 1998b), 21 species of turtle shells (Wu *et al.*, 1998a) and to authenticate crude drugs from snakes (Wang and Zhou, 1996; 1997). Diagnostic primers have been derived from the cytochrome *b* sequence for the authentication of the crude drug *Zaocys dhumnades* (Wang *et al.*, 2000). Several chapters in this book are also devoted to using sequencing as a means for authenticating herbal and animal medicinal materials. Currently, a laboratory can easily generate more than a few thousand base pairs of DNA per day using an automatic DNA sequencer. Coupled with a further decline in the price of consumable materials and equipment used in sequencing, we expect that DNA sequencing will become a more common means for the authentication of Chinese medicinal materials.

1.3. Conclusion

Molecular technology provides an independent approach for the authentication of medicinal materials. Its impact on quality control is just emerging. It will be more fruitful if a concerted effort is made to integrate the existing molecular fingerprinting data and to co-ordinate future projects of molecular authentication of Chinese medicinal materials. One of the immediate tasks for researchers in this field is to compile a molecular