Review

Application of molecular identification techniques in traditional Chinese medicine

Jiankun Jin[a†], Weiwei Zhang[b†], Rui Liu[a], Ruiqi Ma[a], Yuzheng Xiang[a], Yibo Wang[a], Ying Li[a*], Yu Chen[a*]

[a] *School of Life Science and Biopharmaceutics, Shenyang Pharmaceutical University, Shenyang 110016, China;*
[b] *GenScript USA Inc., Piscataway, New Jersey 08854, USA*

Abstract

In recent years, molecular identification technology has become the predominant approach for the identification of traditional Chinese medicine. The molecular identification techniques in recent years were analyzed and summarized in this paper, such as RAPD, ISSR, RFLP, AFLP, SNP, and DNA bar code sequence analysis techniques. By consulting Sciencedirect databases and Web of Science databases, 2348 related articles were found, of which 39 were related to molecular identification techniques and traditional Chinese medicine. The application of the molecular identification techniques in four aspects was reviewed, namely the identification on the authenticity (true or false), multi-source identification and genetic diversity, producing area, and growing year discrimination of traditional Chinese medicine.

Keywords: traditional Chinese medicine; molecular identification techniques; DNA bar code

1 Introduction

The identification methods of traditional Chinese medicine can be categorized into four main developmental phases: sensory evaluation, microscopic identification, physical and chemical identification, and sub-identification [1]. In recent years, molecular identification technology has increasingly become the predominant approach for the identification of traditional Chinese medicine. These techniques focus on the direct analysis of genetic material polymorphisms, specifically DNA, to deduce the inherent genetic variations among species. Molecular identification methods include random primer PCR (RP-PCR), random amplified polymorphic DNA (RAPD), amplified simple sequence repeat (ISSR), restriction fragment length polymorphism (RFLP), amplified restriction fragment length polymorphism (AFLP), and DNA barcoding sequence analysis. In this paper, the main molecular identification techniques were comprehensively analyzed and summarized, and their respective advantages, disadvantages, and primary applications

† These authors contributed equally to this work and should be considered co-first authors.
* Author to whom correspondence should be addressed. Address: School of Life Science and Biopharmaceutics, Shenyang Pharmaceutical University, Shenyang 110016, China; Tel.: +86-18341400530 (Yu Chen), +86-18341400610 (Ying Li); E-mail: gzweishengwu@126.com (Yu Chen), yyswalk@163.com (Ying Li).
These authors have no conflict of interest to declare.

Received: 2024-08-15 Accepted: 2025-05-29

were introduced in detail. The application of molecular identification techniques in the context of traditional Chinese medicine was examined from four perspectives: authenticity verification, multi-origin identification and genetic diversity assessment, age determination, and origin identification. Additionally, the paper addressed existing challenges in molecular identification technology and outlines its potential future direction [2].

2 Types of molecular identification techniques

As shown in Fig. 1, the development of molecular identification technology can be divided into molecular markers based on traditional southern hybridization, molecular markers based on PCR, molecular markers based on repetitive sequences, molecular markers based on mDNA and DNA bar code technology.

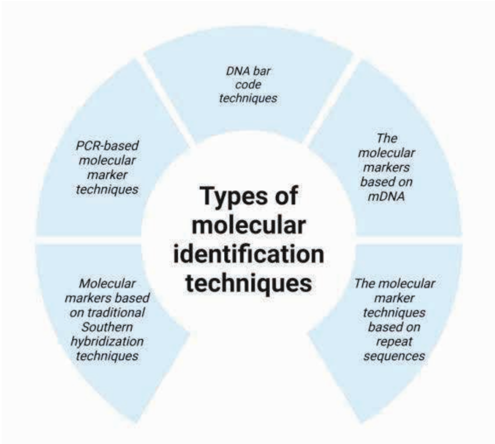

Fig. 1 Types of molecular identification techniques

2.1 Molecular markers based on traditional southern hybridization techniques

The primary methodologies encompassed are restriction fragment length polymorphism (RFLP), single strand conformation polymorphism combined with RFLP (SSCP-RFLP), and denaturing gradient gel electrophoresis in conjunction with RFLP (DGGE-RFLP) [1].

2.2 PCR-based molecular marker techniques

These methods mainly include random amplified polymorphic DNA (RAPD), sequence-tagged site (STS) marker sequencing, sequence characterized amplified region (SCAR) sequencing, reverse transcription polymerase chain reaction (RP-PCR), oligonucleotide primer polymerase chain reaction (OP-PCR), single-strand conformation polymorphism (SSCP-PCR), small oligonucleotide DNA analysis (SODA), DNA amplification fingerprinting (DAF), loop-mediated isothermal amplification (LAMP), and amplified fragment length polymorphism (AFLP). Among them, AFLP is particularly popular. It utilizes polymerase chain reaction (PCR) techniques to amplify genomic DNA restriction fragments. Initially, genomic DNA is cleaved using restriction endonucleases, followed by the ligation of double-stranded adapters to the ends of the resulting DNA fragments. The sequences of these adapters, along with adjacent restriction site sequences, serve as primer binding sites. The restriction fragments are generated using two distinct enzymes, one that cuts infrequently and the other that is commonly employed. This technique integrates the principles of restriction fragment length polymorphism (RFLP) and PCR, thereby combining the reliability associated with RFLP and the efficiency characteristic of PCR. Notably, AFLP amplification can yield specific DNA bands in one variety, but not in another, allowing the DNA polymorphisms generated through primer induction and amplification to function as sub-markers [3].

2.3 The molecular marker techniques based on repetitive sequences

DNA sequences mainly include satellite DNA with repetitive units ranging from hundreds to thousands of base pairs, microsatellite DNA with repetitive units of more than five base pairs, and microsatellite DNA characterized by repetitive units of two to five base pairs, in addition to bead repetitive sequences. Among these, the inter simple sequence repeat (ISSR) technique, which is a form of microsatellite DNA, is widely utilized. The ISSR method was developed by Zietkiewicz et al. in 1994 and is based on the use of anchored microsatellite DNA as primers [4]. Specifically, this involves the addition of two or four random nucleotides to either the 3′ or 5′ end of the simple sequence repeat (SSR). During the polymerase chain reaction (PCR), these anchored primers facilitate annealing at specific sites, resulting in the amplification of DNA fragments located between repetitive sequences that are close enough to the anchored primers. The multiple bands produced from the amplified inter-SSR regions were subsequently analyzed using polyacrylamide gel electrophoresis, with most of the amplified bands exhibited dominant characteristics.

2.4 The molecular markers based on mDNA

These methods mainly include differential display, reverse transcription polymerase chain reaction (RT-PCR), differential display reverse transcription polymerase chain reaction (DDRT-PCR), and characteristic difference analysis (RDA). Among them, RT-PCR is especially popular. RT-PCR represents a common variant of the polymerase chain reaction technology. In this process, an RNA strand is reverse transcribed into complementary DNA, which subsequently serves as a template for DNA amplification through PCR [2].

2.5 DNA bar code techniques

DNA barcoding is a method employed for the rapid and precise identification of biological species by using one or more standardized short genomic DNA fragments179. This technology has several advantages, including high efficiency, accuracy, ease of automation, and standardization. It reduces the dependence on expert knowledge and experience, thereby facilitating the simplified and precise analysis and identification of traditional Chinese medicine. Currently, DNA barcoding is mainly used for the molecular identification of plant-based traditional Chinese medicine, in which ITS2 sequence is commonly used, with the psbA-trnH sequence as the supplementary marker. In contrast, the identification of animal-based traditional Chinese medicine primarily relies on the COI sequence, with ITS2 as the auxiliary sequence.

3 Application of molecular identification technology

Advances in technology have led to the emergence of novel molecular identification techniques. Presently, the most widely utilized methods include inter-simple sequence repeat (ISSR), simple sequence repeat (SSR), and DNA barcoding technology. Compared with traditional methodologies, ISSR and SSR techniques provide several advantages, such as a large number of markers, high allelic variation, co-dominant markers, excellent repeatability, dominant expression, broad applicability, and significant stability. Besides, these techniques do not require prior cloning and sequencing. However, there are also some limitations, including the need for prior knowledge of repetitive sequences, considerable development complexity, high cost, and slightly lower accuracy. Comparatively, DNA barcoding technology boasts advantages, such as insensitivity to morphological characteristics and developmental stages of biological specimens, wide range of sample detection, high efficiency and accuracy, and the potential of automation and standardization. However, not all medicinal materials exhibit identifiable DNA variations .

The application of molecular identification techniques in traditional Chinese medicine is mainly manifested in four aspects: authenticity identification, multi-source identification and genetic diversity, age identification and origin identification of traditional Chinese medicine, as illustrated in Fig. 2.

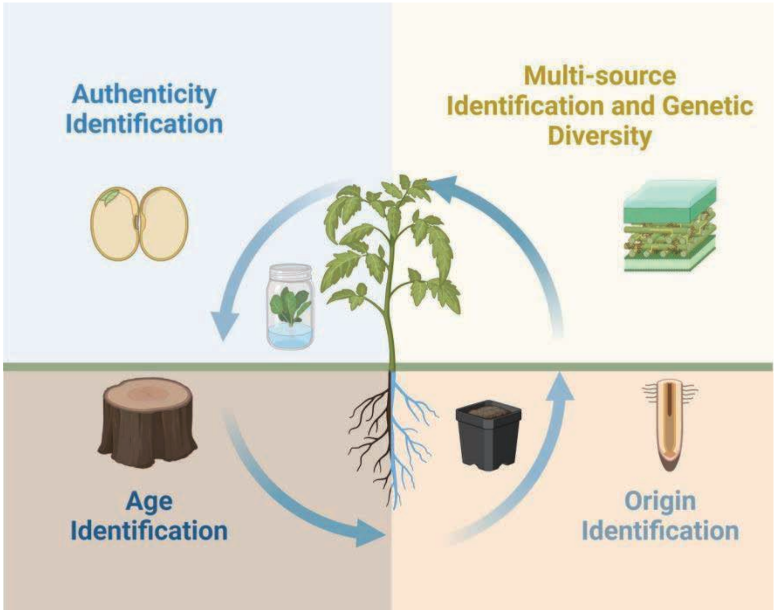

Fig. 2 The application of molecular identification technology in TCM

3.1 Application in the identification of authenticity

The methodology for the microscopic identification of the authenticity of traditional Chinese medicine is relatively straightforward. However, it requires high-level expertise of the appraisers, and solid scientific foundation. Although chemometric approaches are effective, they are usually time-consuming and labor-intensive. In recent years, more and more researchers have realized that molecular identification technology provides a more accurate and efficient means to verify the authenticity of traditional Chinese medicine.

Xing et al. screened 14 pairs of SSR primers and 22 pairs of ISSR primers, focusing on the selection of both types of primers. The amplification results showed that polymorphic loci were abundant, with distinct bands and stable results. The genetic diversity of tissue culture seedlings of morphologically similar Bletilla striata and the authentic purple-flowered Bletilla was assessed by molecular marker technology. Furthermore, a method for rapid identification of Bletilla striata tissue culture plantlets by SSR and ISSR marker techniques was successfully developed [7].

Luo et al. conducted PCR amplification and sequencing of the ITS2 regions from five distinct primitive plant species of Fritillaria chuanensis. They broadened their research scope by retrieving ITS2 sequences for ten species and thirteen samples of Fritillaria thunbergii and its common adulterants from GenBank. MEGA4.1 was used to calculate the Kmuri 2murp distances between species and within species, and the variations in the secondary structures of the ITS2 sequences among the samples were examined. Finally, a phylogenetic tree was reconstructed based on the ITS2 sequences. The findings indicated that the maximum Kmur2murP distance within the original species of Fritillaria thunbergii was 0.0276, while the minimum interspecific distance between Fritillaria thunbergii and its adulterants was 0.0583. The secondary structures of the ITS2 sequences of Fritillaria thunbergii and its adulterants exhibited significant differences. The reconstructed phylogenetic tree demonstrated that various primitive species of Fritillaria thunbergii clustered into a single branch, absolutely different from counterfeit specimens. This study further emphasizes the role of molecular identification techniques in distinguishing authentic traditional Chinese medicine from its adulterated counterparts [8].

3.2 Application in multi-motif identification and genetic diversity

Numerous studies have been conducted to determine the origins of traditional Chinese medicine by molecular identification technologies. These investigations include the analysis of genetic diversity in various medicine, such as Scutellaria baicalensis, Fritillaria thunbergii, loquat, and Radix Pseudostellariae. Additionally, polymerase chain reaction (PCR) direct sequencing technology has been used to facilitate the classification and identification of Chinese yam.

Liu et al. employed the PCR direct sequencing technology to determine the nucleotide sequence of the 18S rRNA gene in Chinese yam and its local varieties, and analyzed the sequence homology. The results showed that the 18S rRNA sequence length for Chinese yam, Guang yam, and native yam was 1810 base pairs, while the length for Fang yam was 1807 base pairs. Sequence comparison showed that the 18S rRNA sequence of Guang yam was identical to that of authentic yam, while the homology with native yam and prescription yam was 99.89% and 97.51%, respectively [9].

Ma et al. extracted and analyzed genomic DNA from 11 dried Bupleurum samples sourced from various provinces using Inter-Simple Sequence Repeat (ISSR) markers. The genetic distance between samples was calculated by DPS (V7.5) software, and a phylogenetic tree was constructed by unweighted pair group method with arithmetic mean (UPGMA). Internal transcribed spacer (ITS) primers were utilized for polymerase chain reaction (PCR) amplification and sequencing of the Bupleurum samples. The ITS sequences were subsequently analyzed by DNAMAN software to assess genetic similarity among the 11 Bupleurum germplasm. The genetic distances ranged from 0.4588 to 0.7822, indicating that the genetic basis among the samples was narrow. Clustering analysis showed that the samples could be categorized into two distinct groups, and the samples from similar geographic origins tended to cluster together. The ITS sequence analysis demonstrated that all samples exhibited a uniform length of 321 base pairs, which also allowed them to be divided into two categories, some of which showed homomorphism. The study suggests that the accuracy and efficiency of Bupleurum germplasm resource identification can be enhanced through the application of ITS technology, especially when combined with ISSR markers. This underscores the potential of molecular identification techniques in the multi-faceted identification and assessment of genetic diversity within traditional Chinese medicine [10].

3.3 Application in the identification of years

The telomerase activity in the main root and cambium cells of Paeonia lactiflora was investigated by Cheng. The results showed that with the increase of the growth years, the telomerase activity in the main root decreased, whereas the telomerase activity in the cambium cells did not change significantly. According to the allometric growth theory, the area of the cambium and the cross-sectional area were calculated by measuring the radius of the cross-section and the outer diameter of the cambium. The mathematical relationship was established in the form of the function $P = 0.02n^{-0.5}$, where P represents the area ratio of the cambium to the cross-section, and n denotes the growth years of Paeonia lactiflora [11].

Liang et al. analyzed the ginseng specimens aged two to eight years from Ji'an and Fusong. They found that the telomerase activity in the taproot was highest, and the cambium and adjacent secondary lignin zone also showed a high level of telomerase activity. This suggests that the taproot is the most active meristematic region in Panax ginseng. Additionally, the researchers employed southern blotting to assess telomere restriction fragment (TRF) length, revealing a significant increase in TRF in the main root corresponding to the growth years. Based on these observations, a mathematical model was proposed in the form of $y = 0.827x + 8.231$, where x represents the growth age and y denotes the TRF. This model can be utilized to estimate the growth years of ginseng specimens aged two to eight years. Furthermore, the study underscores the efficacy of molecular identification techniques in the authentication of traditional Chinese medicine [12].

3.4 Application in origin identification

The source of traditional Chinese medicine (TCM) has a great influece on its quality, making the identification of its origins particularly crucial. Recent studies have increasingly employed molecular identification techniques to ascertain the origins of TCM.

Han et al. employed both morphological seed identification and DNA molecular identification methods to explore the genetic relationships among 28 Bupleurum seeds, aiming to lay a foundation for the identification of Bupleurum seeds by integrating DNA barcoding with seed morphological characteristics. Their analysis revealed that the sequences of Bupleurum chinense seeds clustered into a distinct branch based on ITS2 analysis, while black Bupleurum and another variety with narrow bamboo leaves also formed separate branches, demonstrating a clear monophyletic character. These findings were consistent with the morphological characteristics observed in Bupleurum chinense seeds [13].

The total DNA of Fructus thunbergii was extracted as reported by Long et al. [14]. The ITS2 sequence was subsequently amplified and sequenced in two directions. The species spacing and intraspecific Kimura-2-parameter (K2P) genetic distances were calculated, and a neighbor-joining (NJ) tree was constructed to illustrate systematic relationships. According to the method outlined in the 2025 edition of the Chinese Pharmacopoeia, the quality of medicine derived from Fructus Weilianensis was evaluated. Among the 30 batches analyzed, 7 were unqualified, and 8 batches did not meet total ash standards. Additionally, the cadmium content in 25 batches exceeded the standard, the copper content in 11 batches exceeded the standard, and the lead content in 3 batches exceeded the standard. These findings indicate a significant disparity in genetic distance between Yalian and other related species. This study further confirms that molecular identification using the ITS2 sequence is an effective method to identify Fructus Corydalis species and provides a valuable reference for the identification of other traditional Chinese medicine.

4 Conclusion

The fundamental principle of DNA molecular identification technology is predicated on the difference of genetic material between species, specifically the differences of DNA sequences. Therefore, DNA sequence analysis becomes the primary concern of most researchers. Despite the widespread adoption of numerous high-throughput sequencing techniques and a concomitant decrease in cost, the cost for each reaction is still high, ranging from thousands to tens of thousands of yuan. This financial burden makes the methods impractical for small-scale sequencing endeavors. In the context of identifying proprietary traditional Chinese medicine, the sequencing of PCR products and plasmids requires the analysis of tens to thousands of bases, which challenges many researchers, thereby limiting the acceptance of the technology [14].

With the continuous progress of sequencing technology, it is anticipated that the cost related to high-throughput sequencing will be further reduced, and the ability to handle large datasets will be enhanced. Therefore, high-throughput sequencing is expected to become a standard experimental approach, widely used in the identification of proprietary traditional Chinese medicine. To date, traditional identification methods and molecular identification techniques have operated independently, lacking effective integration. There is no comprehensive identification system to bridge “phenotype” and “heredity”. This disjunction hinders the realization of their complementary technological advantages, resulting in persistent challenges in the identification of proprietary traditional Chinese medicine. If these two methodologies can be integrated to characterize the origins of traditional Chinese medicine from both phenotypic and genetic perspectives, it will not only improve the accuracy of traditional Chinese medicine identification but also open up new avenues for its development.

Acknowledgements

This work was financially supported by National Nature Science Foundation of China (81973284) and Scientific Research Foundation of the Education Department of Liaoning Province (LJKZ0944).

References

[1] Huang LQ, Xiao PG, Guo WP, et al. Molecular Pharmacognosy: An Emerging Marginal Discipline. CN Sci, 2009, 39: 1101-1110.

[2] Shi SM, Pan MJ, Wang J, et al. Application of molecular identification techniques in traditional Chinese medicine. Herb Med, 2016, 47: 3121-3126.

[3] Vos P, Hogers R, Bleeker M, et al. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res, 1995, 23: 4407-4414.

[4] Zietkiewicz E, Rafalski A, Labuda D. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics, 1994, 20: 176-183.

[5] Chen SL, Pan XH, Luo K, et al. DNA barcoding technology for biological resources. Life Sci, 2013, 25: 458-466.

[6] Xing JX, Chen L, Zhang X, et al. SSR and ISSR methods for authenticity identification of Atractylodes macrocephala. Tropic Agricu Sci, 2019, 39: 42-47.

[7] Luo K, Ma P, Yao H, et al. Identification of Fritillaria cirrhosa and its adulterants’ original plants based on ITS2 sequences. Wor Sci Tec (Mod Trad CN Med), 2012, 14: 1153-1158.

[8] Liu YP, He BZ, Cao H, et al. Application of Gene Sequencing Technology in the Quality Research of

Traditional Chinese Medicine (II) - DNA Sequencing Identification of Yam Origin. Herb Med, 2001, 1: 69-73.

[9] Ma YZ, Ke SY. Analysis of ISSR and ITS Sequences of Different Bupleurum Germplasm Resources. J NW A&F Uni: Natu Sci Edit, 2015, 43: 193-200.

[10] Cheng CS, Lai CJS, Wei XY, et al. Research Progress on Identification Methods for the Growth Years of Traditional Chinese medicine. CN Trad Med J, 2021, 46: 1357-1367.

[11] Liang JB, Jiang C, Peng HS, et al. Analysis of the age of Panax ginseng based on telomere length and telomerase activity. Sci Report, 2015, 5: 7985.

[12] Han XW, Feng JX, Jin AX, et al. Microscopic and Molecular Identification of Bupleurum Seeds from Different Origins. North Horticult, 2021, 1: 121-127.

[13] Long L, Huang P, Zhai HW, et al. Comparison of Flavor and Quality from Different Origins and Molecular Identification of Nearby Source Plants. Mod Res Prac Trad CN Med, 2022, 36: 1-7.

[14] Cui ZH, Long P, Wang YL, et al. Application and Prospects of DNA Molecular Marker Technology in the Identification of Traditional Chinese Medicine. CN Med Herb, 2015, 38: 188-192.