Screening and Identification of DNA Barcode Sequences from Source Plants of Curcumae Rhizoma (Postprint)

Authors: Zhang Yuxiu, Liu Yang, Liu Peiwei, Fu Chuan-Kun, Lu Lilan, Wei Jianhe

Date: 2020-03-24T10:21:13+00:00

Abstract

To identify suitable DNA barcode sequences for authenticating the botanical origins of the Chinese medicinal material Curcumae Rhizoma (Ezhu) and to explore novel methods for its rapid and efficient identification, this study first evaluated seven DNA barcode sequences (ITS, ITS2, matK, psbA-trnH, trnL-trnF, rpoB, and atpB-rbcL) across three source plants and nine samples based on amplification and sequencing success rates. The obtained high-quality sequences were subsequently assessed through variable site analysis, genetic distance calculation, and phylogenetic tree analysis using MEGA6.0 software. Finally, the selected DNA barcode sequences were applied to identify the botanical origins of unknown test samples. The results demonstrated that barcode sequences such as ITS, ITS2, and matK exhibited low amplification or sequencing success rates in Ezhu source plants, limiting their practical application. Conversely, psbA-trnH, trnL-trnF, and rpoB barcode sequences contained insufficient variable site information to differentiate among the three distinct botanical origins of Ezhu. Only the atpB-rbcL barcode sequence showed high amplification and sequencing success rates, enabling facile acquisition of high-quality sequences with an ideal length (642–645 bp) and numerous variable sites (11), thus permitting discrimination among the three different botanical origins. The test sample was identified as Curcuma wenyujin based on the phylogenetic tree constructed from the atpB-rbcL sequence. These experimental results indicate that the chloroplast atpB-rbcL sequence can accurately identify different botanical origins of Ezhu and may serve as a DNA barcode sequence for authenticating the source plants of this Chinese medicinal material.

Full Text

Screening and Identification of DNA Barcode Sequences for the Original Plants of Curcumae Rhizoma

Yuxiu Zhang[1], Yang Liu[2], Peiwei Liu[1], Chuankun Fu[1], Lilan Lu[1], Jianhe Wei[1],[2]*

[1] Hainan Provincial Key Laboratory of Resources Conservation and Development of Southern Medicine, Hainan Branch of the Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Haikou 570311, China
[2] Key Laboratory of Bioactive Substances and Resources Utilization of Chinese Herbal Medicine, Ministry of Education & National Engineering Laboratory for Breeding of Endangered Medicinal Materials, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, China

Abstract: To identify suitable DNA barcode sequences for authenticating the original plant species of Curcumae Rhizoma (a traditional Chinese medicine) and explore a rapid and efficient identification method, this study first evaluated seven candidate DNA barcoding sequences (ITS, ITS2, matK, psbA-trnH, trnL-trnF, rpoB, and atpB-rbcL) across three original plant species and nine samples based on PCR amplification and sequencing success rates. The high-quality sequences were then further assessed using MEGA6.0 software through variable site analysis, genetic distance calculation, and phylogenetic tree construction. Finally, the selected barcode sequence was used to identify unknown test samples. The results showed that ITS, ITS2, and matK had low amplification or sequencing success rates in Curcumae Rhizoma original plants, making them unsuitable for practical identification. Meanwhile, psbA-trnH, trnL-trnF, and rpoB contained insufficient variable site information to distinguish among the three different original species. Only the atpB-rbcL barcode sequence demonstrated both high amplification and sequencing success rates, enabling easy acquisition of high-quality sequences with ideal length (642–645 bp) and abundant variable sites (11 sites), thus allowing differentiation of the three original species. Test samples were identified as *Curcuma wenyujin* based on phylogenetic tree analysis. These findings demonstrate that the chloroplast atpB-rbcL sequence can accurately identify different original plant species of Curcumae Rhizoma and serves as a suitable DNA barcode for this purpose.

Keywords: Curcumae Rhizoma, DNA barcode, screening, atpB-rbcL, species identification

---

Curcumae Rhizoma is processed from the rhizomes of three species: *Curcuma kwangsiensis* S. G. Lee et C. F. Liang, *C. phaeocaulis* Val., and *C. wenyujin* Y. H. Chen et C. Ling. It is a commonly used traditional Chinese medicine in China with functions of promoting qi circulation, breaking blood stasis, and reliev-

ing pain by dissipating accumulations (Chinese Pharmacopoeia, 2015). Modern research indicates significant differences in chemical composition and pharmacological effects among Curcumae Rhizoma from different sources (Wu and Li, 2012). The contents of major active compounds—including curdione, curcumol, germacrone, and β-elemene—vary markedly among samples from different origins (Mao et al., 2013). Pharmacological studies have shown that *C. wenyujin* exhibits the best therapeutic effects against gastric and liver cancers (Tang et al., 2013; Zang et al., 2014), and the primary component of the gynecological drug Baofukang suppository, curcuma oil, is mainly derived from *C. wenyujin* (Zhang and Yang, 2010). With increasing market demand for Curcumae Rhizoma both domestically and internationally, artificial cultivation has become the main source of medicinal material (Wu and Zhao, 2003). To ensure correct cultivar selection and safe, effective clinical use, accurate identification of the original plant species is urgently needed.

All three original species belong to the genus *Curcuma*. Traditional identification of this genus relies primarily on morphological characteristics of flowers and leaves (Záveská et al., 2012; Shen et al., 2014; Flora of China, 2004). However, the three original species of Curcumae Rhizoma share highly similar morphological features that are easily confused, making traditional identification and classification difficult (Flora of China, 2004; Xiao et al., 2004). Factors such as artificial cultivation and mutual introduction across regions further complicate classification. Therefore, it is necessary to clarify the original species and germplasm of cultivated Curcumae Rhizoma to ensure proper planting. Researchers have employed various methods including microscopic identification, physicochemical analysis, RAPD analysis, and chemical fingerprinting (Yang et al., 2005; Shen et al., 2014; Xiao et al., 2000/2001; Liu et al., 2016), but these approaches face practical difficulties or excessive sequence length that greatly increases application difficulty (Cao et al., 2010).

This study screened multiple DNA barcode sequences to establish a DNA barcoding identification method suitable for the three original species of Curcumae Rhizoma, while also providing a reference for the classification and identification of *Curcuma* plants.

1.1 Materials

A total of 14 materials were collected, including 9 test samples and 5 samples for identification. *Curcuma phaeocaulis*, *C. kwangsiensis*, and *C. wenyujin* were collected from authentic production regions including Wenzhou, Guangxi, and Sichuan, while test samples were collected from introduction sites in Hainan and Jiangxi. Detailed information is provided in Table 1 . All materials were identified by Professor Zhu Ping from the Hainan Branch of the Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences.

1.2 Methods

1.2.1 DNA Extraction, Amplification, and Sequencing For each sample, 100 mg of fresh leaf tissue was used for DNA extraction following the OMEGA HP Plant DNA Kit protocol (OMEGA, USA). Extracted total DNA quality was assessed using a micro-volume spectrophotometer (Thermo, USA). PCR primers and amplification protocols followed published references (Table 2). The 25 L PCR reaction contained 12.5 L 2× EcoTaq PCR SuperMix (Transgen, China), 1 L each of forward and reverse primers, and 20 ng DNA template. PCR products were detected by 1% agarose gel electrophoresis. Successfully amplified products were purified and bidirectionally sequenced by BGI.

1.2.2 Data Processing Sequencing chromatograms were proofread and assembled using CodonCode Aligner V2.0 (CodonCode Co., USA) to remove low-quality sequences and primer regions. Similarity searches (BLAST) were performed for error checking, and validated sequences were registered in GenBank with new accession numbers (Table 1). MEGA6.0 software was used for multiple sequence alignment, calculation of intra- or inter-specific Kimura 2-parameter (K2P) genetic distances, and construction of NJ phylogenetic trees.

2.1 PCR Amplification and Sequencing

Although ITS and ITS2 sequences achieved 100% PCR success, their sequencing chromatograms showed irregular multiple peaks, preventing acquisition of accurate single sequences. The average amplification success rate for matK was only 18.75%. In contrast, psbA-trnH, rpoB, trnL-trnF, and atpB-rbcL all achieved 100% PCR amplification and sequencing success. Sequence lengths ranged from 325 bp to 645 bp, with psbA-trnH and rpoB showing 0 variable sites, trnL-trnF showing 1 variable site, and atpB-rbcL showing 11 variable sites across 14 samples (Table 3).

2.2 Sequence Characteristic Analysis

Comparative analysis of atpB-rbcL sequences revealed that *C. phaeocaulis* had a sequence length of 642 bp, while both *C. kwangsiensis* and *C. wenyujin* had 645 bp. Compared with *C. wenyujin*, *C. kwangsiensis* showed only a C-to-T transition at position 550, but differed from *C. phaeocaulis* at 8 base positions, including a C-to-T transition at position 76, T insertion at position 259, C-to-T transition at position 310, C-to-T transition at position 397, and a 4-base T deletion at position 430. Among test samples, the atpB-rbcL sequences from Xinyu (Jiangxi) and Chengmai/Lingao (Hainan) showed 100% identity with Wenzhou *C. wenyujin*, while Qiongshan (Hainan) differed by a single base.

2.3 Genetic Distance Calculation and Phylogenetic Tree Based on atpB-rbcL

Using MEGA6.0 software and the K2P distance model, genetic distances among different original species were calculated. Intra-specific genetic distances were 0 for *C. kwangsiensis*, 0.0004 for *C. wenyujin*, and 0.0063 for *C. phaeocaulis*. Inter-specific average distances were 0.0018 between *C. wenyujin* and *C. kwangsiensis*, 0.0081 between *C. wenyujin* and *C. phaeocaulis*, and 0.0095 between *C. kwangsiensis* and *C. phaeocaulis*. For test samples, genetic distances between Hainan samples and Wenzhou *C. wenyujin* ranged from 0 to 0.0016 (average 0.0005), while distances between Jiangxi samples and Wenzhou *C. wenyujin* were all 0.

An NJ phylogenetic tree based on atpB-rbcL sequences was constructed for *C. kwangsiensis*, *C. wenyujin*, and *C. phaeocaulis* (Figure 1 [Figure 1: see original paper]). The tree showed that three populations of *C. wenyujin*, four populations of *C. kwangsiensis*, and two populations of *C. phaeocaulis* each formed monophyletic clades. *Curcuma wenyujin* and *C. kwangsiensis* clustered together as one major branch, while *C. phaeocaulis* formed a separate branch, indicating that the NJ tree based on atpB-rbcL clearly distinguished the three species, with *C. wenyujin* and *C. kwangsiensis* being more closely related than either is to *C. phaeocaulis*. All test samples clustered with Wenzhou *C. wenyujin* (Figure 1).

Since its proposal in 2003, DNA barcoding has become a global hotspot and direction in taxonomic research (Chen et al., 2013). Due to lower nucleotide evolution rates in plants compared to animals, and the prevalence of hybridization and polyploidization events, no single plant region functions as efficiently and universally as the animal COI gene. Therefore, screening one or more suitable DNA barcode sequences for plant classification has become an important research focus (Chen et al., 2010; Group et al., 2011). This study evaluated seven candidate DNA barcode sequences (ITS, ITS2, atpB-rbcL, matK, psbA-trnH, trnL-trnF, and rpoB) and identified atpB-rbcL as suitable for identifying original species of Curcumae Rhizoma based on PCR amplification difficulty, sequencing success rate, and degree of variation.

Nuclear genes ITS and ITS2 showed very low sequencing success in *Curcuma* plants, which Chen et al. attributed primarily to polyploid hybridization and long-term artificial cultivation (Chen et al., 2014). Although ITS2 sequences could be obtained through cloning, this increases cost and complexity while providing only 46.7% resolution within *Curcuma* (Chen et al., 2014). matK is one of the core plant DNA barcode sequences (Leister et al., 1998), but showed the lowest PCR amplification efficiency (18.75%) in this study, a problem frequently reported in other plants (Chen et al., 2014; Sass et al., 2007; Hollingsworth et al., 2009). Chen (2014) improved matK amplification success to 85.4% in *Curcuma* through multiple primer pairs and optimization, yet found no barcode gap in *Curcuma* matK sequences (Chen et al., 2014). Our analysis revealed

that matK sequences could not differentiate the original species of Curcumae Rhizoma. Based on PCR amplification and sequencing success, we conclude that matK, ITS, and ITS2 are unsuitable candidates for identifying original species of Curcumae Rhizoma.

The four sequences psbA-trnH, rpoB, trnL-trnF, and atpB-rbcL all achieved 100% amplification and sequencing success in this study. However, the first three were extremely conserved with slow evolution rates, showing zero or minimal inter-specific differences, while atpB-rbcL exhibited rich variation, making it suitable for differentiating original species. Based on atpB-rbcL sequences, this study clearly distinguished the three original species listed in the Chinese Pharmacopoeia through genetic distance calculation and NJ tree construction, while identification results confirmed that samples introduced to Hainan and Jiangxi were all *C. wenyujin*.

With increasing cultivation of Curcumae Rhizoma in Hainan in recent years, this study applied DNA barcoding technology to identify the introduced species as *C. wenyujin*, providing assurance for its rational development and utilization. This study provides a preliminary exploration of DNA barcoding identification for Curcumae Rhizoma original plants. The selected atpB-rbcL sequence is significant for rapid and accurate identification of different original species, though further sampling expansion is necessary given the numerous cultivated varieties and wide distribution range. This research aims to provide a feasible method for cultivar selection and introduction cultivation of Curcumae Rhizoma original plants, while offering reference for classification and identification of *Curcuma* plants.

References

Cao H, Sasaki Y, Fushimi H, et al., 2010. Authentication of Curcuma species (Zingiberaceae) based on nuclear 18S rDNA and plastid trnK sequences[J]. Acta Pharm Sin, 45(7): 926-933.

Chen J, Zhao J, Erickson DL, et al., 2014. Testing DNA barcodes in closely related species of Curcuma (Zingiberaceae) from Myanmar and China[J]. Mol Ecol Resour, 5(2):337-348.

Chen SL, Yao H, Han JP, et al., 2010. Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species[J]. PLoS ONE, 5(1): e8613.

Chen SL, Yao H, Han JP, et al., 2013. Principles for molecular identification of traditional Chinese materia medica using DNA barcoding[J]. Chin J Chin Mat Med, 38(2):141-148.

Chen SL, 2015. Standard DNA barcodes Chinese for materia medica in Chinese Pharmacopoeia[M]. Beijing: Science Press.

Chiang T, 1998. Universal primers for amplification and sequencing a noncoding spacer between atpB and rbcL genes of chloroplast DNA[J]. Bot Bull Acad Sin,

39(4): 245-250.

Chinese Pharmacopoeia Commission, 2015. Pharmacopoeia of the People's Republic of China[M]. Beijing: China Medical Science Press.

Delectis Florae Reipublicae Popularis Sinicae Agendae Academiae Sinicae Edita, 2004. Flora of China[M]. Beijing: Science Press.

Group CB, Li DZ, Gao LM, et al., 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants[J]. P Natl Acad Sci Usa, 108(49): 11964-19641.

Hollingsworth ML, Andra CA, Forrest LL, et al., 2009. Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants[J]. Mol Ecol Resour, 9(2): 439-457.

Lee SY, Ng WL, Mahat MN, et al., 2016. DNA Barcoding of the Endangered Aquilaria (Thymelaeaceae) and Its Application in Species Authentication of Agarwood Products Traded in the Market[J]. PLoS ONE, 11(4): e0154631.

Leister D, Kurth J, Laurie DA, et al., 1998. Rapid reorganization of resistance gene homologues in cereal genomes[J]. P Natl Acad Sci Usa, 95(1): 370-375.

Liu SL, Wang XH, Jiang CX, et al., 2016. RAPD-PCR analysis on different germplasm resources of Curcumae Rhizoma[J]. Chin Tradd Herb Drugs, 47(17): 3098-3102.

Mao CQ, Lu TL, Jiang HF, et al., 2013. Determination of curdione, curcumol, germacrone, and β-elemene in processed Curcumae Rhizoma from different habitats[J]. Chin Trad Herb Drugs, 44(3): 2834-2837.

Sass C, Little PD, Stevenson WD, et al., 2007. DNA Barcoding in the Cycadales: Testing the potential of proposed barcoding markers for species identification of cycads[J]. PLoS ONE, 2(11): e1154.

Shen LL, Xu LZ, Lü JQ, et al., 2014. Rhizome morphology and flowering characteristic of 5 species Curcuma plants[J]. J Zhongkai Univ Agric Eng, 27(4): 11-14.

Shen Y, Li Y, Liu TT, et al., 2014. Microscopical and physicochemical identification of 3 sources of Curcumae Rhizoma[J]. Chemical Research and Application, 26(2): 267-270.

Tang DC, Zang WH, Feng HH, 2013. Influences of medicated serum of different cultivars of Ezhu (Rhizoma Zedoariae) on proliferation, apoptosis and nucleocytoplasmic ration human gastric carcinoma cells BGC823[J]. J Beijing Univ Trad Chin Med, 36(4): 254-257+26.

Wu BY, Li M, 2012. Study on the correlation between chemical composition and quality of rhizoma zedoary rhizome[J]. J Chengdu Univ TCM, 35(3): 48-49+56.

Wu ZQ, Zhao XW, 2003. A preliminary discussion about introduction of new varieties of chinese herbal medicines from separate regions and their domestication

from wildness[J]. J Agric Sci Technol, 5(3): 57-59.

Xiao XH, Liu SQ, Shi CH, et al., 2000. RAPD polymorphism and authentication of medicinal plants from Turmeric (Curcuma L.) in China[J]. Chin Trad Herb Drugs, 31(3): 209-212.

Xiao XH, Xia WJ, Qin SY, et al., 2001. Pattern recognition of stereoscopic features of the leaves epidermis of Medicinal Curcuma plants in china by image analysis[J]. Chin J Chin Mat Med, 26(8): 523-528.

Xiao XH, Zhong GY, Shu GM, et al., 2004. Numerical taxonomy of medicinal plants of Curcuma in China[J]. Chin J Chin Mat Med, 29(1): 15-24.

Yang FQ, Li SP, Chen Y, et al., 2005. The fingerprint of Ezhu by GC-MS [J]. Acta Pharm Sin, 40(11): 1013-1018.

Zang WH, Tang DC, Ying G, et al., 2014. Inhibition effects of compatibility of Astragali radix and Curcumae Rhizoma and combination with cisplatin on orthotopic tumor in nude mice bearing human hepatocellular carcinoma[J]. Chin J Exp Trad Med Form, 20(5): 131-136.

Záveská E, Fér T, Šída O, et al., 2012. Phylogeny of Curcuma (Zingiberaceae) based on plastid and nuclear sequences: Proposal of the new subgenus Ecomata[J]. Taxon, 61(4): 747-762.

Zhang YW, Yang RQ, 2010. Problems and proposals on the standards of curcuma oil and its preparations[J]. Drug Stand Chin, 11(3): 170-171+23.

Note: Figure translations are in progress. See original paper for figures.