



The Species Identification in Traditional Herbal Patent Medicine, Wuhu San, Based on Shotgun Metabarcoding

Jinxin Liu^{1,2†}, Weishan Mu^{1†}, Mengmeng Shi¹, Qing Zhao¹, Weijun Kong², Hongbo Xie¹ and Linchun Shi^{2*}

¹Hebei Key Laboratory of Study and Exploitation of Chinese Medicine, Chengde Medical University, Chengde, China, ²Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China

OPEN ACCESS

Edited by:

Sukvinder Kaur Bhamra,
University of Kent, United Kingdom

Reviewed by:

Johanna Mahwahwase Bapela,
University of Pretoria, South Africa

Shiv Bahadur,

GLA University, India

Tiziana Sgamma,

De Montfort University,

United Kingdom

*Correspondence:

Linchun Shi
linchun_shi@163.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to

Ethnopharmacology,

a section of the journal

Frontiers in Pharmacology

Received: 16 September 2020

Accepted: 18 January 2021

Published: 16 February 2021

Citation:

Liu J, Mu W, Shi M, Zhao Q, Kong W,
Xie H and Shi L (2021) The Species
Identification in Traditional Herbal
Patent Medicine, Wuhu San, Based on
Shotgun Metabarcoding.
Front. Pharmacol. 12:607200.
doi: 10.3389/fphar.2021.607200

Traditional herbal patent medicine typically consists of multiple ingredients, making it challenging to supervise contamination by impurities and the improper use of raw materials. This study employed shotgun metabarcoding for the species identification of biological ingredients in traditional herbal patent medicine, Wuhu San. The five prescribed herbal materials found in Wuhu San were collected, and their reference sequences were obtained by traditional DNA barcoding using Sanger sequencing. Two lab-made and three commercial Wuhu San samples were collected, and a total of 37.14 Gb of shotgun sequencing data was obtained for these five samples using the Illumina sequencing platform. A total of 1,421,013 paired-end reads were enriched for the Internal Transcribed Spacer 2 (ITS2), *psbA* and *trnH* intergenic spacer region (*psbA-trnH*), maturase *k* (*matK*), and ribulose-1, 5-bisphosphate carboxylase (*rbcl*) regions. Furthermore, 80, 11, 9, and 8 operational taxonomic units were obtained for the ITS2, *psbA-trnH*, *matK*, and *rbcl* regions, respectively, after metagenomic assembly, annotation, and chimeric detection. In the two lab-made mock samples, all labeled ingredients in the Wuhu San prescription were successfully detected, and the positive control, *Panax quinquefolius* L., was detected in the HSZY172 mock sample. Three species, namely *Angelica sinensis* (Oliv.) Diels, *Saposhnikovia divaricata* (Turcz. ex Ledeb.) Schischk., and *Carthamus tinctorius* L., belonging to three labeled ingredients, *Angelicae Sinensis Radix* (Danggui), *Saposhnikoviae Radix* (Fangfeng), and *Carthami Flos* (Honghua), were detected in the three commercial samples. *Angelica dahurica* (Hoffm.) Benth. & Hook. f. ex Franch. & Sav., the original *Angelicae Dahuricae Radix* (Baizhi) species, was only detected in WHS003. *Arisaema erubescens* (Wall.) Schott, *Arisaema heterophyllum* Blume, or *Arisaema amurense* Maxim., the original *Arisaematis Rhizoma* (Tiannanxing) species, were not detected in any of the commercial samples, which could be attributed to the fact that this medicinal material underwent extensive processing. In addition, the *Saposhnikovia divaricata* adulterant was detected in all the commercial samples, while 24 fungal genera, including *Aspergillus*, were identified in both the lab-made and commercial samples. This study showed that shotgun metabarcoding provided alternative strategy and technical

means for identifying prescribed ingredients in traditional herbal patent medicine and displayed the potential to effectively complement traditional methods.

Keywords: Wuhu San, shotgun metabarcoding, DNA barcoding, traditional herbal patent medicine, species identification

INTRODUCTION

In recent years, traditional herbal medicine has been widely used to prevent and treat clinical diseases. Many countries have been using traditional herbs to prevent disease or improve health to varying degrees (Barnes et al., 2016; Job et al., 2016; Sammons et al., 2016; Teng et al., 2016). It is difficult to identify the specific content of traditional herbal patent medicines since they mostly consist of multiple mixed ingredients. Microscopic and physicochemical identification are currently the primary methods used for quality control and the determination of traditional herbal patent medicine content (Committee, 2020). However, the microscopic characteristics of the medicinal materials within multiple original plants may be inconsistent (Chen et al., 1998). Furthermore, insufficient professional talent has also restricted the development of microscopic identification. Physical and chemical identification is based on chemical properties. However, the chemical composition of traditional herbal patent medicine is complex, and the correspondence between the chemical composition and different prescription ingredients may not be clear. In addition, many factors, such as the original plant, environment, harvesting, and processing, may affect the content of active ingredients. It is also possible that some manufacturers illegally add chemical substances, complicating the quality control of traditional herbal patent medicine via chemical composition detection (Xu et al., 2014; Li et al., 2015b). With the development of high-throughput sequencing (HTS) technology, shotgun metagenomics based on the genetic information of species has been successfully applied to identify ingredients in mixed samples. This technique involves the untargeted sequencing of all biological ingredient genomes present in a sample (Quince et al., 2017) to break the metagenomic DNA into small fragments, after which bioinformatics methods are used for assembly without the need for PCR amplification. Therefore, potential biases caused during PCR amplification can be eliminated, and multiple DNA barcodes can be obtained simultaneously for further study to produce more comprehensive data. If the information is used to analyze the traditional DNA barcode region, it is known as shotgun metabarcoding. This was applied here for the species identification of the biological ingredients in traditional herbal patent medicine. Furthermore, shotgun metabarcoding can be a powerful supplement to the conventional identification method used for traditional herbal patent medicine.

Currently, shotgun sequencing technology is primarily used in microbiology to study the composition and functions of microbial communities in different environment samples (Tringe et al., 2005; Warden et al., 2016). A comprehensive study of the microbiome during different processing steps in the beef production chain revealed that the relative abundance of

common pathogenic and non-pathogenic bacteria decreased significantly in the final stage, while the relative abundance of some bacteria or pathogens increased. The study proved that shotgun sequencing technology could be used to evaluate the microbial community composition during beef production, as well as pathogen population shifts (Yang et al., 2016). Several studies have shown that shotgun sequencing technology is also applicable to the study of microbial communities in food or beverages that require fermentation (Ferrocino et al., 2018; Arikan et al., 2020), as well as human microbes found in the skin (Oh et al., 2014), saliva (Hasan et al., 2014), and gastrointestinal tract (Vangay et al., 2018; Zhao et al., 2018). In addition to research in the field of microbiology, shotgun sequencing technology has also been successfully used in animal diet analysis (Srivathsan et al., 2015), animal diversity (Zhou et al., 2013), and ingredient identification in food (Haiminen et al., 2019). The development of high-throughput sequencing technology has allowed the application of research strategies based on DNA barcodes to identify traditional herbal patent medicine. A previous study used high-throughput sequencing to identify the biological components in Yimu Wan, a traditional patent medicine. The results showed that all the prescription ingredients could be detected based on the ITS2 sequences, indicating that this technique can be used effectively to detect the legality and safety of Yimu Wan (Jia et al., 2017). Another study used single-molecule, real-time sequencing to identify multiple ingredients in Jiwei Qianghuo Wan, and the result showed that seven prescription ingredients and positive controls were successfully detected in the two reference samples. Adulterants and potential contaminant species were also found in the commercial samples, indicating that this method can effectively detect the biological components of Chinese patent medicines (Xin et al., 2018b). Furthermore, a study based on high-throughput sequencing and ITS2 regions detected some prescription ingredient adulterants (Cangzhu and Tiannanxing) in traditional Ruyi Jinhuang San medicine (Shi et al., 2018). However, minimal studies are available involving the species identification in traditional herbal patent medicine based on shotgun sequencing. Xin et al. reported the first systematic study involving species detection in traditional herbal patent medicine based on shotgun sequencing (Xin et al., 2018a). The results showed that the ITS2 region could detect all the prescription ingredients, as well as the positive control in the mock samples of Longdan Xiegan Wan. This confirms that shotgun metagenomic sequencing can be used to identify the biological ingredients in traditional herbal patent medicine.

Wuhu San was first recorded in the book, *Si He Ting Ji Fang*, written by Ling Huan during the Qing Dynasty (Deng et al., 2000). It is a type of powder patent medicine prepared by mixing five Chinese medicinal materials, namely *Angelicae Sinensis*

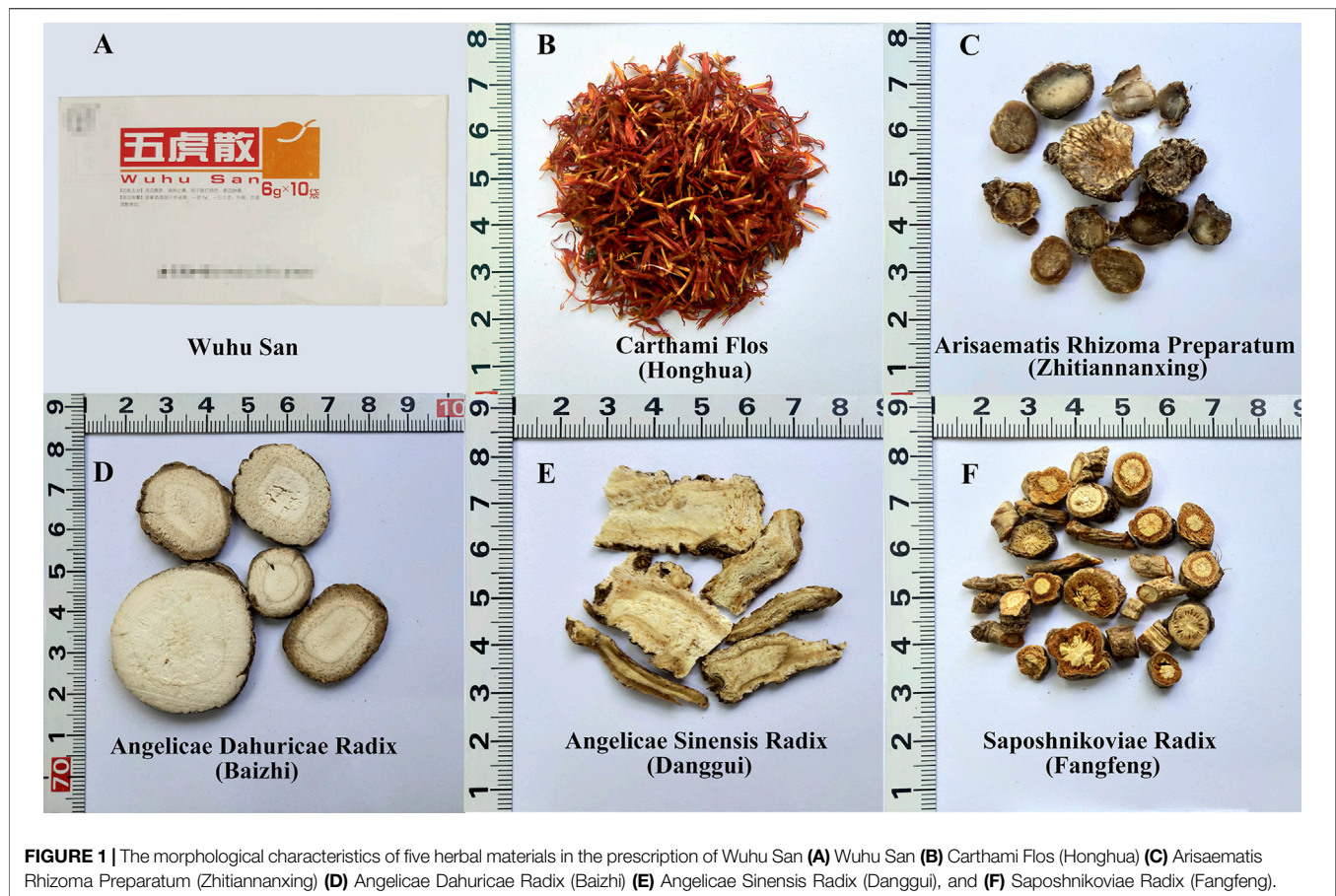


FIGURE 1 | The morphological characteristics of five herbal materials in the prescription of Wuhu San (A) Wuhu San (B) Carthami Flos (Honghua) (C) Arisaematis Rhizoma Preparatum (Zhitiannanxing) (D) Angelicae Dahuricae Radix (Baizhi) (E) Angelicae Sinensis Radix (Danggui), and (F) Saposhnikoviae Radix (Fangfeng).

Radix (Danggui), Saposhnikoviae Radix (Fangfeng), Angelicae Dahuricae Radix (Baizhi), Carthami Flos (Honghua), and Arisaematis Rhizoma Preparatum (Zhitiannanxing). The corresponding Latin names of the original species are shown in **Supplementary Table S1**. It promotes blood circulation, relieves pain, reduces swelling, and disperses blood stasis. Not only can it be used externally with white wine, but it can also be administered with warm yellow rice wine or warm, boiled water. Pharmacological studies have shown that Wuhu San has a relatively apparent anticoagulant effect (Jia et al., 1997), while its alcohol extract displays excellent anti-inflammatory and analgesic properties (Yang et al., 1990). Although both the active substances in Angelicae Sinensis Radix (Danggui) and Carthami Flos (Honghua) have an excellent inhibitory effect on platelet aggregation and can improve blood flow (Li et al., 2009; Zhang et al., 2009), combining these two medicinal materials is more effective in promoting blood circulation (Yue et al., 2017). Moreover, prescription ingredients, such as Saposhnikoviae Radix (Fangfeng), Angelicae Dahuricae Radix (Baizhi), and Arisaematis Rhizoma (Tiannanxing) display certain anti-inflammatory and analgesic properties (Okuyama et al., 2001; Kang et al., 2008; Chunna et al., 2015).

This study uses Wuhu San as an example to evaluate the feasibility and efficacy of using shotgun metabarcoding to identify the biological ingredients in traditional herbal patent medicine.

Two mock samples were prepared according to the official species composition listed in the 2015 edition of the Chinese Pharmacopoeia and used to verify the feasibility of the shotgun metabarcoding method. This technique was then employed to determine the biological species composition in the commercial Wuhu San samples, aiming to provide different strategies and technical means for the prescription ingredient identification and quality control of traditional herbal patent medicine, such as Wuhu San.

MATERIALS AND METHODS

Herbal Material, Lab-Made Mock Wuhu San, and Commercial Wuhu San Samples

Four kinds of Wuhu San herbal materials, namely Angelicae Sinensis Radix (Danggui), Saposhnikoviae Radix (Fangfeng), Angelicae Dahuricae Radix (Baizhi), Carthami Flos (Honghua), were collected from the Beijing TRT pharmaceutical company. Arisaematis Rhizoma (Tiannanxing) was obtained from Chengde, Hebei Province (**Supplementary Table S2** and **Figure 1**). The herbal materials were authenticated using the morphological and traditional DNA barcoding methods. The lab-made mock samples were prepared according to the prescription ingredients and manufacturing

TABLE 1 | The proportion of the five ingredients listed in the prescription of Wuhu San.

Pinyin name	Latin name	Proportion in the preparation (%)
Danggui	Angelicae Sinensis Radix	21.3
Honghua	Carthami Flos	21.3
Fangfeng	Saposhnikovia Radix	21.3
Zhitiannanxing	Arisaematis Rhizoma Preparatum	21.3
Baizhi	Angelicae Dahuricae Radix	14.6

method of Wuhu San listed in the 2015 edition of the Chinese Pharmacopoeia (Table 1). Of these, *Arisaematis Rhizoma* (Tiannanxing) was processed in advance following the method described in the Chinese Pharmacopoeia for *Arisaematis Rhizoma Preparatum* (Zhitiannanxing). The two lab-made mock samples were prepared and labeled as HSZY160 and HSZY172. The *Panaxis Quinquefolii Radix* (Xiyangshen) powder was added to HSZY172 as a positive control at an amount equal to *Angelicae Dahuricae Radix* (Baizhi), representing the lowest herbal ingredient in the Wuhu San prescription. In addition, the three commercial Wuhu San samples were acquired from pharmacies and labeled WHS001, WHS002, and WHS003.

DNA Extraction, PCR Amplification, Sanger Sequencing, and HTS

The DNA extraction of the herbal material samples was performed according to previous research (Liu et al., 2017) and the DNA barcoding principles for traditional Chinese herbal medicine (Chen et al., 2014) using a plant genomic DNA extraction kit (Tiangen Biochemical Technology (Beijing) Co., Ltd., China). The meta-genomic DNA of Wuhu San was extracted according to the previously published protocols of the CTAB-based method (Cheng et al., 2015) with some changes. A pre-wash buffer was used for pretreatment (Xin et al., 2018a), after which lysis buffer was added. The samples were placed in a 56°C water bath overnight for lysis. Extraction was performed using chloroform/isoamyl alcohol (volume ratio 24:1), and phenol/chloroform/isoamyl alcohol (volume ratio 25:24:1). The DNA was purified by adding 50 µL of sodium acetate and 1,250 µL of 100% methanol. The extracted DNA quality was estimated using a NanoDrop one ultra-micro spectrophotometer (Thermo Fisher Scientific Inc., USA). The traditional DNA barcoding regions of ITS2, *psbA-trnH*, *matK*, and *rbcl* were amplified with DNA barcoding primer sets and conditions proposed by the barcodes of the traditional Chinese herbal medicine data system (TCM-BOL) (Chen et al., 2014), the CBOL plant working group (Group et al., 2009), and the barcode of life data system (BOLD) (Ratnasingham and Hebert, 2007) using 2 × Taq master mix (AidLab Biotechnologies Co., Ltd., China). The PCR products were bi-directionally sequenced on an ABI 3730xL DNA Analyzer (ThermoFisher Co., Ltd., United States). After constructing a PCR-free library, the Wuhu San DNA was sheared into fragments and sequenced using the Illumina NovaSeq platform.

Data Analysis

The Sanger sequencing results were obtained according to the *Standard DNA barcodes of Chinese Materia Medica in Chinese Pharmacopoeia* edited by Chen Shilin (Chen, 2015). The sequence chromatograms were assembled, and the primers were removed using Codoncode aligner v 9.0.1 (CodonCode Corp., Dedham, MA, United States). For the Illumina sequencing data, the sequencing adapter and low-quality reads were filtered using Trimmomatic v0.38 (Bolger et al., 2014). The paired-end reads were enriched using local python scripts (Shi et al., 2019). The enriched reads belonging to ITS2, *psbA-trnH*, *matK*, and *rbcl* were assembled using MEGAHIT v1.2.9 and MetaSPAdes v3.13.2 (Li et al., 2015a; Nurk et al., 2017). Contigs obtained via the two types of software were merged, and duplicates were removed with cd-hit at 100% identity (Li and Godzik, 2006). The ITS2 regions were obtained using the hidden Markov model (HMM)-based annotation method (Keller et al., 2009). The traditional DNA barcoding regions of *psbA-trnH*, *matK*, and *rbcl* were acquired by removing primer sequences based on Cutadapt v2.10 (Kechin et al., 2017). The chimera detection of the annotated contigs was performed using UCHIME v4.2 (Edgar et al., 2011). Sequences belonging to each marker were clustered into OTUs at 100% identity using Usearch v11 (<https://www.drive5.com/usearch/>), and the representative sequence for each OTU was selected for further analysis. The shotgun paired-end reads were mapped to the OTU representative sequences using bowtie2 v2.4.1 (Langmead and Salzberg, 2012), while the sequencing depth and coverage values were calculated using Samtools v1.10 (Etherington et al., 2015). Poor-quality OTUs were removed when its representative sequences displayed a sequencing depth ≤ 3 or coverage $\leq 95\%$. The remaining high-quality OTUs were used for species assignment by searching the TCM-BOL (Chen et al., 2014), BOLD (Ratnasingham and Hebert, 2007), and GenBank (Benson et al., 2018) databases using the basic local alignment search tool, BLAST (Camacho et al., 2009). Finally, the statistics and taxonomic visualization of the species composition of the traditional herbal patent medicine were performed using MEGAN v6.18.9 (Huson et al., 2016).

After the species in the traditional herbal patent medicine were identified via DNA barcodes, some terms related to the identified species were defined as follows:

Authentic

The species in the medicinal materials are authentic if it is identified as one of the labeled ingredients in the prescription of the traditional herbal patent medicine.

Substitution

Substitutions refer to the species in the medicinal materials with similar characteristics such as efficacy, chemical composition, pharmacological effect, and clinical effect, which are selected instead of authentic medicinal materials according to the clinical medication plan when there is a shortage of these materials (Tang and Huang, 1994; Suo and Chen, 2006).

Adulterant

Adulterants refer to the species in the medicinal materials that are used as authentic although they are similar in appearance or have the same name as the authentic material, but are different regarding the

TABLE 2 | The GenBank accession numbers of the five herbal materials in Wuhu San and the positive control, *Panacis Quinquefolii Radix* (Xiyangshen).

Herb medicinal material	ITS2	<i>psbA-trnH</i>	<i>matK</i>	<i>rbcl</i>
Angelicae Sinensis Radix (Danggui)	MN727081	MT994327	MN729559	MN746764
Carthami Flos (Honghua)	MN727076	MT994328	MN729561	MN746766
Saposhnikoviae Radix (Fangfeng)	MT821449	MT994331	MW000338	MW000334
Arisaematis Rhizoma (Tiannanxing)	MT821451	—	MW000340	MW000335
Angelicae Dahuricae Radix (Baizhi)	MT821450	MT994330	MW000339	—
Panacis Quinquefolii Radix (Xiyangshen)	MT102865	MT994329	MW000341	MW000333

Note: "—" indicates that the sequence was not successfully obtained.

original plant source, chemical composition, pharmacological effect, and clinical effect (Tang and Huang, 1994).

Contaminant

Contaminants include fungal contamination and impurities.

RESULTS

The Authentication of the Five Herbal Materials in Wuhu San and Their ITS2, *psbA-trnH*, *matK*, and *rbcl* DNA Barcodes

The five herbal materials labeled on Wuhu San prescription were collected from Chengde (Hebei province). They were first identified using the morphological method and then authenticated using DNA barcoding to ensure the accuracy of the mock samples. High-quality DNA was extracted from these materials, after which the ITS2, *psbA-trnH*, *matK*, and *rbcl* DNA barcodes were amplified using their corresponding universal primers. Except for the *psbA-trnH* sequence of *Arisaematis Rhizoma* (Tiannanxing) and the *rbcl* sequence of *Angelicae Dahuricae Radix* (Baizhi), all the ITS2, *psbA-trnH*, *matK*, and *rbcl* DNA barcodes of the five herbal materials were successfully amplified and then bi-directionally sequenced using Sanger sequencing technology. The GenBank accession numbers of these sequences are shown in **Table 2**. The ITS2 and *psbA-trnH* sequences obtained via Sanger sequencing were assigned to species by blasting to the TCM-BOL system, while the *matK* and *rbcl* DNA barcodes obtained using the same method were assigned to a species or genus using the BOLD system and GenBank NT database. By combining the identification results of the four DNA barcodes, all five herbal materials were authenticated, and their original *Angelicae Sinensis Radix* (Danggui), *Saposhnikoviae Radix* (Fangfeng), *Angelicae Dahuricae Radix* (Baizhi), *Carthami Flos* (Honghua), and *Arisaematis Rhizoma Preparatum* (Zhitiannanxing) species were assigned to *Angelica sinensis* (Oliv.) Diels, *Saposhnikovia divaricata* (Turcz. ex Ledeb.) Schischk., *Carthamus tinctorius* L., *Angelica dahurica* (Hoffm.) Benth. & Hook. f. ex Franch. & Sav., and *Arisaema amurense* Maxim., respectively.

HTS and Shotgun Metabarcoding Data Assembly

The average DNA concentration of the lab-made mock samples and commercial samples was 144.06 ng/μL, while the A_{260}/A_{280} ranged between 1.8 and 2.0 (**Supplementary Table S3**). This

TABLE 3 | The results of the data analysis of four marker types.

	ITS2	<i>psbA-trnH</i>	<i>matK</i>	<i>rbcl</i>
Number of unique contig	378	6,416	44	46
Number of DNA barcodes after annotation chimera	136	26	16	21
Number of OTU	80	11	9	8
Average length (bp)	207.5	321.1	836.3	703
GC content (%)	54.4	32.2	34.7	44.1

indicated that the concentration and purity of the DNA extracted from the traditional herbal patent medicine samples were high. A total of 37.14 G of raw data was obtained via HTS, while 8.54 G and 9 G of raw data were acquired from the HSZY160 and HSZY172 lab-made mock samples, respectively. Additionally, 6.81 G, 6.78 G, and 6.01 G of raw sequencing data were acquired from the WHS001, WHS002, and WHS003 commercial samples. A total of 123,799,141 paired-end sequencing reads were obtained. After removing low-quality sequences, a total of 1,421,013 paired-end sequencing reads were enriched for the ITS2, *psbA-trnH*, *matK*, and *rbcl* regions. The detailed sequencing results are shown in **Supplementary Table S4**. A total of 6,884 unique contigs were generated by assembling and then removing duplications using MEGAHIT v1.2.9 and MetaSPAdes v3.13.2. The DNA barcoding regions of ITS2, *psbA-trnH*, *matK*, and *rbcl* yielded 136, 26, 16, and 21 unique contigs, respectively, after annotating and removing the primers. The cluster analysis of the ITS2 region yielded a total of 80 OTUs, with an average length of 207.5 bp and an average GC content of 54.4%. The number of OTUs obtained via nuclear ITS2 was more than seven times that of chloroplast *psbA-trnH*, *matK*, and *rbcl*. Moreover, the GC content of the ITS2 sequences was higher than that of the *psbA-trnH*, *matK*, and *rbcl* sequences. The specific data results of the four markers are shown in **Table 3**.

The Accuracy Verification of the DNA Barcoding Sequences Assembled Using the Shotgun Sequencing Data of the Lab-Made Samples

The DNA barcode assembly results of the labeled ingredients in the prescription of the lab-made mock samples are shown in **Table 4**. To determine the assembly accuracy of the DNA barcode

TABLE 4 | The DNA barcode sequences of five ingredients in the prescription of the lab-made Wuhu San samples, and the positive control, *Panax quinquefolius*, obtained via shotgun metabarcoding.

Species	HSZY160				HSZY172			
	ITS2	<i>psbA-trnH</i>	<i>matK</i>	<i>rbcl</i>	ITS2	<i>psbA-trnH</i>	<i>matK</i>	<i>rbcl</i>
<i>Angelica sinensis</i>	√	√	√	√	√	√	√	√
<i>Carthamus tinctorius</i>	√	√	√	√	√	√	√	√
<i>Saposhnikovia divaricata</i>	√	√	√	√	√	√	√	√
<i>Arisaema amurense</i>	√	—	√	√	√	—	√	√
<i>Angelica dahurica</i>	√	√	—	—	√	√	—	—
<i>Panax quinquefolius</i>					√	√	√	√

Note: "√" shows that the corresponding assembly sequence of this species was obtained; "—" indicates that the corresponding assembly sequence of this species cannot be obtained.

regions assembled via shotgun sequencing, the sequences obtained using shotgun metabarcoding and the reference sequences of ITS2, *psbA-trnH*, *matK*, *rbcl* obtained via Sanger sequencing were analyzed for consistency. The *psbA-trnH* sequence of *Arisaema amurense* and the *rbcl* sequence of *Angelica dahurica* were not obtained with Sanger sequencing.

Regarding the ITS2 sequences, the assembly sequences of all the ingredients in the prescriptions of the two mock samples were obtained. The sequence bases of *Angelica sinensis* and *Angelica dahurica* were identical to their reference sequences. A one base difference was evident between the sequences of *Arisaema amurense* obtained via two sequencing methods. Two *Saposhnikovia divaricata* assembly sequences were obtained, one of which was identical to the reference sequence bases, while the other differed by three bases. Two *Carthamus tinctorius* assembly sequences were obtained, which differed from the reference sequence by 0 and one base, respectively.

For the *psbA-trnH* sequences, the assembly sequences of *Angelica sinensis*, *Saposhnikovia divaricata*, *Carthamus tinctorius*, and *Angelica dahurica* were successfully obtained, but shotgun metabarcoding failed to acquire the *Arisaema amurense* sequence. The assembled sequences of *Angelica sinensis*, *Angelica dahurica*, and *Carthamus tinctorius* were identical to the reference sequences obtained via Sanger sequencing. Two *psbA-trnH* sequences of *Saposhnikovia divaricata* were acquired via shotgun sequencing. Compared with the sequences obtained via traditional DNA barcoding, one is identical, while the other displays two base differences.

Regarding the *matK* sequences, the assembly sequences were acquired for all the species in the two mock samples except for *Angelica dahurica*. A comparison between the *matK* assembly and reference sequences of the four species showed that the sequence bases of *Arisaema amurense* were the same. The two assembled sequences of *Carthamus tinctorius* and *Saposhnikovia divaricata* were obtained via shotgun sequencing and displayed a 0–3 base difference from the reference sequences. A total of two assembled *matK* sequences of *Angelica sinensis* were obtained, which differed from the reference sequences by five and seven bases, respectively.

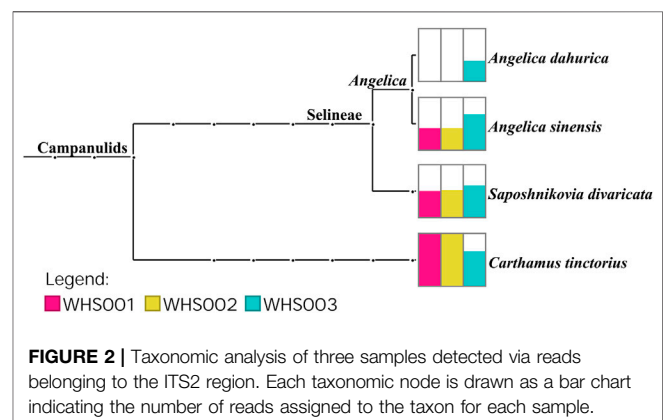
The assembly sequences of *Carthamus tinctorius* and *Arisaema amurense* were successfully acquired for the *rbcl* region. There were only five mutation sites among the *rbcl* sequences of *Angelica sinensis*, *Saposhnikovia divaricata*, and

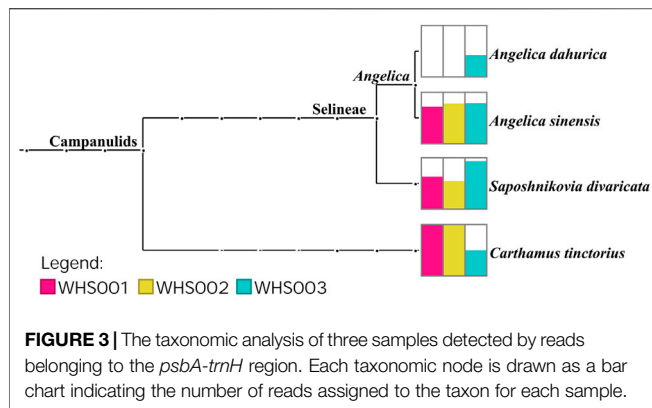
Angelica dahurica. Therefore, these sequences of the three species could not be assembled separately. Only the *rbcl* sequences of *Carthamus tinctorius* and *Arisaema amurense* were obtained using both shotgun and Sanger sequencing. The assembled sequences of the two species were completely consistent with the reference sequences.

In addition, for the positive control, *Panax quinquefolius*, all the ITS2, *psbA-trnH*, *matK*, and *rbcl* sequences that were acquired using the shotgun metabarcoding method were identical to those obtained via the traditional DNA barcode method.

The Plant Species Composition of Commercial Wuhu San Samples Identified Through Shotgun Metabarcoding

Regarding the labeled ingredients in the prescription, combined with the ITS2, *psbA-trnH*, *matK*, and *rbcl* regions, the three commercial samples contained prescription medicinal materials *Angelica sinensis*, *Carthamus tinctorius*, and *Saposhnikovia divaricata*. None of the samples contained *Arisaema erubescens*, *Arisaema heterophyllum*, or *Arisaema amurense*, representing the original *Arisaematis Rhizoma* (Tiannanxing) species, while only *Angelica dahurica* was detected in WHS003 (Figures 2, 3, Supplementary Figures S1, S2). For the ITS2 and *psbA-trnH* sequences, the results showed that *Angelica sinensis* and *Saposhnikovia divaricata* were detected in all three





commercial samples, while *Angelica dahurica* was only detected in WHS003 (Figures 2, 3). A total of six OTUs were obtained from the *matK* region. Of these, two OTUs were identified as *Carthamus tinctorius*, while the remaining four were only identified as belonging to the Apiaceae family, but the species could not be authenticated (Supplementary Figure S1). Six OTUs were obtained from the *rbcL* sequences in the three commercial samples, of which one was identified to species level, namely *Carthamus tinctorius*. The remaining five OTUs could only be identified as belonging to the Apiaceae family, but no species could be determined (Supplementary Figure S2). Detailed reads of the prescription ingredients in the three commercial samples based on four barcodes are shown in Supplementary Tables S5–S8.

As for the adulterants of the labeled ingredients, *Ferula bungeana* Kitag. was detected in two of the commercial samples (WHS001 and WHS002) based on the ITS2 sequences. In addition to these labeled ingredients and their adulterants, several other potential impurities were found in the commercial Wuhu San samples. Based on the ITS2 sequences, *Scutellaria baicalensis* Georgi was detected in two commercial samples (WHS001 and WHS002), while *Salix* L. was detected in WHS002 and WHS003. Impurities, such as *Convolvulus arvensis* L., *Chenopodium album* L., and *Citrus* L., were also found in WHS003.

The Fungal Contamination of the Lab-Made and Commercial Wuhu San Samples Detected via ITS2

A total of 36 fungal OTUs were obtained based on ITS2 sequences, including 22 families and 24 genera. The reads number of the *Rhizopus* genus was the highest of the 24 detected genera, accounting for 94.33% of the total number of fungal reads. It was the predominant genera in the two lab-made mock samples and three commercial samples. Fungi belonging to the *Macrophomina*, *Fusarium*, *Aspergillus*, and *Alternaria* genera were the dominant abundant in this study. Most of these fungi were molds that were present during the storage of herbal medicines, while some were soil habitat fungi.

Here, 8, 7, 13, 16, and 20 fungal genera were detected in HSZY160, HSZY172, WHS001, WHS002, and WHS003,

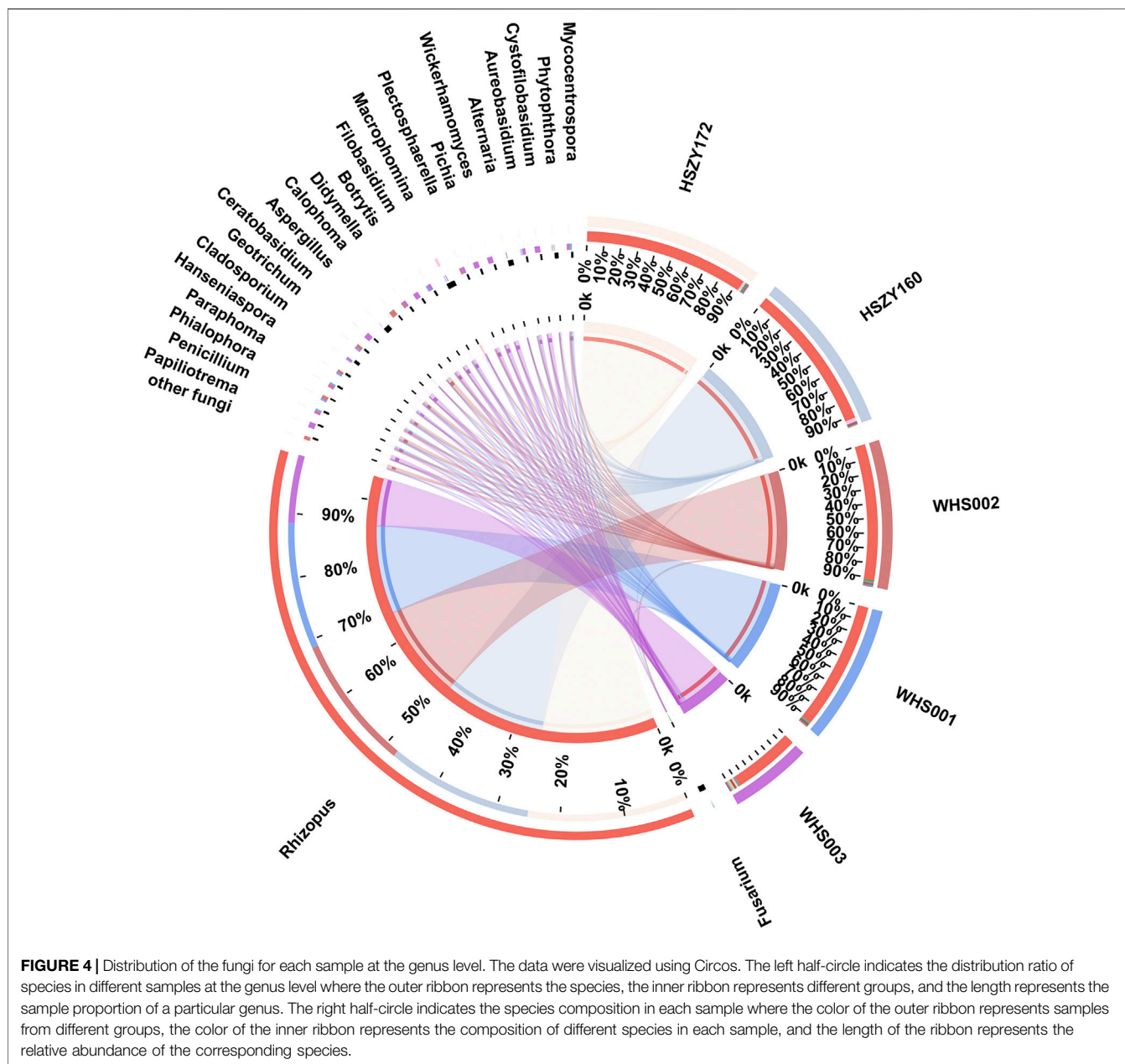
respectively. Fungi belonging to the *Fusarium*, *Rhizopus*, and *Alternaria* genera were detected in all five samples. In addition, the composition of the identified fungi in the two lab-made samples (HSZY160 and HSZY172) was similar at the genus level. The number of fungal species found in the commercial samples was significantly higher than in the mock samples, and some differences were evident between the fungal compositions of the three commercially available samples (Figure 4, Supplementary Tables S9). Besides the fungi belonging to genera found in all five samples, *Aspergillus*, *Penicillium*, *Geotrichum*, and *Mycocentrospora* were detected in all three commercial samples. WHS001 and WHS002 displayed a higher similarity in fungal composition than WHS003. Most of the fungal species were detected in WHS003, while *Aspergillus flavus*, which may produce aflatoxin that is harmful to human health, was detected in this sample (Supplementary Figure S3, Supplementary Tables S10).

DISCUSSION

The Feasibility of Shotgun Metabarcoding Technology in Authenticating the Herbal Ingredients of Wuhu San

DNA metabarcoding is currently the most widely used detection method for mixed biological samples. Many studies are available that involve the identification of biological ingredients of traditional herbal medicine based on DNA metabarcoding technology (Jia et al., 2017; Xin et al., 2018b; Shi et al., 2018; Zhang et al., 2020). However, the PCR amplification efficiency of universal DNA barcode primers is affected by the severe DNA degradation of traditional herbal medicine (Xin et al., 2018a). Furthermore, it may result in potential bias during PCR amplification using primers (Berry et al., 2011). Shotgun metabarcoding directly performs library construction and sequencing of the total DNA of mixed samples (Quince et al., 2017) and can obtain ITS2 sequences and multiple chloroplast DNA barcode sequences through the assembly for species identification (Xin et al., 2018a). This method has also been applied for studying clinical or complex environmental samples (Nielsen et al., 2014; Vangay et al., 2018). The shotgun metabarcoding method can reduce or eliminate the potential bias caused by PCR amplification and obtain a longer DNA barcode sequence interval than DNA metabarcoding. Several analytical biodiversity studies based on shotgun sequencing technology have shown that this technique can be used for biodiversity assessment. This method avoids the PCR amplification of particular gene markers to display species richness with high fidelity, while there is a significant correlation between the reads and biomass of most species (Zhou et al., 2013; Bista et al., 2018). However, this technology is more expensive, while the DNA quality requirements are also higher. Furthermore, the related bioinformatics analysis also presents a significant challenge.

This study shows that four barcode sequences can be successfully obtained in most of the medicinal materials



through shotgun metabarcoding technology in the lab-made mock samples. The mutual verification between the results obtained via different markers further confirmed the accuracy of the method. The results revealed that the *psbA-trnH* sequence of plants belonging to Araceae presented a low success rate via Sanger sequencing due to the adenine (A) and thymine (T) base content of over 70%, while their *psbA-trnH* sequences were difficult to obtain (Luo et al., 2009). It is speculated that the failure to obtain the *psbA-trnH* sequences of *Arisaema amurense* via shotgun metabarcoding technology is due to sequencing reads not being enriched enough for assembly or assembly failure caused by continuous AT repetition. Furthermore, the *Panax quinquefolius* barcode sequences were obtained in the HSZY172 sample, proving the sensitivity of the method for detecting

prescription ingredient species. *Arisaematis Rhizoma* (Tiannanxing) was not detected in the three commercially available samples through any of the markers. It is speculated that this is due to the difference in the degree of processing of *Arisaematis Rhizoma Preparatum* (Zhitiannanxing) when preparing the lab-made samples and commercial products. The DNA degradation of *Arisaematis Rhizoma Preparatum* (Zhitiannanxing) in the commercial samples may be more severe. *Ferula bungeana*, a plant belonging to the Apiaceae, was detected in two of the commercial samples. Studies have shown that the dried roots of the *Ferula bungeana* are used as *Saposhnikovia Radix* (Fangfeng) in the market (Yang, 2006; Chen and Chen, 2010). However, the efficacy of the two medicinal materials is quite different, and the identification of

the herbal materials should be enhanced to ensure the efficacy of traditional herbal patent medicine. Moreover, herbal impurities, such as *Salix* sp., *Chenopodium album*, *Convolvulus arvensis*, *Citrus* sp. and *Scutellaria baicalensis* were also detected. Their presence may be due to weeds that have been mixed in during harvesting (Geng et al., 2018) or accidental cross-contaminants from the same production line (Xin et al., 2018b). Furthermore, fungi were found in all the samples. During the planting, harvesting, transportation, and storage of herbal medicines, improper methods may cause fungal growth or the accumulation of mycotoxins (Ying et al., 2016), directly affecting the quality, efficacy, and safety of herbal medicines. This necessitates the examination of optimal storage conditions or preparation methods of herbal medicines contaminated by fungi (Wang, 2016). In summary, this indicated that shotgun metabarcoding could not only detect the adulteration of herbal materials in traditional herbal patent medicine, but it can also detect exogenous contamination, such as fungi and impurities. This proves the feasibility of shotgun metabarcoding for detecting biological ingredients in the traditional herbal patent medicine, Wuhu San.

The Challenges of the Current Shotgun Metabarcoding Method During Data Analysis

False Positives Caused by Reads Mapping in Conserved Regions

In this study, the ITS2 sequence of *Peucedanum japonicum* Thunb. was assembled from the lab-made mock samples. The coverage of the reads mapping was 100%, and the sequencing depth was 1,445.61. However, *Peucedanum japonicum* was not added to the lab-made mock samples. Visual reads mapping based on Codoncode Aligner indicated that the tail of the ITS2 sequence had a mapping depth exceeding 2000 ×, but only two mapping reads were present at the front end. Although the coverage of the fragments was uneven, the exceptionally high coverage of the tail significantly increased the overall coverage of the sequence, leading to the occurrence of false-positive sequences. Intercepting the tail segment, the NCBI BLAST analysis indicated that it was a 28S conserved sequence (Supplementary Figure S4). Further investigation revealed that the ITS2 sequence tail assembly was not accurate, preventing the ITS2 sequence annotation process from recognizing the 28S section, partially cutting it off. The 28S region is exceptionally conservative. Bowtie2 software randomly maps the reads to the reference genome with the same sequences during the mapping process (Langmead and Salzberg, 2012), resulting in an extremely high mapping depth for the conservative 28S region and a high average sequencing depth for the ITS2 sequences. This problem highlights the necessity to perform sequence annotation and primer removal accurately. Furthermore, CodonCode Aligner software can also verify the annotation results to reduce the occurrence of false-positive sequences.

The Accuracy of High Similarity Sequence Assembly

The assembly of high similarity sequences or low variability sequences is a challenge during shotgun metabarcoding data analysis (Quince et al., 2017). The optimized metagenomic data assembly software and more extensive k-mer parameters may overcome the assembly errors of lower similarity sequences to some extent. However, some difficulties remain when assembling the *matK* and *rbcl* sequences of some species in the same family, especially in the same genus. In this study, the ITS2 and *psbA-trnH* sequences of *Angelica dahurica* were obtained from the lab-made mock samples and commercial samples, but the *matK* and *rbcl* sequences of *Angelica dahurica* were not assembled. The prescriptions of Wuhu San contain *Angelica dahurica*, *Angelica sinensis*, and *Saposhnikovia divaricata* of the same family. The *matK* and *rbcl* sequences of the three species exhibited a similarity of more than 98%. It is speculated that the assembly of the three species may be incorrect due to insufficient assembly accuracy. The *matK* sequences of *Angelica dahurica*, *Angelica sinensis*, and *Saposhnikovia divaricata* obtained via Sanger sequencing were compared. There were six base differences between *Angelica dahurica* and *Saposhnikovia divaricata*, and 12 base differences between *Angelica dahurica* and *Angelica sinensis*. Furthermore, 14 base differences were evident between the *matK* sequences of *Saposhnikovia divaricata* and *Angelica sinensis* (Supplementary Figure S5). The *matK* sequence of *Angelica dahurica* displayed a higher similarity to that of *Saposhnikovia divaricata*, and differences were apparent in the bases at sites 42, 376, 415, 588, 722, and 758, indicating an average of 139 bp in a variant site. Analysis of the visual reads mapping results based on Codoncode Aligner showed that the base sites mentioned above have specific bases representing *Angelica dahurica* and *Saposhnikovia divaricata*, respectively (Supplementary Figure S6). Therefore, it is inferred that the assembly has not reached a high level of accuracy due to the small difference in sequence bases. The *matK* sequences of two species may be assembled into one sequence, representing the species sequence with more extensive sequencing depth. This is the same in the case of the *rbcl* sequence. Analysis performed via Sanger sequencing revealed that the bases of the *rbcl* sequences of *Angelica dahurica* and *Saposhnikovia divaricata* were T and A at the 270 base site, T and C at the 130 base site, and T and G at the 635 base site, respectively. The bases of the *rbcl* sequences of *Angelica dahurica* and *Angelica sinensis* at base sites 237, 270, and 475 were G/A, T/A, and C/T, respectively (Supplementary Figure S7). The base differences between the *rbcl* sequences of the three species were smaller. There were only five mutation sites among the *rbcl* sequences of *Angelica dahurica*, *Angelica sinensis*, and *Saposhnikovia divaricata*, that is, one variant site appeared on average 141 bp. The *rbcl* sequence mapping results of *Angelica dahurica* showed that different bases represented these three species at the various base positions (Supplementary Figure S8). The average length of a variation site in *matK* and *rbcl* sequences exceeds the length of commonly used k-mer (Quince et al., 2017). Using a more extended k-mer parameter may solve the problem of species distinction when the sequence

similarity exceeds 98%. However, the k-mer length may exceed the standard analysis length, requiring a massively distributed metagenome assembler, such as Ray, for *de novo* assembly to solve the computational time and memory challenge (Boisvert et al., 2012).

The Identity Threshold of DNA Barcodes for Constructing OTUs

This study initially conducted OTU sequence clustering according to 99% similarity to improve the efficiency and accuracy of the analysis. The results showed that *Angelica sinensis* was detected based on the ITS2, *psbA-trnH*, and *matK* regions, while the *Angelica sinensis* sequence was not detected based on the *rbcL* region. However, the *rbcL* sequence of *Saposhnikovia divaricata*, which belongs to the same family as *Angelica sinensis*, was detected. The Codoncode Aligner was used to further analyze the *rbcL* sequences of *Angelica sinensis* and *Saposhnikovia divaricata* obtained via Sanger sequencing. The results revealed that the *rbcL* sequences of the two species only displayed a 4-base difference. When the similarity was set to 99% for OTU clustering, the *rbcL* sequences of *Angelica sinensis* and *Saposhnikovia divaricata* was artificially divided into an OTU cluster. Therefore, the assembled *rbcL* sequences obtained via shotgun sequencing and the *rbcL* sequences of *Angelica sinensis* and *Saposhnikovia divaricata* acquired with Sanger sequencing were re-analyzed by building phylogenetic trees. Two sequences, namely "WHS001_rbcL_0189_k141_7" and "WHS002_rbcL_0047_k141_8" were found and grouped with the *rbcL* sequences of *Angelica sinensis* (HSYC2002 and HSYC2022) (**Supplementary Figure S9**). Therefore, the *Angelica sinensis* sequence could be detected by the *rbcL* region. This study further revealed that the *Angelica dahurica*, *Angelica sinensis*, and *Saposhnikovia divaricata* sequences were similar, especially those of *matK* and *rbcL*. Therefore, different levels of similarity should be set for OTU clustering when using DNA barcodes with varying species resolutions. For homologous species, the similarity should be further adjusted to 100% to avoid the undetectable phenomenon of sequences with similarities that are too high, which is also consistent with the current analysis strategy recommended by USEARCH.

The Species-Discriminating Power of the ITS2, *psbA-trnH*, *matK*, and *rbcL* DNA Barcodes

The four DNA barcodes displayed differences in the species discriminating power for the ingredients of Wuhu San prescriptions. All the ITS2 sequences obtained in this study can accurately identify species after BLAST. However, the *psbA-trnH* sequences of *Panax quinquefolius* and *Panax ginseng* did not have a variable site and could not be accurately distinguished. The regions of *rbcL* and *matK* exhibited certain limitations in identifying Apiaceae species in this study. Another study indicated that the efficiency of *rbcL* and *matK* sequences in identifying Apiaceae species was much lower than that of ITS2 sequences (Liu et al., 2014). Of the labeled ingredients in the Wuhu San prescription, *Angelica dahurica*, *Angelica sinensis*, and *Saposhnikovia divaricata* were all Apiaceae

plants. In addition to a small base difference and insufficient assembly accuracy, the low discriminating power of the *rbcL* and *matK* sequences for Apiaceae may also be one reason for the failure to detect *Angelica dahurica* in Wuhu San samples based on these two regions.

The analytical results of this study indicated that the ITS2 sequences displayed the strongest species discriminating power, while that of *psbA-trnH* sequences was lower than the ITS2 sequences. The *matK* and *rbcL* sequences demonstrated the worst species discrimination. Although the discriminating efficiency of the chloroplast *psbA-trnH*, *matK*, and *rbcL* sequences in this study was lower than that of the nuclear ITS2 sequences, the sequences of the chloroplast genome also exhibited certain advantages. The chloroplast genome is mostly maternally inherited and represents single-copy sequences in plant cells (Chen, 2016; Daniell et al., 2016). Moreover, chloroplast DNA sequences can be used as a versatile tool for plant identification (Nock et al., 2011). Most plants contain a significant number of chloroplasts, making DNA easy to obtain. In addition, this study showed that the species obtained from the chloroplast sequences were relatively simple. The obtained sequences generally represented prescription ingredients or obvious adulterants. A combination of multiple DNA barcodes can improve the resolution and accuracy of species discrimination (Group et al., 2009; Arulandhu et al., 2017). In addition, the ITS2 sequences can also detect fungi, which can be used to monitor the potential risk of the fungal contamination of traditional herbal patent medicine (Sweeney and Dobson, 1998; Guo et al., 2020). Therefore, to take advantage of shotgun metabarcoding, combining multiple barcodes obtained via the technology can increase the reliability and applicability of the experimental results. It helps monitor the quality of traditional herbal patent medicine.

DATA AVAILABILITY STATEMENT

The high-throughput sequencing datasets presented in this study can be found in the National Center for Biotechnology Information (NCBI) SRA online repository. The accession number of the BioProject is PRJNA663116. The accession numbers of the BioSample specimens are SAMN16124456, SAMN16124457, SAMN16124458, SAMN16124459, and SAMN16124460. And the SRA accession numbers for the above five BioSample specimens are SRR12632599, SRR12632598, SRR12632597, SRR12632596, and SRR12632595, respectively.

The DNA barcoding sequences assembled from the Sanger sequencing datasets presented in this study can be found in the NCBI GenBank online repository. The accession numbers for these DNA barcoding sequences are MN727081, MN727076, MT821449, MT821451, MT821450, MT102865, MT994327, MT994328, MT994331, MT994330, MT994329, MN729559, MN729561, MW000338, MW000340, MW000339, MW000341, MN746764, MN746766, MW000334, MW000335, and MW000333.

AUTHOR CONTRIBUTIONS

LS and JL conceived and designed the study. WM, QZ, and HX performed the experiment. WM, LS, and MS analysed the data. WM, LS, and JL wrote the paper. LS and JL revised the paper. All authors read and approved the final manuscript.

FUNDING

This work was supported by The National Key Research and Development Program of China (2019YFC1604705), Beijing

REFERENCES

- Arulandhu, A. J., Staats, M., Hagelaar, R., Voorhuijzen, M. M., Prins, T. W., Scholtens, I., et al. (2017). Development and validation of a multi-locus DNA metabarcoding method to identify endangered species in complex samples. *GigaScience*. 6 (10), 1–18. doi:10.1093/gigascience/gix080
- Arkan, M., Mitchell, A. L., Finn, R. D., and Gürel, F. (2020). Microbial composition of Kombucha determined using amplicon sequencing and shotgun metagenomics. *J. Food Sci.* 85 (2), 455–464. doi:10.1111/1750-3841.14992
- Barnes, J., McLachlan, A. J., Sherwin, C. M., and Enioutina, E. Y. (2016). Herbal medicines: challenges in the modern world. Part 1. Australia and New Zealand. *Expet Rev. Clin. Pharmacol.* 9 (7), 905–915. doi:10.1586/17512433.2016.1171712
- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K. D., et al. (2018). *GenBank*. Nucleic Acids Research. 46 (D1), D41–d47. doi:10.1093/nar/gkx1094
- Berry, D., Ben Mahfoudh, K., Wagner, M., and Loy, A. (2011). Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Appl. Environ. Microbiol.* 77 (21), 7846–7849. doi:10.1128/aem.05220-11
- Bista, I., Carvalho, G. R., Tang, M., Walsh, K., Zhou, X., Hajibabaei, M., et al. (2018). Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples. *Molecular Ecology Resources*. 18 (5), 1020–1034. doi:10.1111/1755-0998.12888
- Boisvert, S., Raymond, F., Godzaridis, E., Laviolette, F., and Corbeil, J. (2012). Ray Meta: scalable de novo metagenome assembly and profiling. *Genome Biol.* 13 (12), R122. doi:10.1186/gb-2012-13-12-r122
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15), 2114–2120. doi:10.1093/bioinformatics/btu170
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinf.* 10, 421. doi:10.1186/1471-2105-10-421
- Chen, H. L., Sun, Y. J., Zhao, A. M., and Yang, W. C. (1998). On the microscopic identification of Chinese patent medicine and its development direction (in Chinese). *Liaoning Journal of Traditional Chinese Medicine*. 25 (12), 579–580
- Chen, R. S., and Chen, X. Y. (2010). Identification of saposchnikovia divaricata (in Chinese). *Capital Medicine*. 17 (17), 36
- Chen, S., Pang, X., Song, J., Shi, L., Yao, H., Han, J., et al. (2014). A renaissance in herbal medicine identification: from morphology to DNA. *Biotechnol. Adv.* 32 (7), 1237–1244. doi:10.1016/j.biotechadv.2014.07.004
- Chen, S. L. (2015). *Standard DNA barcodes of Chinese materia medica in Chinese Pharmacopoeia*. Beijing: Science Press
- Chen, X. C. (2016). *Barcoding Chinese herbal medicines: from Gene to genome*. Doctor. Beijing: Peking Union Medical College
- Cheng, C. S., Tan, T. Q., Long, Z., Liu, Z. Z., Wu, W. R., Wang, Y. P., et al. (2015). Optimization of DNA extraction for Chinese patent medicine and its application on molecular identification of ginseng preparations by MAS-PCR. *Chin. Tradit. Herb. Drugs* 46 (17), 2549–2555. 10.7501/j.issn.0253-2670.2015.17.008
- Municipal Natural Science Foundation (7202136), National Natural Science Foundation of China (81703659), Technology Innovation Guidance Project-Science and Technology Work Conference of Hebei Provincial Department of Science and Technology, and China Postdoctoral Science Foundation (2017M610815).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2021.607200/full#supplementary-material>.

- Chunna, L. I., Liu, Y., Pengshou, L. I., Shi, X., Tunhai, X. U., and Liu, T. (2015). Chemical constituents and pharmacological activities of arisaema amurense maxim. *Jilin Journal of Traditional Chinese Medicine*. 35 (3), 293–296
- Committee, S. P. (2020). *Pharmacopoeia of the people's Republic of China Part I*. Beijing: China Medical Science Press
- Daniell, H., Lin, C. S., Yu, M., and Chang, W. J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17 (1), 134. doi:10.1186/s13059-016-1004-2
- Deng, H. Z., Chen, Y. Y., Liu, H. H., Chen, F. L., and Huang, X. L. (2000). Pharmacodynamic studies of Wuhu koufuye. *J. Fourth Mil. Med. Univ.* 20 (3), 269–271
- Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27 (16), 2194–2200. doi:10.1093/bioinformatics/btr381
- Etherington, G. J., Ramirez-Gonzalez, R. H., and MacLean, D. (2015). bio-stools 2: a package for analysis and visualization of sequence and alignment data with SAMtools in Ruby. *Bioinformatics* 31 (15), 2565–2567. doi:10.1093/bioinformatics/btv178
- Ferrocino, I., Bellio, A., Giordano, M., Macori, G., Romano, A., Rantsiou, K., et al. (2018). Shotgun metagenomics and volatilome profile of the microbiota of fermented sausages. *Appl. Environ. Microbiol.* 84 (3), 02117–e02120. doi:10.1128/aem.02120-17
- Geng, Y. L., Yuan, L. B., Zhang, P., Zhang, C. J., and Hun, Z. Y. (2018). Characterization of weed communities in medicinal plant fields in Anguo, Hebei. *J. Chin. Med. Mater.* 41 (05), 1048–1053
- Group, C. P. W., Hollingsworth, P. M., Forrest, L. L., Spouge, J. L., Hajibabaei, M., Ratnasingham, S., et al. (2009). A DNA barcode for land plants. *Proc. Natl. Acad. Sci. U.S.A.* 106 (31), 12794–12797. doi:10.1073/pnas.0905845106
- Guo, M., Jiang, W., Yang, M., Dou, X., and Pang, X. (2020). Characterizing fungal communities in medicinal and edible Cassiae Semen using high-throughput sequencing. *Int. J. Food Microbiol.* 319, 108496. doi:10.1016/j.ijfoodmicro.2019.108496
- Haiminen, N., Edlund, S., Chambliss, D., Kunitomi, M., Weimer, B. C., Ganesan, B., et al. (2019). Food authentication from shotgun sequencing reads with an application on high protein powders. *NPJ Sci Food*. 3, 24. doi:10.1038/s41538-019-0056-6
- Hasan, N. A., Young, B. A., Minard-Smith, A. T., Saeed, K., Li, H., Heizer, E. M., et al. (2014). Microbial community profiling of human saliva using shotgun metagenomic sequencing. *PLoS One* 9 (5), e97699. doi:10.1371/journal.pone.0097699
- Huson, D. H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., et al. (2016). MEGAN community edition - interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.* 12 (6), e1004957. doi:10.1371/journal.pcbi.1004957
- Jia, J., Xu, Z., Xin, T., Shi, L., and Song, J. (2017). Quality control of the traditional patent medicine Yimu wan based on SMRT sequencing and DNA barcoding. *Front. Plant Sci.* 8, 926. doi:10.3389/fpls.2017.00926
- Jia, Y., Hong, L., Qi, H., Peng, S., and Lu, F. (1997). Chinese herbal medicine research OF antithrombotic effect. *Natural Product Research and Development*. 9 (2), 17–20

- Job, K. M., Kiang, T. K., Constance, J. E., Sherwin, C. M., and Enioutina, E. Y. (2016). Herbal medicines: challenges in the modern world. Part 4. Canada and United States. *Expet Rev. Clin. Pharmacol.* 9 (12), 1597–1609. doi:10.1080/17512433.2016.1238762
- Kang, O. H., Chae, H. S., Oh, Y. C., Choi, J. G., Lee, Y. S., Jang, H. J., et al. (2008). Anti-nociceptive and anti-inflammatory effects of *Angelicae dahuricae radix* through inhibition of the expression of inducible nitric oxide synthase and NO production. *Am. J. Chin. Med.* 36 (5), 913–928. doi:10.1142/s0192415x0800634x
- Kechin, A., Boyarskikh, U., Kel, A., and Filipenko, M. (2017). cutPrimers: a new tool for accurate cutting of primers from reads of targeted next generation sequencing. *J. Comput. Biol.* 24 (11), 1138–1143. doi:10.1089/cmb.2017.0096
- Keller, A., Schleicher, T., Schultz, J., Müller, T., Dandekar, T., and Wolf, M. (2009). 5.8S-28S rRNA interaction and HMM-based ITS2 annotation. *Gene*. 430 (1–2), 50–57. doi:10.1016/j.gene.2008.10.012
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9 (4), 357–359. doi:10.1038/nmeth.1923
- Li, D., Liu, C. M., Luo, R., Sadakane, K., and Lam, T. W. (2015a). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31 (10), 1674–1676. doi:10.1093/bioinformatics/btv033
- Li, H. X., Han, S. Y., Wang, X. W., Ma, X., Zhang, K., Wang, L., et al. (2009). Effect of the carthamins yellow from *Carthamus tinctorius* L. on hemorrhological disorders of blood stasis in rats. *Food Chem. Toxicol.* 47 (8), 1797–1802. doi:10.1016/j.fct.2009.04.026
- Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22 (13), 1658–1659. doi:10.1093/bioinformatics/btl158
- Li, Z. H., Qian, Z. Z., and Cheng, Y. Y. (2015b). [The technological innovation strategy for quality control of Chinese medicine based on Big Data]. *Zhongguo Zhongyao Zazhi* 40 (17), 3374–3378
- Liu, J., Shi, L., Han, J., Li, G., Lu, H., Hou, J., et al. (2014). Identification of species in the angiosperm family Apiaceae using DNA barcodes. *Mol Ecol Resour.* 14 (6), 1231–1238. doi:10.1111/1755-0998.12262
- Liu, J., Shi, L., Song, J., Sun, W., Han, J., Liu, X., et al. (2017). BOKP: a DNA barcode reference library for monitoring herbal drugs in the Korean Pharmacopoeia. *Front. Pharmacol.* 8, 931. doi:10.3389/fphar.2017.00931
- Luo, K. C. S., Chen, K. L., Song, J. Y., and Yao, H. (2009). Application of DNA barcoding to the medicinal plants of the Araneae family. *Planta Med.* 75 (04), 416
- Nielsen, H. B., Almeida, M., Juncker, A. S., Rasmussen, S., Li, J., Sunagawa, S., et al. (2014). Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat. Biotechnol.* 32 (8), 822–828. doi:10.1038/nbt.2939
- Nock, C. J., Waters, D. L., Edwards, M. A., Bowen, S. G., Rice, N., Cordeiro, G. M., et al. (2011). Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnol. J.* 9 (3), 328–333. doi:10.1111/j.1467-7652.2010.00558.x
- Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 27 (5), 824–834. doi:10.1101/gr.213959.116
- Oh, J., Byrd, A. L., Deming, C., Conlan, S., Program, N. C. S., Kong, H. H., et al. (2014). Biogeography and individuality shape function in the human skin metagenome. *Nature*. 514 (7520), 59–64. doi:10.1038/nature13786
- Okuyama, E., Hasegawa, T., Matsushita, T., Fujimoto, H., Ishibashi, M., and Yamazaki, M. (2001). Analgesic components of *saposhnikovia* root (*Saposhnikovia divaricata*). *Chem Pharm Bull.* 49 (2), 154–160. doi:10.1248/cpb.49.154
- Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., and Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* 35 (9), 833–844. doi:10.1038/nbt.3935
- Ratnasingham, S., and Hebert, P. D. (2007). Bold: the barcode of life data system (<http://www.barcodinglife.org>). *Mol. Ecol. Notes*. 7 (3), 355–364. doi:10.1111/j.1471-8286.2007.01678.x
- Sammons, H. M., Gubarev, M. I., Krepkova, L. V., Bortnikova, V. V., Corrick, F., Job, K. M., et al. (2016). Herbal medicines: challenges in the modern world. Part 2. European Union and Russia. *Expet Rev. Clin. Pharmacol.* 9 (8), 1117–1127. doi:10.1080/17512433.2016.1189326
- Shi, L., Chen, H., Jiang, M., Wang, L., Wu, X., Huang, L., et al. (2019). CPGAVAS2, an integrated plasmid sequence annotator and analyzer. *Nucleic Acids Res.* 47 (W1), W65–w73. doi:10.1093/nar/gkz345
- Shi, L. C., Liu, J. X., Wei, M. J., Xie, L. F., and Song, J. Y. (2018). DNA metabarcoding identification of prescription ingredients in traditional medicine *Ruyi Jinhuang San*. *Scientia Sinica Vitae*. 48 (04), 490–497
- Srivathsan, A., Sha, J. C., Vogler, A. P., and Meier, R. (2015). Comparing the effectiveness of metagenomics and metabarcoding for diet analysis of a leaf-feeding monkey (*Pygathrix nemaeus*). *Mol Ecol Resour.* 15 (2), 250–261. doi:10.1111/1755-0998.12302
- Suo, F., and Chen, S. (2006). Study on substitute material of endangered Chinese traditional medicine. *Asia Pacific Journal of traditional Chinese Medicine*. (04), 68–72
- Sweeney, M. J., and Dobson, A. D. (1998). Mycotoxin production by *Aspergillus*, *Fusarium* and *Penicillium* species. *Int. J. Food Microbiol.* 43 (3), 141–158. doi:10.1016/s0168-1605(98)00112-3
- Tang, G., and Huang, Z. (1994). Discussion on ten common concepts such as authentic and adulterants of traditional Chinese medicine (in Chinese). *J. Chin. Med. Mater.* 17 (03), 47–48
- Teng, L., Zu, Q., Li, G., Yu, T., Job, K. M., Yang, X., et al. (2016). Herbal medicines: challenges in the modern world. Part 3. China and Japan. *Expet Rev. Clin. Pharmacol.* 9 (9), 1225–1233. doi:10.1080/17512433.2016.1195263
- Tringe, S. G., von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., et al. (2005). Comparative metagenomics of microbial communities. *Science*. 308 (5721), 554–557. doi:10.1126/science.1107851
- Vangay, P., Johnson, A. J., Ward, T. L., Al-Ghalith, G. A., Shields-Cutler, R. R., Hillmann, B. M., et al. (2018). US immigration westernizes the human gut microbiome. *Cell* 175 (4), 962–e10. doi:10.1016/j.cell.2018.10.029
- Wang, S. (2016). *Research on storage specification of traditional Chinese medicines of being moldy with malt, lotus seeds and nutmeg as the models*. Master. Beijing: Peking Union Medical College
- Warden, J. G., Casaburi, G., Omelon, C. R., Bennett, P. C., Breecker, D. O., and Foster, J. S. (2016). Characterization of microbial mat microbiomes in the modern thrombolite ecosystem of lake Clifton, western Australia using shotgun metagenomics. *Front. Microbiol.* 7, 1064. doi:10.3389/fmicb.2016.01064
- Xin, T., Su, C., Lin, Y., Wang, S., Xu, Z., and Song, J. (2018a). Precise species detection of traditional Chinese patent medicine by shotgun metagenomic sequencing. *Phytomedicine* 47, 40–47. doi:10.1016/j.phymed.2018.04.048
- Xin, T., Xu, Z., Jia, J., Leon, C., Hu, S., Lin, Y., et al. (2018b). Biomonitoring for traditional herbal medicinal products using DNA metabarcoding and single molecule, real-time sequencing. *Acta Pharm. Sin. B* 8 (3), 488–497. doi:10.1016/j.apsb.2017.10.001
- Xu, F. Q., Zhang, X. R., Guo, F. W., Wen, A. D., and Tian, X. R. (2014). Research progress on the quality control of Chinese patent medicine. *Prog. Mod. Biomed.* 14 (31), 6159–6163
- Yang, C. Y. (2006). Identification of *Saposhnikovia divaricata* adulterants *ferula bungeana* kitag., *libam-otis laticalcina* shna et sheh and *carum carvi* L. (root) (in Chinese) *Heilongjiang J. Tradit. Chin. Med.* (5), 50
- Yang, X., Noyes, N. R., Doster, E., Martin, J. N., Linke, L. M., Magnuson, R. J., et al. (2016). Use of metagenomic shotgun sequencing technology to detect foodborne pathogens within the microbiome of the beef production chain. *Appl. Environ. Microbiol.* 82 (8), 2433–2443. doi:10.1128/AEM.00078-16
- Yang, X. L., Cui, H. S., and Li, M. (1990). Anti-inflammatory and analgesic effect of alcohol extract of Wuhu San (in Chinese). *Fujian Medical Journal*. 12 (4), 31–32
- Ying, G. Y., Zhao, X., Wang, J. L., Yang, M. H., Guo, W. Y., and Kong, W. J. (2016). [Application and prospect of "couplet medicine" techniques in preservation of Chinese medicinal materials]. *Zhongguo Zhongyao Zazhi* 41 (15), 2768–2773
- Yue, S. J., Xin, L. T., Fan, Y. C., Li, S. J., Tang, Y. P., Duan, J. A., et al. (2017). Herb pair Danggui-Honghua: mechanisms underlying blood stasis syndrome by system pharmacology approach. *Sci. Rep.* 7, 40318. doi:10.1038/srep40318
- Zhang, G., Liu, J., Gao, M., Kong, W., Zhao, Q., Shi, L., et al. (2020). Tracing the edible and medicinal plant *pueraria montana* and its products in the marketplace yields subspecies level distinction using DNA barcoding and DNA metabarcoding. *Front. Pharmacol.* 11, 336. doi:10.3389/fphar.2020.00336

- Zhang, L., Du, J. R., Wang, J., Wang, J., Yu, D. K., Chen, Y., et al. (2009). Z-ligustilide extracted from radix angelica synesis decreased platelet aggregation induced by ADP *ex vivo* and arterio-venous shunt thrombosis *in vivo* in rats. *Yakugaku Zasshi*. 129 (7), 855–859. doi:10.1248/yakushi.129.855
- Zhao, L., Zhang, F., Ding, X., Wu, G., Lam, Y. Y., Wang, X., et al. (2018). Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes. *Science*. 359 (6380), 1151–1156. doi:10.1126/science.aao5774
- Zhou, X., Li, Y., Liu, S., Yang, Q., Su, X., Zhou, L., et al. (2013). Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience* 2 (1), 4. doi:10.1186/2047-217x-2-4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Liu, Mu, Shi, Zhao, Kong, Xie and Shi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.