Biological Ingredient Analysis of Traditional Herbal Patent Medicine Fuke Desheng Wan Using the Shotgun Metabarcoding Approach

Hongbo Xie[1†], Qing Zhao[1,2†], Mengmeng Shi[1], Weijun Kong[3], Weishan Mu[1], Baoli Li[3], Jingyi Zhao[1], Chunying Zhao[1], Jing Jia[4], Jinxin Liu[1,3]* and Linchun Shi[3]*

[1]Hebei Key Laboratory of Study and Exploitation of Chinese Medicine, Chengde Medical University, Chengde, China, [2]Department of Pharmacy, Baoding First Central Hospital, Baoding, China, [3]Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China, [4]Department of Pharmacy, Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan, China

With the widespread use of traditional medicine around the world, the safety and efficacy of traditional herbal patent medicine have become an increasing concern to the public. However, it is difficult to supervise the authenticity of herbal materials in mixed herbal products according to the current quality standards, especially for traditional herbal patent medicine, with a distinct variance in the dosage of herbal materials. This study utilized the shotgun metabarcoding approach to analyze the biological ingredients of Fuke Desheng Wan (FKDSW), which is an effective traditional herbal product for the treatment of dysmenorrhea. Six herbal materials were collected, and a lab-made mock FKDSW sample was produced to establish a method for the authentication assessment of biological ingredients in traditional herbal patent medicine based on shotgun metabarcoding. Furthermore, four commercial FKDSW samples were collected to verify the practicality of the shotgun metabarcoding approach. Then, a total of 52.16 Gb raw data for 174 million paired-end reads was generated using the Illumina NovaSeq sequencing platform. Meanwhile, 228, 23, and 14 operational taxonomic units (OTUs) were obtained for the ITS2, *matK*, and *rbcL* regions, respectively, after bioinformatic analysis. Moreover, no differences were evident between the assembly sequences obtained *via* shotgun metabarcoding and their corresponding reference sequences of the same species obtained *via* Sanger sequencing, except for part of the ITS2 and *matK* assembly sequences of *Paeonia lactiflora* Pall., *Saussurea costus* (Falc.) Lipsch. and *Bupleurum chinense* DC. with 1–6 different bases. The identification results showed that all six prescribed ingredients were successfully detected and that the non-authentic ingredient of Bupleuri Radix (Chaihu, *Bupleurum chinense* DC. or *Bupleurum scorzonerifolium* Willd.) was found in all the commercial samples, namely *Bupleurum falcatum* L. Here, 25 weed species representing 16 genera of ten families were detected. Moreover, 26 fungal genera belonging to 17 families were found in both lab-made and commercial FKDSW samples. This study demonstrated that the shotgun metabarcoding approach could overcome the biased PCR amplification and authenticate the biological ingredients of traditional herbal patent medicine with a distinct variance in the dosage of the herbal materials. Therefore, this provides an appropriate evaluation method for improving the safety and efficacy of traditional herbal patent medicine.

Keywords: shotgun metabarcoding, traditional herbal patent medicine, biological ingredients, weeds, fungi
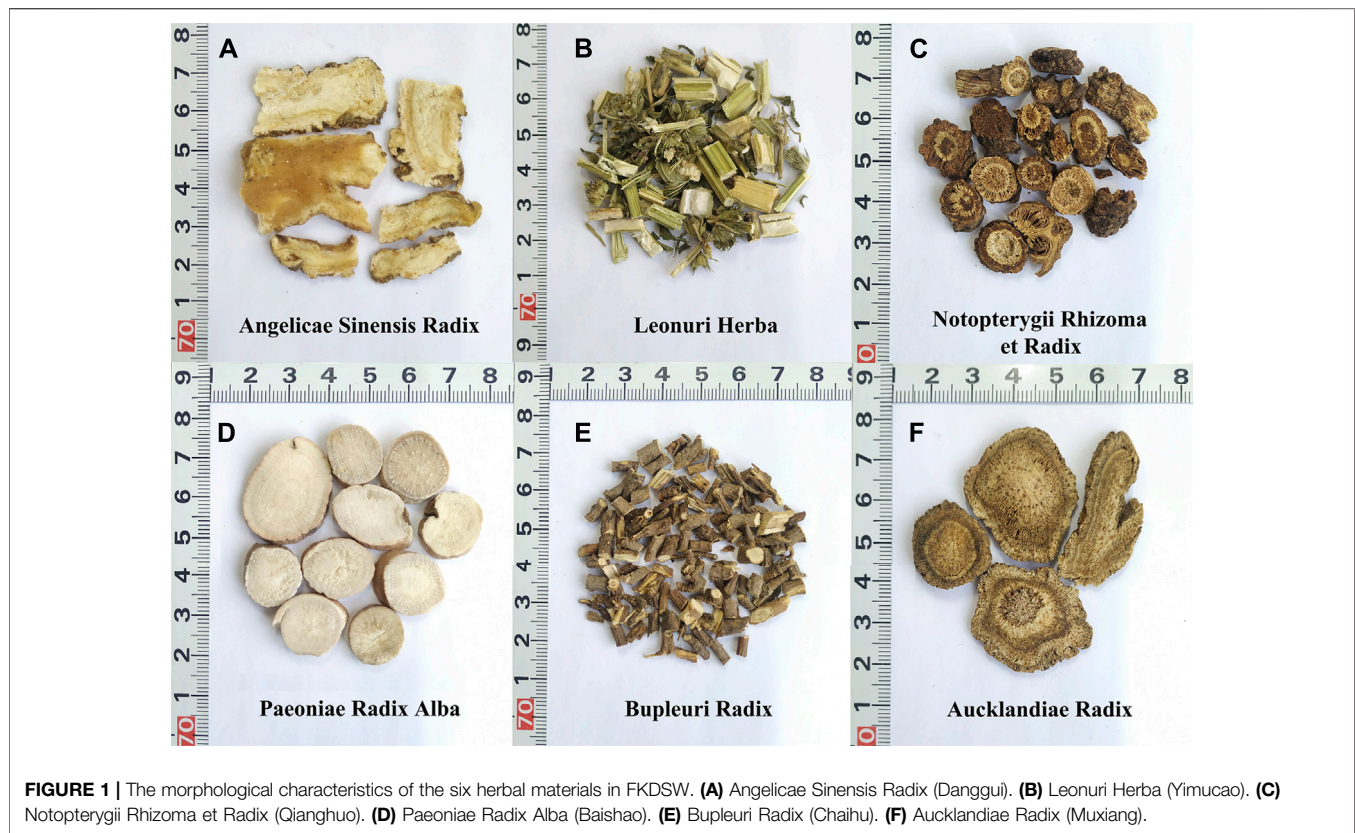
INTRODUCTION

Dysmenorrhea is a common gynecological disease that can harm the health, work status, and quality of life of women. Research was conducted in various countries to determine the number of adolescents and young women suffering from dysmenorrhea, the results of which indicated that the prevalence rate of this condition ranged from 34% (Egypt) to 94% (Oman) (De Sanctis et al., 2016). Although the prescribed first-line therapy for dysmenorrhea is non-steroidal anti-inflammatory drugs, which usually alleviate menstrual pain by inhibiting peripheral, systemic prostaglandins, and their corresponding downstream effects, there are still approximately 18% of women who continue to suffer from the condition who display a distinct resistance to this treatment (Iacovides et al., 2015; Oladosu et al., 2018). As a result, traditional herbal medicines have been proposed as alternative therapies for dysmenorrhea (Oladosu et al., 2018). Fuke Desheng Wan (FKDSW) has been recommended as a gynecological medicine for curing dysmenorrhea caused by liver depression or the stagnancy of both blood and qi (Commission, 1994). This treatment exhibits a unique curative effect in relieving dysmenorrhea, regulating abnormal menstruation, and improving complications (Commission, 1994). FKDSW is a honeyed pill consisting of six herbal materials, including Angelicae Sinensis Radix (Danggui), Paeoniae Radix Alba (Baishao), Aucklandiae Radix (Muxiang), Notopterygii Rhizoma et Radix (Qianghuo), Leonuri Herba (Yimucao), and Bupleuri Radix (Chaihu), at significantly different and very precise dosages. Leonuri Herba (Yimucao) is the primary ingredient in FKDSW at a level of 59.26%, whereas the content of Aucklandiae Radix (Muxiang), Notopterygii Rhizoma et Radix (Qianghuo), and Bupleuri Radix (Chaihu) displayed the lowest dosage level at 3.70%. The quality control and biological ingredient assessment of FKDSW are challenging compared to other traditional herbal patent medicines with similar dosage levels of each ingredient. In this case, the characteristics of medicinal materials with low content will be overwhelmed by the characteristics of medicinal materials with high content. Although the thin-layer chromatography identification method used for FKDSW, involving Angelicae Sinensis Radix (Danggui) and Paeoniae Radix Alba (Baishao), is recorded in the current standard (Commission, 1994), it is not enough to authenticate all of its labeled ingredients due to the current methods don't cover all the ingredients. For example, an authenticity survey revealed that approximately 35.3% of Bupleuri Radix (Chaihu) samples in herbal markets were identified to be adulterants in 85 samples of Bupleuri Radix (Chaihu) (Wang et al., 2017). Moreover, there are lots of researches proved that the widespread adulteration of commercial herbal products have been found throughout the global market (van der Valk et al., 2017; Ichim, 2019; Ichim et al., 2020). For example, Clematis armandii, an adulterant of Akebiae Caulis (Mutong, Akebia trifoliata), was detected in the commercial herbal product of Longdan Xiegan Wan (Xin et al., 2018a). And there are 7% of herbal products of the Lonicerae japonicae Flos contained both of two adulterants Eucommiae Folium and Lonicerae Flos (Gao et al., 2017).

Ichim (2019) and Ichim et al. (2020) found that herbal products containing undeclared contaminant, substitute, and filler species, or none of the labeled species were distributed across all continents and regions.

Compared to traditional pharmacopoeial identification methods including macroscopic, microscopic, chemical authentication approaches, DNA-based technology is more universal and accurate and can discriminate different species based on specific DNA fragments or even complete genomic information without the influence of environmental modification or the limitation of experiences, in spite of DNA-based methods do not provide any quantitative nor qualitative information of the active compounds in the herbal materials or the herbal preparations (Chen et al., 2012; Chen et al., 2018; Raclariu et al., 2018; Grazina et al., 2020). The protocol of DNA-based species identification methods like DNA metabarcoding mainly includes three steps: 1) total DNA extraction, 2) PCR amplification of the target DNA regions with universal primers, and 3) the identification and biodiversity assessment of multiple species in complex herbal preparations and products using high-throughput sequencing technology (Taberlet et al., 2012; Coghlan et al., 2015; Arulandhu et al., 2017; Gao et al., 2017; Song et al., 2017). However, the dependence of all these methods on PCR amplification limits their application in identifying the biological ingredients in traditional herbal patent medicine. For example, there are a large number of substances in traditional herbal materials or herbal products, such as polysaccharides, polyphenols, or excipients, which may inhibit PCR amplification and even lead to false-negative PCR results (Schrader et al., 2012). Furthermore, PCR bias, due to the differential binding of PCR primers to DNA templates, may lead to the loss of target sequences of some taxa or introduce chimeric sequences and other errors (Berry et al., 2011; Porter and Hajibabaei, 2018).

Shotgun metagenomics is capable of untargeted sequencing of all biological genomes from a single bulk sample without PCR amplification (Quince et al., 2017), which has been recognized as an unbiased method for investigating multiple species in various environmental or clinical samples (Bovo et al., 2018; Vijayvargiya et al., 2019; Ye et al., 2019). For example, Bovo et al. (2018) analyzed the environmental DNA samples of honey bees using shotgun metagenomic sequencing and reported five major biological groups in the two samples, including arthropods, plants, fungi, bacteria, and viruses, indicating that this method could be applied in large-scale experiments. Moreover, shotgun metagenomic sequencing holds additional benefits, such as species detection, the discovery of novel species, exploring the potential functions and relative abundance of the species of an organism, while reducing the reliance on independent cultures and PCR amplification approaches (Jovel et al., 2016; Ranjan et al., 2016; Porter and Hajibabaei, 2018; Ye et al., 2019). Although shotgun metagenomic sequencing has the potential of becoming a powerful analytical tool and has been applied to monitor the biological ingredients of traditional herbal patent medicine (Xin et al., 2018a), it presents two inherent difficulties that need to be overcome, namely the considerable amount of data that is generated and the limited number of whole genomic reference sequences. Since the purpose of the traditional herbal patent medicine identification is only to determine the species

FIGURE 1 | The morphological characteristics of the six herbal materials in FKDSW. **(A)** Angelicae Sinensis Radix (Danggui). **(B)** Leonuri Herba (Yimucao). **(C)** Notopterygii Rhizoma et Radix (Qianghuo). **(D)** Paeoniae Radix Alba (Baishao). **(E)** Bupleuri Radix (Chaihu). **(F)** Aucklandiae Radix (Muxiang).

composition of biological ingredients, selecting the phylogenetic or DNA barcoding markers may provide an alternative, fast, accurate, and lightweight approach (Segata et al., 2012; Sunagawa et al., 2013). With the development of high-throughput sequencing and bioinformatics, shotgun metabarcoding is expected to become a highly effective regulatory approach for the authentication assessment of biological ingredients in herbal product mixtures.

FKDSW is selected as the research object in this study. Shotgun metagenomic sequencing is used in conjunction with DNA barcoding (defined as "shotgun metabarcoding" in the present research) to determine the species composition in FKDSW and discuss the feasibility of this approach for monitoring the biological ingredients in traditional herbal patent medicine with distinct variance in the dosage of herbal materials. Meanwhile, this study combines multiple materials evidence to make its strategies and techniques for clearly and efficiently authenticating the herbal materials of FKDSW, effectively supplementing traditional identification methods, which will be useful for improving the holistic quality control of traditional herbal patent medicine.

MATERIALS AND METHODS

Sample Collection

Six herbal materials were purchased from the Beijing Tongrentang herbal store to prepare the mock FKDSW sample

and included Angelicae Sinensis Radix (Danggui, HSYC 2002), Paeoniae Radix Alba (Baishao, HSYC 2036), Aucklandiae Radix (Muxiang, HSYC 2037), Notopterygii Rhizoma et Radix (Qianghuo, HSYC 2038), Leonuri Herba (Yimucao, HSYC 2039), and Bupleuri Radix (Chaihu, HSYC 2078). Then, these materials were authenticated *via* morphological and DNA barcoding identification. In particular, the morphological identification was based on the herbal materials of form, shape and size, color, external markings and texture, organoleptic characters (odor, taste, and mouthfeel) recorded in the Chinese Pharmacopoeia (Commission, 2020). And these materials are deposited in the Institute of Medicinal Plant Development herbarium (herbarium code "IMD", NYBG: https://www.nybg.org/). The six medicinal materials and their prescribed proportions are shown in **Figure 1**, **Table 1** and **Supplementary Table S1**. To verify the feasibility and accuracy of the shotgun metabarcoding approach, the mock FKDSW sample was handcrafted in the laboratory according to the preparation method recorded in the pharmaceutical standard (Volume 9, WS3-B-1743-94) (Commission, 1994), and named HSZY167. Four batches of commercial FKDSW samples were purchased from drug stores to test the applicability of shotgun metabarcoding and numbered as HSZY003, HSZY140, HSZY153, and HSZY154.

DNA Extraction and Quantification

The DNA was extracted from the herbal materials following the previously published DNA barcoding protocol (Liu et al., 2017)

TABLE 1 | The prescribed proportion of each herbal ingredient that was labeled in FKDSW.

Herbal material Latin name (pinyin name)	Prescribed proportion (%)
Angelicae Sinensis Radix (Danggui)	14.81
Paeoniae Radix Alba (Baishao)	14.81
Aucklandiae Radix (Muxiang)	3.70
Notopterygii Rhizoma et Radix (Qianghuo)	3.70
Leonuri Herba (Yimucao)	59.26
Bupleuri Radix (Chaihu)	3.70

and principles for traditional Chinese herbal medicine (Chen et al., 2014). Detailed steps were carried out according to the protocol supplied by the manufacturer of the plant genomic DNA extraction kit [Tiangen Biochemical Technology (Beijing) Co., Ltd, China]. Then, the extracted DNA was amplified *via* PCR with primer sets of ITS2, *matK*, and *rbcL* regions using 2 × Taq master mix (AidLab Biotechnologies Co., Ltd., China). The primer selection and PCR conditions adhered to the barcode of traditional Chinese herbal medicine data system (TCM-BOL) (Chen et al., 2014), the consortium for the barcode of life (CBOL) plant working group (Group, 2009), and the barcode of life data system (BOLD) (Ratnasingham and Hebert, 2007). Then the PCR products were electrophoresed on 1.0% agarose gel using GelRed nucleic acid gel stain (Biotium Biotechnologies Co., Ltd., United States) to determine the integrity.

The metagenomic DNA was extracted from the traditional herbal patent medicine samples according to a method reported (Cheng et al., 2015), with some minor modifications during the pretreatment steps (Xin et al., 2018a). Finally, the concentration and quality of the DNA in all the samples of the herbal materials and traditional herbal patent medicine were determined using a NanoDrop ONE ultra-micro spectrophotometer (Thermo Fisher Scientific, Inc., United States).

Sequencing, Bioinformatic Analysis, and Species Identification

The PCR products of ITS2, *matK*, and *rbcL* were bi-directionally sequenced using an ABI 3730xL DNA Analyzer (Thermofisher Co., Ltd., United States). Then, the sequencing output files generated *via* Sanger sequencing were aligned and assembled using Codoncode aligner v 9.0.1 (CodonCode Corp., Dedham, MA, United States).

The Illumina NovaSeq platform was used for the shotgun sequencing of the traditional herbal patent medicine samples after a meta-genomic DNA paired-end library was constructed with sheared fragments. The sequencing output files generated by high-throughput sequencing were used to remove the adapters, and low quality reads using Trimmonmatic v0.38 (Bolger et al., 2014). The ITS2, *matK*, and *rbcL* sequences were downloaded from the National Center for Biotechnology Information (NCBI) to construct the reference database. And the reads after quality control were aligned to the reference database using basic local alignment search tool (BLAST), then the targeted paired-end reads of ITS2, *matK*, and *rbcL* were enriched by local python

scripts (Shi et al., 2019). The enriched reads were assembled with MEGAHIT v1.2.9 and MetaSPAdes v3.13.2 (Li et al., 2015; Nurk et al., 2017), while the k-mer values ranged from 31 to 127. The assembled contigs were clustered, and redundant data were removed using cd-hit with 100% identity (Li and Godzik, 2006). Cutadapt v2.10 was employed to remove the primer sequences of the conventional DNA barcoding *matK* and *rbcL* regions (Martin, 2011), while the ITS2 region was annotated using the hidden Markov model (HMM)-based annotation method (Keller et al., 2009). Chimeric sequences were removed by UCHIME v4.2 (Edgar et al., 2011). The remaining sequences were clustered into operational taxonomic units (OTUs) at 100% sequence similarity using Usearch v11 (https://www.drive5.com/usearch/). The OTUs of the representative sequences were further processed with Bowtie2 v2.4.1 and samtools v1.10, where the former was used to map the shotgun paired-end reads to the representative sequences, and the latter was employed to calculate the sequencing depth and coverage values (Langmead and Salzberg, 2012; Etherington et al., 2015). Then, the OTUs were manually removed, satisfying the following parameters: sequencing depth ≤3 and/or coverage ≤95%. The BLAST search (Camacho et al., 2009) was applied according to the information in the GenBank (Benson et al., 2014), TCM-BOL (Chen et al., 2014), and BOLD (Ratnasingham and Hebert, 2007) databases to obtain the remaining OTUs and facilitate species identification. All sequences acquired *via* Sanger sequencing were deposited in the GenBank database. The GenBank accession numbers are shown in **Table 2**. Finally, the resultant BLAST files were imported into MEGAN v6.18.9 for taxonomic analysis and to obtain the statistics regarding the species composition of the traditional herbal patent medicine (Huson et al., 2016).

RESULTS

The Authentication of the Six Herbal Materials in FKDSW and Their Reference ITS2, *matK*, and *rbcL* DNA Barcodes

The authentication of the six herbal materials was further verified *via* DNA barcoding to ensure their accuracy. All the ITS2 sequences of the six herbal ingredients used for producing the lab-made mock FKDSW sample were successfully obtained and assigned to species by blasting to the TCM-BOL database. The *matK* sequences of Paeoniae Radix Alba (Baishao) and Notopterygii Rhizoma et Radix (Qianghuo), and the *rbcL* sequences of Paeoniae Radix Alba (Baishao), Notopterygii Rhizoma et Radix (Qianghuo), and Bupleuri Radix (Chaihu) failed to be amplified by the universal primers, while only the *matK* and *rbcL* sequences of the remaining herbal materials were used for species assignment by blasting to the BOLD and GenBank nt databases. The comprehensive species identification results of three DNA barcodes showed that the six herbal materials were authentic species recorded in the Chinese Pharmacopoeia, which were consistent with the morphological identification results. Finally, the original

TABLE 2 | The species of herbal materials assigned by classic DNA barcoding and their GenBank accession numbers.

Sample ID	Species	GenBank accession number		
		ITS2	*matK*	*rbcL*
HSYC2002	*Angelica sinensis* (Oliv.) Diels	MN712234	MN746764	MN729559
HSYC2036	*Paeonia lactiflora* Pall.	MN712235	—	—
HSYC2037	*Saussurea costus* (Falc.) Lipsch.	MN712236	MW000342	MW000336
HSYC2038	*Hansenia weberbaueriana* (Fedde ex H.Wolff) Pimenov & Kljuykov	MN712237	—	—
HSYC2039	*Leonurus japonicus* Houtt.	MN712238	MW000343	MW000337
HSYC2078	*Bupleurum chinense* DC.	MN712239	MW000344	—

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

species of Angelicae Sinensis Radix (Danggui), Paeoniae Radix Alba (Baishao), Aucklandiae Radix (Muxiang), Notopterygii Rhizoma et Radix (Qianghuo), Leonuri Herba (Yimucao), and Bupleuri Radix (Chaihu) were verified to be *Angelica sinensis* (Oliv.) Diels, *Paeonia lactiflora* Pall., *Saussurea costus* (Falc.) Lipsch. (synonym: *Aucklandia costus* Falc.), *Hansenia weberbaueriana* (Fedde ex H.Wolff) Pimenov & Kljuykov, *Leonurus japonicus* Houtt., and *Bupleurum chinense* DC. The quantity and quality of the extracted DNA are shown in **Supplementary Table S2**. The identification results and GenBank accession numbers are presented in **Table 2**.

An Overview of the High-Throughput Sequencing Data and Shotgun Metabarcoding Assembly

The average extracted DNA concentration from the five FKDSW samples was 110.5 ng/μL, while the A_{260}/A_{280} values ranged from 1.8 to 1.9, indicating that the quantity and quality of the metagenomic DNA extracted from FKDSW adequately met the requirements of high throughput sequencing (**Supplementary Table S3**). A total of 52.16 Gb of clean data was generated from about 174 million paired-end reads using the Illumina NovaSeq sequencing platform. On average, 10.43 Gb of clean data was acquired for each FKDSW sample. The commercial HSZY003 sample displayed the largest quantity of sequencing data at 18.14 Gb, while HSZY140 had the least at 6.6 Gb. Furthermore, 371,553 paired-end reads were enriched for ITS2, *matK*, and *rbcL* from the clean data, while the number of paired-end reads belonging to the ITS2, *matK*, and *rbcL* regions was 217,085, 52,642, and 101,826, respectively. Further analysis showed that the number of reads enriched for each marker was consistent with the clean data of its corresponding sample (**Supplementary Table S4**). In total, 1,248 unique contigs were generated, in which the number of unique contigs of ITS2, *matK*, and *rbcL* was 888, 150, and 210, respectively. Moreover, 466 DNA barcodes were obtained from ITS2, *matK*, and *rbcL* after conventional DNA barcoding region annotation and chimera detection. Subsequently, these DNA barcodes in the conventional DNA barcoding region were clustered into 265 OTUs comprising 228, 23, and 14 OTUs for ITS2, *matK*, and *rbcL*, respectively (**Table 3**). Considering the potential sequencing or assembly errors, further accuracy verification was performed for each OTU *via* read mapping. Finally, 133 reliable OTUs were

obtained, including 98 ITS2 OTUs, 21 *matK* OTUs, and 14 *rbcL* OTUs.

The Species Detection of in the Lab-Made Mock Sample and the Verification of the Assembly Results

A total of 58 reliable OTUs were obtained from the ITS2, *matK*, and *rbcL* regions in the lab-made mock FKDSW sample, including 36 ITS2 OTUs, 12 *matK* OTUs, and 10 *rbcL* OTUs. The species assignment results showed that all the original species of the six prescribed herbal ingredients were successfully detected and included *Angelica sinensis*, *Paeonia lactiflora*, *Saussurea costus*, *Hansenia weberbaueriana*, *Leonurus japonicas,* and *Bupleurum chinense* (**Table 4**). Of these OTUs in the ITS2, *matK*, and *rbcL* regions, seven were identified as *Bupleurum chinense*, which exhibits the most significant number of OTUs among the six labeled species. Although there are only three OTUs belonging to *Leonurus japonicas*, which contains the largest number of mapping reads. And 4,528, 9,906, and 10,263 reads could be mapped to the ITS2, *matK*, and *rbcL* OTUs of *Leonurus japonicas*, respectively.

To determine the assembly accuracy of ITS2, *matK*, and *rbcL* from the shotgun sequencing data, this study compared these sequences (called "assembly sequences" here) with their corresponding reference sequences of the same species obtained *via* Sanger sequencing. For the ITS2 region, all assembly sequences and reference sequences of the original species belonging to six prescribed herbal ingredients were obtained *via* the two sequencing methods, respectively (**Supplementary Figure S1**). The assembly sequences obtained from *Angelica sinensis*, *Hansenia weberbaueriana*, and *Leonurus japonicas* were absolutely consistent with the reference sequences obtained *via* Sanger sequencing. Three types of ITS2 assembly sequences were evident for *Paeonia lactiflora*, one of which was consistent with the reference sequence, while the others were 1–2 base substitutions for the reference sequences. Two types of haplotypes of ITS2 assembly sequences were obtained from *Saussurea costus* and *Bupleurum chinense*, respectively, with 0–4 base difference from the reference sequences. For the *matK* sequences, four species shared sequences that were generated *via* the two sequencing methods. The assembly sequences from the *matK* region of *Angelica sinensis*, *Saussurea costus,* and *Leonurus japonicas* were identical to the

TABLE 3 | The assembly results of three DNA barcodes in the FKDSW samples.

	ITS2	matK	rbcL
Number of unique contigs	888	150	210
Number of DNA barcodes after annotation and chimera detection	386	47	33
Number of OTUs	228	23	14
Average length (bp)	210.5	844.2	703
GC content (%)	59	33.2	42.8

TABLE 4 | Species detection of the ITS2, matK, and rbcL DNA barcoding regions obtained *via* shotgun metabarcoding in the lab-made mock FKDSW sample.

Herbal ingredients	Original species	ITS2	matK	rbcL
Angelicae Sinensis Radix (Danggui)	*Angelica sinensis*	√	√	√
Paeoniae Radix Alba (Baishao)	*Paeonia lactiflora*	√	√	√
Aucklandiae Radix (Muxiang)	*Saussurea costus*	√	√	√
Notopterygii Rhizoma et Radix (Qianghuo)	*Hansenia weberbaueriana*	√	√	√
Leonuri Herba (Yimucao)	*Leonurus japonicus*	√	√	√
Bupleuri Radix (Chaihu)	*Bupleurum chinense*	√	√	√

Note: "√" means that the assembly sequence of this species was obtained.

reference sequences. As for the three *matK* assembly sequences of *Bupleurum chinense*, there were 0, 1, and 6 base differences in these assembly sequences compared with its reference sequences. The assembly sequences obtained *via* shotgun metabarcoding were identical to their corresponding reference sequences of the same species obtained *via* Sanger sequencing for *Angelica sinensis, Saussurea costus,* and *Leonurus japonicas* in the *rbcL* region.

The Authentic and Non-Authentic Ingredients Detected in the Commercial Samples

In total, 98 reliable ITS2 OTUs were obtained from the commercial samples, of which 11 OTUs were identified as the original plant species of six prescribed herbal ingredients. The BLAST results demonstrated that *Angelica sinensis, Paeonia lactiflora, Saussurea costus, Hansenia weberbaueriana, Leonurus japonicus,* and *Bupleurum chinense* were detected based on the ITS2 sequences in four commercial samples. *Hansenia forbesii* (H.Boissieu) Pimenov & Kljuykov, which is another botanical source species for Notopterygii Rhizoma et Radix (Qianghuo), was also detected in the HSZY003, HSZY153, and HSZY154 samples. In addition, *Bupleurum falcatum* L. was detected in four commercial samples and is a non-authentic ingredient for Bupleuri Radix (Chaihu), as shown in **Figure 2** and **Supplementary Table S5**.
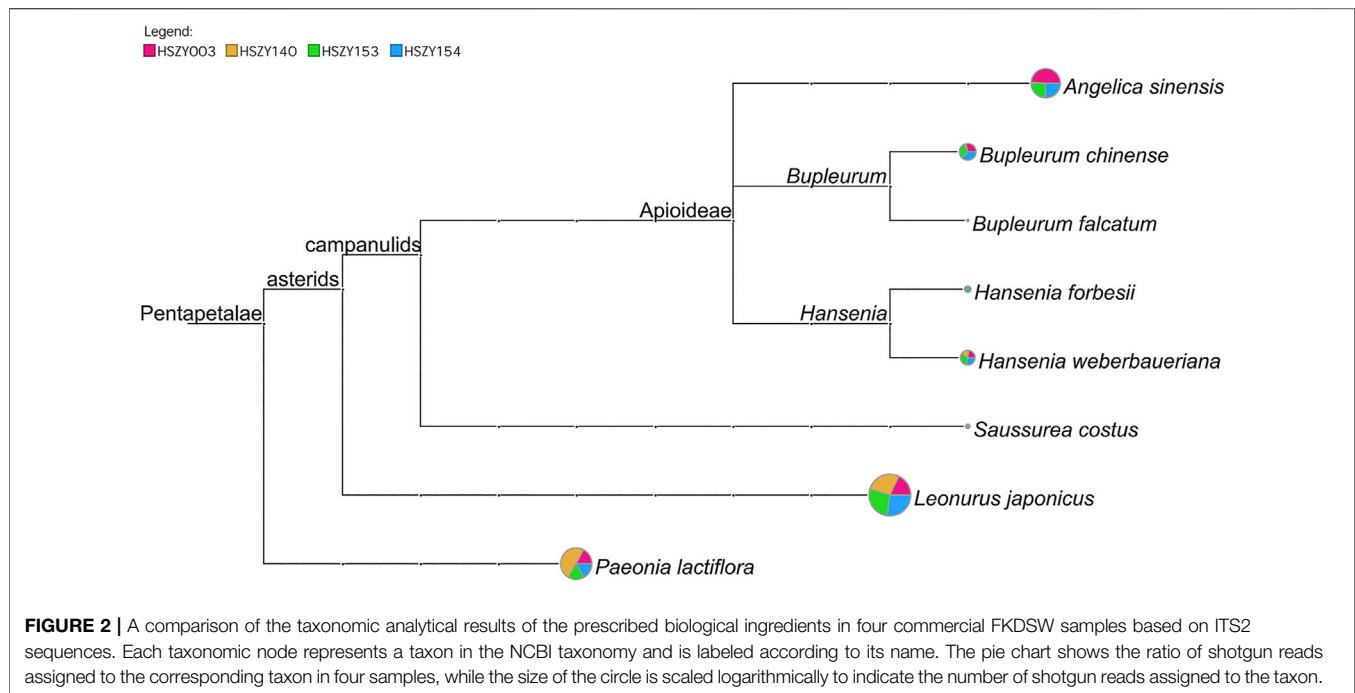
Among the obtained 21 *matK* OTUs and 14 *rbcL* OTUs, *Paeonia lactiflora* and *Leonurus japonicus* were detected in the four commercial FKDSW samples. The original plant species of Angelicae Sinensis Radix (Danggui) and Bupleuri Radix (Chaihu) were determined at the genus level in the commercial FKDSW samples. For example, one OTU obtained from the *matK* region was identified as several species belonging to the genus *Angelica* in the four commercial FKDSW samples, but could not be assigned to a specific species. Moreover, five OTUs obtained

from the *matK* and *rbcL* regions belonged to *Bupleurum* in the three commercial FKDSW samples, of which three were shared by HSZY003, HSZY153, and HSZY154. In addition, there are other contaminating species were found in the commercial FKDSW samples, which are described in detail in **Supplementary Figures S2, S3**, and **Supplementary Table S5**.

In general, six prescribed ingredients in the four commercial FKDSW samples were successfully detected with the combination of the ITS2, *matK*, and *rbcL* sequences. The ITS2 region displayed higher identification efficiency at the species level, while it was difficult to determine some Asteraceae and Apiaceae at the species level using the *matK* and *rbcL* sequences.

Unlabeled Plant Species Detected in the Lab-Made Mock and Commercial Samples

The contaminating plant species found in this study were represented by 68 OTUs from the ITS2, *matK*, and *rbcL* regions, and were classified into 14 families that included 24 genera and 30 species, while the remainder could be resolved at the genera or family levels, as shown in **Supplementary Tables S6–S8**. Ten plant families with 16 genera that included 25 species denoted common field weeds, such as *Artemisia annua* L., *Artemisia argyi* H.Lév. & Vaniot, *Artemisia scoparia* Waldst. & Kit., and *Erigeron canadensis* L. from Asteraceae, *Humulus scandens* and *Cannabis sativa* L. from Cannabaceae, *Ipomoea purpurea* (L.) Roth and *Ipomoea nil* (L.) Roth from Convolvulaceae, *Setaria viridis* (L.) P.Beauv. from Poaceae, as well as *Abutilon theophrasti* Medik. and *Hibiscus trionum* L. from Malvaceae. Of these, *Artemisia, Humulus,* and *Ipomoea* signified the three most common plant genera of these potential species. For the ITS2 region, a total of 11 families with 22 genera that included 28 species were identified for the contaminating plant species. Additionally, six families with six genera that included six species, as well as four families with four genera that included three species, were detected for the *matK* and *rbcL* regions,

FIGURE 2 | A comparison of the taxonomic analytical results of the prescribed biological ingredients in four commercial FKDSW samples based on ITS2 sequences. Each taxonomic node represents a taxon in the NCBI taxonomy and is labeled according to its name. The pie chart shows the ratio of shotgun reads assigned to the corresponding taxon in four samples, while the size of the circle is scaled logarithmically to indicate the number of shotgun reads assigned to the taxon.

respectively. Furthermore, there were distinct differences in the relative species abundance of the contaminating plant species between the mock and four commercial FKDSW samples. For example, only four genera of botanical contamination were found in the mock FKDSW sample, including *Artemisia*, *Humulus*, *Ipomoea*, and *Erigeron*. However, all botanically contaminating species found in this study were detected in the commercial samples, containing a small number of other arbor species that included *Ulmus* of Ulmaceae, *Populus* of Salicaceae, *Robinia* of Fabaceae, and *Broussonetia* of Moraceae. The HSZY003 sample comprised a large proportion of botanical contamination, including 23 genera with 29 species.

## The Fungal Communities in the Lab-Made Mock and Commercial Samples

This study analyzed the fungal communities in five FKDSW samples *via* ITS2. In total, 40 fungal ITS2 OTUs were obtained, and the BLAST results indicated that a total of 17 families that included 26 genera of fungal species were detected in all the FKDSW samples (**Supplementary Table S9**). Pleosporaceae represented the most abundant family, accounting for 17.54–41.29% of the fungal reads in five FKDSW samples, followed by Aspergillaceae, Cladosporiaceae, and Nectriaceae, as shown in **Figure 3**. Further taxonomical classification at the genus level demonstrated that *Alternaria*, *Aspergillus*, *Cladosporium*, and *Fusarium* were the most dominant genera among the 27 detected genera, with a relative abundance of 17.54–41.29%, 9.68–23.39%, 6.05–21.64%, and 7.56–19.30%, respectively, as shown in **Figure 4**. There were differences in fungal communities between the lab-made mock and four commercial FKDSW samples. *Alternaria* was the dominant

genera in the four commercial samples, while *Aspergillus* was the most abundant genera in the lab-made mock sample. In addition, the fungal communities in the HSZY003, HSZY140, HSZY153, and HSZY154 commercial samples were composed of 21, 10, 12, and 8 genera, respectively, while the lab-made mock sample only contained seven genera, namely, *Alternaria*, *Aspergillus*, *Cladosporium*, *Colletotrichum*, *Didymella*, *Fusarium*, and *Plectosphaerella*.

DISCUSSION

Challenges and Biotechnical Advances of Shotgun Metabarcoding in Traditional Herbal Product Authentication

Shotgun metagenomics is widely used for the detection and characterization of microbial community structures and functions, including those found in the human intestinal tract (Laudadio et al., 2018), in soil (Khodakova et al., 2014), in food (Yang et al., 2016), and in marine water (Venter et al., 2004), as well as terrestrial arthropod communities (Zhou et al., 2013). Xin et al. (2018a) applied shotgun metagenomic sequencing for the first time to detect the biological ingredients in Longdan Xiegan Wan, verifying the feasibility of the method for monitoring the species composition in traditional herbal patent medicine. A significant difference was evident compared with Longdan Xiegan Wan since the content of the highest dosage of the all the herbal materials exceeded that of the lowest dosage in FKDSW more than 16 times. In this study, shotgun metabarcoding was performed directly to analyze the biological ingredients in lab-made mock and commercial

FIGURE 3 | The fungal composition detected in lab-made mock and commercial FKDSW samples at the family and genus levels.
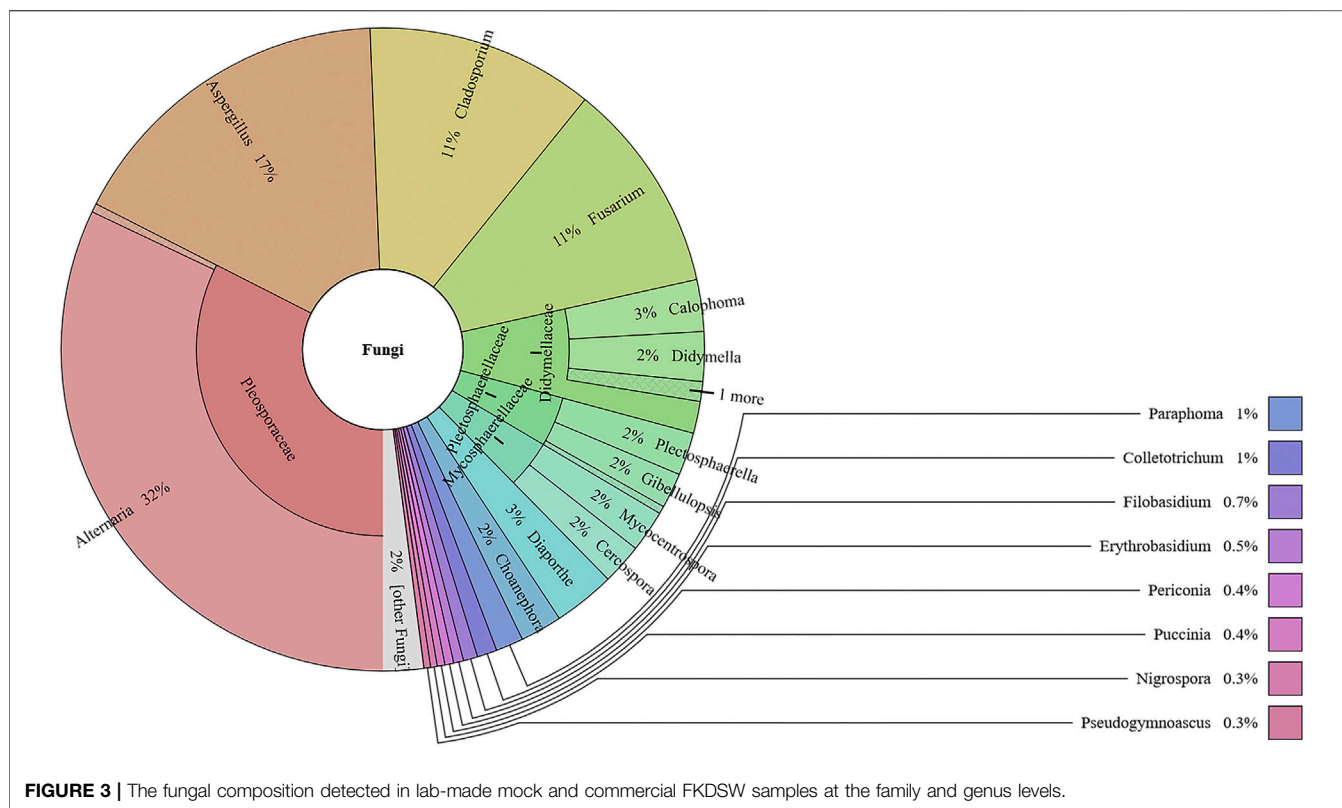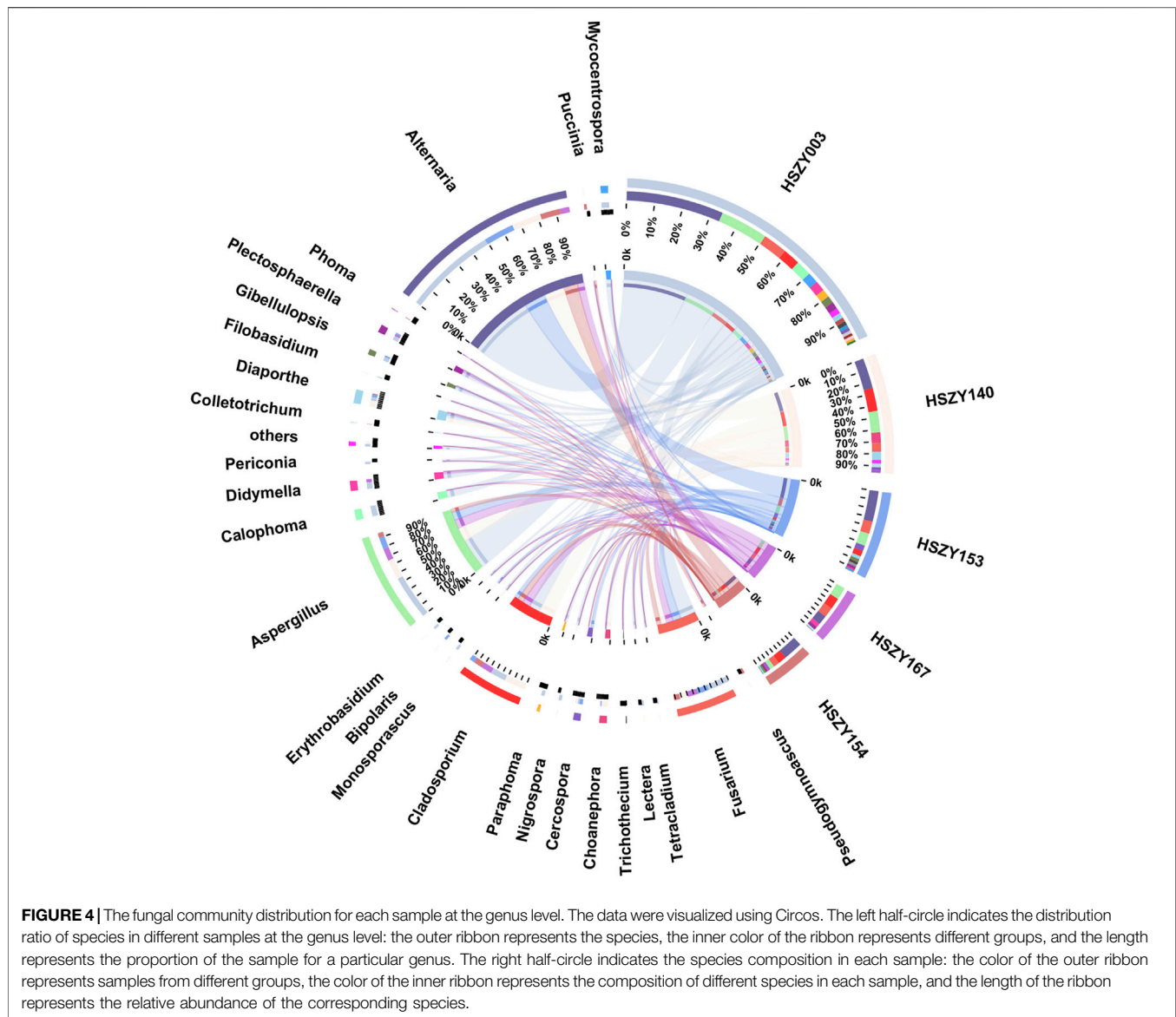
FKDSW samples, while the raw data obtained were qualified for subsequent data processing, such as assembly, annotation, and alignment.

Thereafter, all the assembly sequences of the ITS2, *matK*, and *rbcL* barcodes were successfully obtained for the six prescribed herbal ingredients in the mock FKDSW sample. The obtained assembly sequences were compared with each reference sequence, and the results indicated that all the assembly sequences were identical to the corresponding reference sequences except for part of the ITS2 and *matK* assembly sequences denoting *Paeonia lactiflora, Saussurea costus,* and *Bupleurum chinense.* A possible reason for the base differences between the assembly sequences and the reference sequences may be the multiple ITS2 copies or individual differences among plants (Álvarez, 2003). Although there are frequently hundreds of ITS2 copies with potentially dozens of different sequences in each genome, all ITS2 variants are sufficient for species identification in most cases (Song et al., 2012). Furthermore, there are positive and significant correlations between the amount of content of each herbal material and the read depth of ITS2, *matK*, *rbcL* barcodes (Zhou et al., 2013; Bista et al., 2018), however, the change ratio of the read depth with the amount of content is not with a good linear relationship. Finally, the analytical results suggested that all the prescribed and non-prescribed ingredients in the four commercial samples were detected based on the combination of three DNA barcodes even at low level amount of content. This verified that shotgun metabarcoding can overcome the biased PCR amplification and is particularly suitable for authenticating the biological

ingredients in traditional herbal patent medicine with distinct variances in the herbal material dosages.

A total of 51 species were identified at the species level based on the ITS2 sequences that were obtained, which included all the prescribed medicinal materials and most of the exogenous biological contamination. Nevertheless, only nine species were determined at the species level based on the *matK* and *rbcL* sequences. Some Asteraceae and Apiaceae species were resolved at the genera level or higher, such as the original plant species of Angelicae Sinensis Radix (Danggui) and Bupleuri Radix (Chaihu). For this phenomenon, it may be caused by the identification efficiency of different DNA barcodes (Gao et al., 2010; China Plant et al., 2011). In 2010, Chen et al. tested the performance of seven candidate DNA barcodes (*psbA-trnH*, *matK*, *rbcL*, *rpoC1*, *ycf5*, ITS2, and ITS) from medicinal plants. The results showed that the identification efficiency of the ITS2 region at the species level was 92.7%, and it was proposed as the standard DNA barcoding for medicinal plants (Chen et al., 2010). Compared to the ITS/ITS2 regions, the identification efficiency of the *matK* and *rbcL* DNA barcodes was too low to distinguish more species in Asteraceae and Apiaceae (Gao et al., 2010; Liu et al., 2014). Although the chloroplast gene in *matK* and *rbcL* did not recognize the species in some cases, they displayed an excellent ability to identify orchids (Lahaye et al., 2008) or discriminate among species in congeneric pair-wise comparisons (Newmaster et al., 2006). Moreover, chloroplast genomes have the advantage of maternal inheritance, avoiding genetic recombination, and the high-copy number of plastids per cell are accessible to extract genomic DNA (Twyford and Ness, 2017). Therefore, we recommend that *matK* and *rbcL* as the

FIGURE 4 | The fungal community distribution for each sample at the genus level. The data were visualized using Circos. The left half-circle indicates the distribution ratio of species in different samples at the genus level: the outer ribbon represents the species, the inner color of the ribbon represents different groups, and the length represents the proportion of the sample for a particular genus. The right half-circle indicates the species composition in each sample: the color of the outer ribbon represents samples from different groups, the color of the inner ribbon represents the composition of different species in each sample, and the length of the ribbon represents the relative abundance of the corresponding species.

complementary barcodes to ITS2 for the identification of medicinal plants. To sum up, this study established a biological method for monitoring traditional herbal patent medicine *via* shotgun metabarcoding with ITS2 as the core barcode, and *matK* and *rbcL* as the supplementary barcodes, effectively supplementing traditional identification methods.

Shotgun Metabarcoding is a "Best Practice" Method for the Detection of Biological Contamination Like Non-Authentic Ingredient, Weed, and Fungi Species

In addition to the prescribed herbal medicines, this study identified some unlabeled biological species, such as *Bupleurum falcatum*, weeds, and several fungi species. Bupleuri Radix was officially derived from the dried root of the Asteraceae plants, *Bupleurum*

chinense, and *Bupleurum scorzonerifolium* Willd., according to the Chinese Pharmacopeia (Commission, 2020). However, there are still more than 20 additional species of the genus *Bupleurum* that are habitually utilized in China as a non-authentic ingredient for Bupleuri Radix (Liu et al., 2011). *Bupleurum falcatum* was introduced from Japan into China in the early 1970s and has been recorded as the legal original species of Bupleuri Radix in Japanese and Korean Pharmacopoeia (Yuan et al., 2017; Yeom et al., 2018). In 2007, Zhu et al. (2017) analyzed the chemical profiles of *Bupleurum chinense*, *Bupleurum yinchowense* R.H.Shan & Y.Li, and *Bupleurum falcatum* using UHPLC-QTOF-MS. The results indicated that the chemical profiles of the root samples from the three *Bupleurum* species were similar, especially the characteristic saikosaponins. However, a study suggested that the extracts of *Bupleurum chinense*, *Bupleurum falcatum,* and *Bupleurum scorzonerifolium* exhibit different pharmacological activities (Yuan et al., 2017). Using the FKDSW with the dried root of *Bupleurum*

falcatum instead of the root of *Bupleurum chinense* or *Bupleurum scorzonerifolium* poses a potential threat to its clinical efficacy.

A total of 10 plant families with 16 genera that included 25 weed species were detected in five FKDSW samples, such as species from *Artemisia* (*Artemisia annua*), *Humulus* (*Humulus scandens*), and *Ipomoea* (*Ipomoea purpurea*), which are all common field weeds found on the farmland where traditional herbal materials are cultivated (Ya-ling et al., 2018). It is not unusual for weed species to be mixed into traditional herbal products (Xin et al., 2018b). Jia et al. (2017) detected a variety of non-listed plant species in the commercial Yimu Wan samples using single-molecule realtime sequencing and DNA barcoding. Of these, species from *Humulus*, *Ipomoea*, *Artemisia,* and *Amaranthus* are mostly denoted common weeds, and sequences belonging to these species accounted for a large proportion of the total sequences. Controlling field weeds is not only an essential factor that impacts the artificial standardized cultivation of traditional herbal materials but is also a significant problem affecting global agricultural production. In addition, there are strict standards about using pesticides and fertilizers during the cultivation process of traditional herbal materials (Sang, 2017). Therefore, weed control is challenging during the cultivation of traditional herbal materials because of strict requirements, high standards, and high labor costs. "Leonuri Herba (Yimucao)" is the dried ground part of *Leonurus japonicas*. And it easily mixes with the leaves of some *Artemisia* plants during the manufacturing process due to their overlapping morphological features. *Ipomoea purpurea*/*Ipomoea nil,* and *Humulus scandens* are twining herbs (Zhengyi et al., 2013) that twine around the stem of *Leonurus japonicus* while growing. This might be the reason why *Artemisia*, *Humulus,* and *Ipomoea* were the most common plant genera found in five FKDSW samples.

Similar to agricultural products, plants cultivated for herbal medicines are also vulnerable to fungal contamination during the process of cultivation, processing, transportation, and storage (Xing et al., 2016; Guo et al., 2020; He et al., 2020). This study identified a total of 17 families that included 26 genera of fungal species in five FKDSW samples, and there is a certain difference in fungal communities between the mock and the commercial FKDSW products. The reason for this difference in fungal communities may be due to that the herbal materials for making the mock sample were purchased from medicinal stores, while the raw herbal materials of the commercial products were collected by the FKDSW manufacturers (He et al., 2020). Furthermore, most of the detected fungi species were common pathogenic fungi typically present during the growth period of field plants. For example, *Cercospora canescens* is an important pathogen of Cercospora leaf spot that can lead to serious yield loss of Yardlong bean (Duangsong et al., 2018). It is a remarkable fact that the *Alternaria*, *Aspergillus,* and *Fusarium* genera represent the most common infective fungi in agricultural products, food, and herbal medicines among the dominant bacteria detected, and are potential mycotoxin-producing microbial flora (Xing et al., 2016; Alshannaq and Yu, 2017; Guo et al., 2020). Of these, aflatoxins (AFs) and ochratoxin A, produced by *Aspergillus,* denote the most important contaminants due to their strong carcinogenicity (Perrone and Gallo, 2017).

Therefore, the 2020 edition of the Chinese Pharmacopoeia (Commission, 2020) expressly limit the content of AFs in Chinese medicinal materials, such as Jujubae Fructus (Dazao), Cassiae Semen (Juemingzi), and Hordei Fructus Germinatus (Maiya), which are moldy-prone Chinese medicinal materials, expressed as aflatoxin B_1 (AFB$_1$)$\leq$ 5 µg/kg, and the total amount of aflatoxin G_1, aflatoxin G_2, AFB$_1$ and aflatoxin $B_2 \leq$ 10 µg/kg. Consequently, it is suggested that the quality control of the entire industrial chain should be reinforced using DNA barcoding technology during the cultivation, harvesting, processing, transportation, and storage processes of herbal medicines to ensure the safety and efficacy of traditional herbal medicines.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: NCBI Bioproject: PRJNA663116, BioSample: SAMN16132426, SAMN16132427, SAMN16132428, SAMN16132429, SAMN16132430. SRA accession numbers are: SRR12640731, SRR12640730, SRR12640729, SRR12640728, SRR12640727. 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 MN712234, MN712235, MN712236, MN712237, MN712238, MN712239, MN746764, MW000342, MW000343, MW000344, MN729559, MW000336, MW000337.

AUTHOR CONTRIBUTIONS

LS and JL conceived and designed the study. CZ, WK, and BL collected samples. QZ, WM, and HX performed the experiment. LS, MS, and HX analysed the data. HX and QZ wrote the manuscript. JZ and JJ revised the manuscript. All authors read and approved the final manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2021.607197/full#supplementary-material

REFERENCES

Alshannaq, A., and Yu, J. H. (2017). Occurrence, Toxicity, and Analysis of Major Mycotoxins in Food. *Int. J. Environ. Res. Public Health* 14 (6), 632. doi:10.3390/ijerph14060632

Álvarez, I., and Wendel, J. F. (2003). Ribosomal ITS Sequences and Plant Phylogenetic Inference. *Mol. Phylogenet. Evol.* 29 (3), 417–434. doi:10.1016/s1055-7903(03)00208-2

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

Benson, D. A., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. (2014). GenBank. *Nucleic Acids Res.* 42 (Database issue), D32–D37. doi:10.1093/nar/gkt1030

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. *Mol. Ecol. Resour.* 18, 1020–1034. doi:10.1111/1755-0998.12888

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

Bovo, S., Ribani, A., Utzeri, V. J., Schiavo, G., Bertolini, F., and Fontanesi, L. (2018). Shotgun Metagenomics of Honey DNA: Evaluation of a Methodological Approach to Describe a Multi-Kingdom Honey Bee Derived Environmental DNA Signature. *PLoS One* 13 (10), e0205575. doi:10.1371/journal.pone.0205575

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: Architecture and Applications. *BMC Bioinformatics* 10, 421. doi:10.1186/1471-2105-10-421

Chen, S., Guo, B., Zhang, G., Yan, Z., Luo, G., Sun, S., et al. (2012). [Advances of Studies on New Technology and Method for Identifying Traditional Chinese Medicinal Materials]. *Zhongguo Zhong Yao Za Zhi* 37 (8), 1043–1055. doi:10.4268/cjcmm20120801

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., Yao, H., Han, J., Liu, C., Song, J., Shi, L., et al. (2010). Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. *PLoS One* 5 (1), e8613. doi:10.1371/journal.pone.0008613

Chen, X., Zhou, J., Cui, Y., Wang, Y., Duan, B., and Yao, H. (2018). Identification of Ligularia Herbs Using the Complete Chloroplast Genome as a Super-barcode. *Front. Pharmacol.* 9, 695. doi:10.3389/fphar.2018.00695

Cheng, C. S., Tan, T. Q., Long, Z., Liu, Z. Z., and Zhou, H. (2015). Optimization of DNA Extraction for Chinese Patent Medicine and its Application on Molecular Identification of Ginseng Preparations by MAS-PCR. *Chin. Traditional Herbal Drugs* 46 (17), 2549–2555. doi:10.7501/j.issn.0253-2670.2015.17.008

China Plant, B. O. L. G., Li, D. Z., Gao, L. M., Li, H. T., Wang, H., Ge, X. J., 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. *Proc. Natl. Acad. Sci. U S A.* 108 (49), 19641–19646. doi:10.1073/pnas.1104551108

Coghlan, M. L., Maker, G., Crighton, E., Haile, J., Murray, D. C., White, N. E., et al. (2015). Combined DNA, Toxicological and Heavy Metal Analyses Provides an Auditing Toolkit to Improve Pharmacovigilance of Traditional Chinese Medicine (TCM). *Sci. Rep.* 5, 17475. doi:10.1038/srep17475

Commission, C. P. (1994). *Pharmaceutical Standard of the Ministry of Health of the People's Republic of China, Preparation of Traditional Chinese Medicine*. Beijing: People's Medical Publishing House.

Commission, C. P. (2020). *Pharmacopoeia of the People's Republic of China, Part I.* Beijing: China Medical Science Press.

De Sanctis, V., Soliman, A. T., Elsedfy, H., Soliman, N. A., Soliman, R., and El Kholy, M. (2016). Dysmenorrhea in Adolescents and Young Adults: a Review in Different Country. *Acta Biomed.* 87 (3), 233–246.

Duangsong, U., Laosatit, K., Somta, P., and Srinives, P. (2018). Genetics of Resistance to Cercospora Leaf Spot Disease Caused by Cercospora Canescens and Psuedocercospora Cruenta in Yardlong Bean (Vigna Unguiculata Ssp. Sesquipedalis) × Grain Cowpea (V. Unguiculata Ssp. Unguiculata) Populations. *J. Genet.* 97 (5), 1451–1456. doi:10.1007/s12041-018-1003-z

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-Samtools 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

Gao, T., Yao, H., Song, J., Zhu, Y., Liu, C., and Chen, S. (2010). Evaluating the Feasibility of Using Candidate DNA Barcodes in Discriminating Species of the Large Asteraceae Family. *BMC Evol. Biol.* 10, 324. doi:10.1186/1471-2148-10-324

Gao, Z., Liu, Y., Wang, X., Song, J., Chen, S., Ragupathy, S., et al. (2017). Derivative Technology of DNA Barcoding (Nucleotide Signature and SNP Double Peak Methods) Detects Adulterants and Substitution in Chinese Patent Medicines. *Sci. Rep.* 7 (1), 5858. doi:10.1038/s41598-017-05892-y

Grazina, L., Amaral, J. S., and Mafra, I. (2020). Botanical Origin Authentication of Dietary Supplements by DNA-based Approaches. *Compr. Rev. Food Sci. Food Saf.* 19 (3), 1080–1109. doi:10.1111/1541-4337.12551

Group, C. P. W. (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

He, L., Bai, L., Shan, L., Yu, P., Zhang, L., Dou, X., et al. (2020). Variations in Fungal Microbiota and Aflatoxin Contamination during the Processing of Yuanzhi, a Traditional Chinese Medicine. *Ind. Crops Prod.* 152, 112509. doi:10.1016/j.indcrop.2020.112509

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

Iacovides, S., Avidon, I., and Baker, F. C. (2015). What We Know about Primary Dysmenorrhea Today: a Critical Review. *Hum. Reprod. Update* 21 (6), 762–778. doi:10.1093/humupd/dmv039

Ichim, M. C., Häser, A., and Nick, P. (2020). Microscopic Authentication of Commercial Herbal Products in the Globalized Market: Potential and Limitations. *Front. Pharmacol.* 11, 876. doi:10.3389/fphar.2020.00876

Ichim, M. C. (2019). The DNA-Based Authentication of Commercial Herbal Products Reveals Their Globally Widespread Adulteration. *Front. Pharmacol.* 10, 1227. doi:10.3389/fphar.2019.01227

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

Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., et al. (2016). Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics. *Front. Microbiol.* 7, 459. doi:10.3389/fmicb.2016.00459

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

Khodakova, A. S., Smith, R. J., Burgoyne, L., Abarno, D., and Linacre, A. (2014). Random Whole Metagenomic Sequencing for Forensic Discrimination of Soils. *PLoS One* 9 (8), e104996. doi:10.1371/journal.pone.0104996

Lahaye, R., van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., et al. (2008). DNA Barcoding the Floras of Biodiversity Hotspots. *Proc. Natl. Acad. Sci. U S A.* 105 (8), 2923–2928. doi:10.1073/pnas.0709936105

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

Laudadio, I., Fulci, V., Palone, F., Stronati, L., Cucchiara, S., and Carissimi, C. (2018). Quantitative Assessment of Shotgun Metagenomics and 16S rDNA Amplicon Sequencing in the Study of Human Gut Microbiome. *OMICS* 22 (4), 248–254. doi:10.1089/omi.2018.0013

Li, D., Liu, C. M., Luo, R., Sadakane, K., and Lam, T. W. (2015). 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, 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

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 Pharmacopeia. *Front. Pharmacol.* 8, 931. doi:10.3389/fphar.2017.00931

Liu, X. J., Hu, J., Li, Z. Y., Qin, X. M., Zhang, L. Z., and Guo, X. Q. (2011). Species Classification and Quality Assessment of Chaihu (Radix Bupleuri) Based on High-Performance Liquid Chromatographic Fingerprint and Combined Chemometrics Methods. *Arch. Pharm. Res.* 34 (6), 961–969. doi:10.1007/s12272-011-0613-2

Martin, M. (2011). Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet j.* 17 (1), 10. doi:10.14806/ej.17.1.200

Newmaster, S. G., Fazekas, A. J., and Ragupathy, S. (2006). DNA Barcoding in Land Plants: Evaluation of rbcL in a Multigene Tiered Approach. *Can. J. Bot.* 84 (3), 335–341. doi:10.1139/b06-047

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

Oladosu, F. A., Tu, F. F., and Hellman, K. M. (2018). Nonsteroidal Antiinflammatory Drug Resistance in Dysmenorrhea: Epidemiology, Causes, and Treatment. *Am. J. Obstet. Gynecol.* 218 (4), 390–400. doi:10.1016/j.ajog.2017.08.108

Perrone, G., and Gallo, A. (2017). Aspergillus Species and Their Associated Mycotoxins. *Methods Mol. Biol.* 1542, 33–49. doi:10.1007/978-1-4939-6707-0_3

Porter, T. M., and Hajibabaei, M. (2018). Scaling up: A Guide to High-Throughput Genomic Approaches for Biodiversity Analysis. *Mol. Ecol.* 27 (2), 313–338. doi:10.1111/mec.14478

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

Raclariu, A. C., Heinrich, M., Ichim, M. C., and de Boer, H. (2018). Benefits and Limitations of DNA Barcoding and Metabarcoding in Herbal Product Authentication. *Phytochem. Anal.* 29 (2), 123–128. doi:10.1002/pca.2732

Ranjan, R., Rani, A., Metwally, A., McGee, H. S., and Perkins, D. L. (2016). Analysis of the Microbiome: Advantages of Whole Genome Shotgun versus 16S Amplicon Sequencing. *Biochem. Biophys. Res. Commun.* 469 (4), 967–977. doi:10.1016/j.bbrc.2015.12.083

Ratnasingham, S., and Hebert, P. D. (2007). Bold: The Barcode of Life Data System. *Mol. Ecol. Notes* 7 (3), 355–364. doi:10.1111/j.1471-8286.2007.01678.x

Sang, B. (2017). *Law of the People's Republic of China on Traditional Chinese Medicine,Chinese-English*. Beijing: People's Medical Publishing House.

Schrader, C., Schielke, A., Ellerbroek, L., and Johne, R. (2012). PCR Inhibitors - Occurrence, Properties and Removal. *J. Appl. Microbiol.* 113 (5), 1014–1026. doi:10.1111/j.1365-2672.2012.05384.x

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. (2012). Metagenomic Microbial Community Profiling Using Unique Clade-specific Marker Genes. *Nat. Methods* 9 (8), 811–814. doi:10.1038/nmeth.2066

Shi, L., Chen, H., Jiang, M., Wang, L., Wu, X., Huang, L., et al. (2019). CPGAVAS2, an Integrated Plastome Sequence Annotator and Analyzer. *Nucleic Acids Res.* 47 (W1), W65–w73. doi:10.1093/nar/gkz345

Song, J., Shi, L., Li, D., Sun, Y., Niu, Y., Chen, Z., et al. (2012). Extensive Pyrosequencing Reveals Frequent Intra-genomic Variations of Internal Transcribed Spacer Regions of Nuclear Ribosomal DNA. *PLoS One* 7 (8), e43971. doi:10.1371/journal.pone.0043971

Song, M., Dong, G. Q., Zhang, Y. Q., Liu, X., and Sun, W. (2017). Identification of Processed Chinese Medicinal Materials Using DNA Mini-Barcoding. *Chin. J. Nat. Med.* 15 (7), 481–486. doi:10.1016/s1875-5364(17)30073-0

Sunagawa, S., Mende, D. R., Zeller, G., Izquierdo-Carrasco, F., Berger, S. A., Kultima, J. R., et al. (2013). Metagenomic Species Profiling Using Universal Phylogenetic Marker Genes. *Nat. Methods* 10 (12), 1196–1199. doi:10.1038/nmeth.2693

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., and Willerslev, E. (2012). Towards Next-Generation Biodiversity Assessment Using DNA Metabarcoding. *Mol. Ecol.* 21 (8), 2045–2050. doi:10.1111/j.1365-294X.2012.05470.x

Twyford, A. D., and Ness, R. W. (2017). Strategies for Complete Plastid Genome Sequencing. *Mol. Ecol. Resour.* 17 (5), 858–868. doi:10.1111/1755-0998.12626

van der Valk, J. M. A., Leon, C. J., and Nesbitt, M. (2017). Macroscopic Authentication of Chinese Materia Medica (CMM) : A UK Market Study of Seeds and Fruits. *J. Herbal Med.* 8, 40–51. doi:10.1016/j.hermed.2017.03.007

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., et al. (2004). Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science* 304 (5667), 66–74. doi:10.1126/science.1093857

Vijayvargiya, P., Jeraldo, P. R., Thoendel, M. J., Greenwood-Quaintance, K. E., Esquer Garrigos, Z., Sohail, M. R., et al. (2019). Application of Metagenomic Shotgun Sequencing to Detect Vector-Borne Pathogens in Clinical Blood Samples. *PLoS One* 14 (10), e0222915. doi:10.1371/journal.pone.0222915

Wang, Y. D., Han, X. N., Zhao, Y. D., Lei, T. L., and Lu, J. C. (2017). Identification of Commercial Bupleuri Radix and its Adulterants Based on ITS2 Barcode. *Chin. Traditional Herbal Drugs* 48 (17), 3590–3596.

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

Xing, F., Ding, N., Liu, X., Selvaraj, J. N., Wang, L., Zhou, L., et al. (2016). Variation in Fungal Microbiome (Mycobiome) and Aflatoxins during Simulated Storage of In-Shell Peanuts and Peanut Kernels. *Sci. Rep.* 6, 25930. doi:10.1038/srep25930

Ya-ling, G., Li-bing, Y., Peng, Z., Cheng-jun, Z., and Zhi-ying, H. (2018). Characterization of Weed Communities in Medicinal Plant Fields in Anguo,Hebei. *J. Chin. Med. Mater.* 41 (05), 1048–1053. doi:10.13863/j.issnl001-4454.2018.05.008

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

Ye, S. H., Siddle, K. J., Park, D. J., and Sabeti, P. C. (2019). Benchmarking Metagenomics Tools for Taxonomic Classification. *Cell* 178 (4), 779–794. doi:10.1016/j.cell.2019.07.010

Yeom, M., Kim, E. Y., Kim, J. H., Jung, H. S., and Sohn, Y. (2018). High Doses of Bupleurum Falcatum Partially Prevents Estrogen Deficiency-Induced Bone Loss with Anti-osteoclastogenic Activity Due to Enhanced iNOS/NO Signaling. *Front. Pharmacol.* 9, 1314. doi:10.3389/fphar.2018.01314

Yuan, B., Yang, R., Ma, Y., Zhou, S., Zhang, X., and Liu, Y. (2017). A Systematic Review of the Active Saikosaponins and Extracts Isolated from Radix Bupleuri and Their Applications. *Pharm. Biol.* 55 (1), 620–635. doi:10.1080/13880209.2016.1262433

Zhengyi, W., Raven, P. H., and Deyuan, H. (2013). *Flora of China*. Beijing: Science Press.

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

Zhu, L., Liang, Z. T., Yi, T., Ma, Y., Zhao, Z. Z., Guo, B. L., et al. (2017). Comparison of Chemical Profiles between the Root and Aerial Parts from Three Bupleurum

Species Based on a UHPLC-QTOF-MS Metabolomics Approach. *BMC Complement. Altern. Med.* 17 (1), 305. doi:10.1186/s12906-017-1816-y

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.