High-throughput detection and screening of biological ingredients in Shenling Baizhu San by shotgun metabarcoding

Weishan Mu[1], Zhaolei Zhang[1], Qing Zhao[1], Weijun Kong[1], Hongye Zhao[1], Yanze Liu[1], Jinxin Liu[2], and Linchun Shi[1]

[1]Affiliation not available
[2]Chinese Academy of Medical Sciences & Peking Union Medical College

April 16, 2024

Abstract

High-throughput detection and screening of biological ingredients in a single bulk sample containing a wide range of taxa still have considerable problems. Recently, a new approach named "shotgun metabarcoding" combining shotgun sequencing and multiple barcodes has been proposed. Here, Shenling Baizhu San (SLBZS), a traditional patent medicine composed of ten plant and fungus ingredients was used to test its species detection capability. Mock and pharmaceutical samples of SLBZS were collected, and a total of 39.52 Gb of raw data were obtained via PCR-free shotgun sequencing. All ingredients made for mock samples can be successfully re-detected except Coix lacryma-jobi was failed to be detected in the second mock sample, and the positive control, the roots of Panax quinquefolius, can be re-detected in the second mock sample. For pharmaceutical samples, not only labeled ingredients but also adulterants, e.g. P. quinquefolius, were detected in a sample. The presence of P. quinquefolius was verified using the SNP detection method with a pair of Panax specific primers. In addition, 18 genera of fungal species were detected in both mock and pharmaceutical samples, however, the total number of fungal reads was relatively small. As a result, this study confirmed that shotgun metabarcoding could authenticate all the biological ingredients of SLBZS, has the potential to enable genomic analyses ideally of all species a single bulk sample.

High-throughput detection and screening of biological ingredients in Shenling Baizhu San by shotgun metabarcoding

Weishan Mu[1], Zhaolei Zhang[1], Qing Zhao[1], Weijun Kong[2], Hongye Zhao[1], Yanze Liu[2], Jinxin Liu[1, 2*], Linchun Shi[2*]

[1] Hebei Key Laboratory of Study and Exploitation of Chinese Medicine, Chengde Medical University, Chengde, 067000, China

[2] Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, 100193, China

*** Correspondence** : Jinxin Liu, liujx_23@163.com; Linchun Shi, linchun_shi@163.com

Abstract

High-throughput detection and screening of biological ingredients in a single bulk sample containing a wide range of taxa still have considerable problems. Recently, a new approach named "shotgun metabarcoding" combining shotgun sequencing and multiple barcodes has been proposed. Here, Shenling Baizhu San (SLBZS), a traditional patent medicine composed of ten plant and fungus ingredients was used to test its species detection capability. Mock and pharmaceutical samples of SLBZS were collected, and a total of 39.52 Gb of raw data were obtained *via* PCR-free shotgun sequencing. All ingredients made for mock samples can

be successfully re-detected except *Coix lacryma-jobi* was failed to be detected in the second mock sample, and the positive control, the roots of *Panax quinquefolius* , can be re-detected in the second mock sample. For pharmaceutical samples, not only labeled ingredients but also adulterants, e.g. *P. quinquefolius* , were detected in a sample. The presence of *P. quinquefolius* was verified using the SNP detection method with a pair of *Panax* specific primers. In addition, 18 genera of fungal species were detected in both mock and pharmaceutical samples, however, the total number of fungal reads was relatively small. As a result, this study confirmed that shotgun metabarcoding could authenticate all the biological ingredients of SLBZS, has the potential to enable genomic analyses ideally of all species a single bulk sample.

Keywords: shotgun metabarcoding, High-throughput sequencing, multiple loci, Shenling Baizhu San, species detection,adulterant

1. Introduction

Shenling Baizhu San (SLBZS), composed of ten herbal materials for the treatment of gastrointestinal diseases, was first recorded in the*Taiping Huimin Heji Jufang* in the Song Dynasty. This product is a blended powder of ten herbal materials (The national pharmacopoeia committee of China, 2020b), nine of them are plant ingredients including roots, fruits, and seeds and another one is the sclerotium of a fungus. Each ingredient needs to go through certain processing steps before use to achieve the purpose of purity, flavor, lower toxicity, and drying without deterioration. They are then mixed in certain proportion and crushed into a fine powder. Finally, it is dispensed in certain doses for people to use. Modern clinical and pharmacological studies have verified that SLBZS can regulate the composition of intestinal flora (Feng et al., 2020; Shi et al., 2019; Zhang et al., 2019), exert significant anti-acid capacity (Wu et al., 2010), and improve the symptoms of gastrointestinal diseases (Zhang et al., 2019). At the same time, it also showed certain anti-inflammatory effect (Yang et al., 2014b)to reduce inflammation in the lung, liver, and can reduce liver and lung damage (Feng et al., 2020; Yang et al., 2014a). Besides, it also exerted the functions to improve metabolism and immunity (Chen & Xue, 2014b). Each ingredient of SLBZS is thought to combine and work together to achieve the effects of nourishing the spleen and stomach, and profit the lungs based on the theory of traditional Chinese medicine (TCM). Therefore, ensuring the accuracy of the herbal materials is the first step to ensure the efficacy. Currently, in the 2020 edition of the Pharmacopoeia of the People's Republic of China, the identification method of SLBZS contains the microscopic identification and thin layer chromatography (TLC). Under the microscopic identification for the item of SLBZS, only the microscopic characteristics of nine herbal materials other than Coicis Semen were described, and TLC was only for the detection of the characteristic components of Ginseng Radix et Rhizoma (the roots of *Panax ginseng*) and Glycyrrhizae Radix et Rhizoma (the roots and rhizomes with or without the periderm of *Glycyrrhiza uralensis* , *G. glabra,* or *G. inflata*). In addition, there are researches conducted to quantify the active components of the four ingredients of SLBZS, i.e. Ginseng Radix et Rhizoma, Poria (the sclerotium of *Wolfiporia cocos*), Atractylodis Macrocephalae Rhizoma (the rhizomes with or without periderm of *Atractylodes macrocephala*), Platycodonis Radix (the roots with or without periderm of *Platycodon grandiflorus*), using Ultra Performance Liquid Chromatography (UPLC) to provide a basis for quality control of their prescriptions (Li et al., 2019). A more comprehensive method for monitoring the quality of all herbal ingredients is urgently to be developed.

High throughput sequencing (HTS) has been becoming a hot spot in the identification of biological ingredients of mixtures. A study by Vangay et al., published on the journal of *Cell* using HTS to test stool samples from Thai immigrants in the US and US-born European American individuals, revealed that the gut microbiome diversity and function of non-Western immigrants westernises after immigration to the US (Vangay et al., 2018). A study used four eDNA barcoding assays to detect the biological components (plants, vertebrates, and arthropods) in five common terrestrial substrates, demonstrated that the ability of HTS-based eDNA metabarcoding could be used as a powerful tool for terrestrial biomonitoring (Van Der Heyde et al., 2020). Another study applied the method of combining capture enrichment and HTS to test two freshwater macrozoobenthos mock communities with different biodiversity, and found that capture enrichment provides a more reliable and accurate representation of species occurrences and relative biomasses in comparison with

polymerase chain reaction (PCR) enrichment for bulk DNA (Gauthier et al., 2020). At the same time, the HTS also realizes the identification of ingredients and the detection of pollutants in food. A recent study by Haiminen developed a bioinformatics pipeline that used HTS to detect several samples of high protein powders contained not only in chicken but also pork and beef that were not expected (Haiminen et al., 2019). Another study on the identification of food ingredients showed that HTS could detect the ingredients in sausages, as well as the presence of sequences of microorganisms, plants, and other animal species (Ripp et al., 2014). The successful application of HTS in the identification of mixed species also provides a new idea for the identification of biological ingredients of compound traditional herbal patent medicine. The DNA metabarcoding technique has been successfully used to identify the species of different dosage forms of traditional herbal patent medicine such as Yimu Wan (Jia et al., 2017), Jiuwei Qianghuo Wan (Xin et al., 2018b), and Ruyi Jinhuang San (Shi et al., 2018). Although it is feasible to use PCR-based methods such as DNA metabarcoding to identify traditional herbal patent medicine (Shi et al., 2018; Xin et al., 2018b), PCR bias may lead to false-negatives in certain species (Piñol et al., 2015), and universal primer design is also a challenge for such methods. Researchers used shotgun sequencing technology combined with DNA barcoding region to analyze the species composition of traditional herbal patent medicine Wuhu San and proposed the concept of "shotgun metabarcoding" (Liu et al., 2021a). This method did not depend on PCR amplification, and could successfully detected all labeled ingredients in the mock samples of Wuhu San prescription, which proved the feasibility of the method, and the applicability of the method was verified with pharmaceutical samples (Liu et al., 2021a). In the same year, the researcher demonstrated that the method was also applicable to the quality control of the traditional herbal patent medicine Qingguo Wan, which has a different dosage form from Wuhu San (Liu et al., 2021b).

Here, this study obtained four barcode regions of ITS2,*psbA-trnH* , *matK* , and *rbcL* through shotgun sequencing and bioinformatics analysis, and analyzed the biological ingredients in SLBZS. Mock samples of SLBZS were used to establish a standard analytical method and verify its feasibility. Subsequently, the method was successfully used in the biological ingredients analysis of pharmaceutical samples to verify the applicability of the method, and the biological ingredients of the pharmaceutical samples were evaluated.

2. Materials and methods

2.1. Sample collectionand preparation of lab-made mock SLBZS samples

Ten herbal materials of SLBZS were purchased from a Beijing Tongrentang drug store (**Figure 1)** . DNA barcoding techniques were used to obtain ITS2, *psbA-trnH* , *matK* , and *rbcL* of the herbal materials. These barcodes are used to verify the accuracy of the original species of herbal materials on the one hand, and on the other hand to verify the accuracy of subsequent sequencing assembly (**Supplementary Table 1**). Subsequently, these authenticated herbal materials were used to make mock samples according to the preparation method under the "Shenling Baizhu San" item of the Chinese Pharmacopoeia (2020 edition). The proportion of each medicinal material listed in the preparation is shown in **Supplementary Table 2** . Two mock samples were made and named HSZY159 and HSZY171, respectively. To verify the sensitivity of the method, 6.4 mg powder of *Panax quinquefolius* root was added to each 100 mg of the mock sample HSZY171 in the same proportion as the lowest proportion of the prescribed ingredient in SLBZS. In addition, four pharmaceutical samples of SLBZS with different production batch numbers were collected from different pharmacy stores, and were marked as A03, HSZY152, HSZY191, and HSZY192.

2.2. DNA extraction and PCR amplification

The DNA of herbal materials was extracted with a plant genomic DNA extraction kit (Tiangen Biochemical Technology (Beijing) Co., Ltd, China), and partially optimized regarding the DNA barcoding principles for traditional Chinese herbal medicine (Chen et al., 2014a) and previous research (Liu et al., 2017). Then four regions, ITS2, *psbA-trnH* , *matK,* and *rbcL* , were selected for DNA barcodes. Their primers and amplification conditions were set in accordance with the previous study (Chen et al., 2010). PCR products were confirmed by gel electrophoresis on 1.0% agarose gels after staining with GelRed nucleic acid gel stain (Biotium, USA).

The sample pretreatment and meta-genomic DNA extraction of the mock and pharmaceutical samples were

performed following the previously published related research protocols (Liu et al., 2021a).100 mg of each sample was collected for sample pretreatment. DNA concentration was evaluated using spectrophotometric measurement of absorbance at 260 nm wavelength (NanoDrop one, Thermo Fisher Scientific Inc., USA).

2.3. Sanger sequencing, HTS, and data analysis

The PCR products of medicinal materials were bi-directionally sequenced using an ABI 3730xL DNA Analyzer (Thermo Fisher Scientific Inc., USA). After constructing a PCR-free library, the meta-genomic DNA of SLBZS was sequenced using the Illumina NovaSeq platform. For the Sanger sequencing data, Codoncode Aligner v 9.0.1 (CodonCode Corp., Dedham, MA, USA) was used to assemble the sequencing chromatograms and remove the primer sequences. For the Illumina sequencing data, bioinformatics analysis methods including low-quality sequence filtering, reads enrichment, sequences assembly, short sequences mapping, chimera detection and other steps were performed as described in the previous study (Liu et al., 2021a). Usearch v11 (https://www.drive5.com/usearch/) was used to cluster sequences with 100% identity into operational taxonomic units (OTUs), and representative sequences in each OTU were selected for further analysis. The sequencing depth and coverage values were calculated using samtools v1.10 (Etherington et al., 2015). The representative sequence of each OTU e was used for further analysis when the sequence depth is [?]3 and the coverage is [?]95%, otherwise, it was removed as a poor quality OTU. The remaining high-quality OTUs were used for species assignment by searching the DNA barcoding system for traditional Chinese medicines (TCM-BOL) (Chen et al., 2014a), the barcode of life data system (BOLD) (Ratnasingham & Hebert, 2007) and the nucleotide of GenBank databases (Nt) using the basic local alignment search tool (BLAST) (Camacho et al., 2009). Finally, the statistics and taxonomic visualization of the species composition of prescription ingredients in the SLBZS, were performed using MEGAN v 6.19.4 (Huson et al., 2016).

2.4. The validation of adulterant ingredient, *P. quinquefolius,* in pharmaceutical samples based on DNA Barcoding and SNP detection technologies

To verify the reliability of the presence of *P. quinquefolius* in the pharmaceutical samples, SNP-based identification method was used for distinguishing the species of *P. ginseng* and *P. quinquefolius* from each other based on ITS2 sequences (Chen et al., 2013). Specific primers for the ITS2 region of *P. ginseng* and *P. quinquefolius* were designed using Oligo 6 in this study. The forward and reverse primers are 5'-AGTCTTTGAACGCAAGTTG-3' and 5'-CAAGGACTCGCATTTGGGC-3', respectively. Subsequently, powders of *P. ginseng* and *P. quinquefolius*roots were mixed in different ratios of 1: 9, 3: 7, 5: 5, 7: 3, and 9: 1, to verify whether the method has the ability to detect the presence of *P. quinquefolius* in the mock mixed powder. DNA extraction, PCR amplification, and sequencing were performed according to the relevant methods described in Sections 2.2 and 2.3. Sequence alignment, SNP detection, and peak map observation were performed using CodonCode Aligner v 9.0.1. Finally, this primer pair was used to amply the specific ITS2 regions of *P. ginseng* and *P. quinquefolius* , and to observe whether there were double peaks in the sequencing traces.

3. Results

3.1. Summary for the data analysis results of the shotgun sequencing

Through DNA barcoding technology, except for Poria, the other nine herbal materials and the positive control are all identified to be authentic medicinal materials (**Supplementary Table 1**). The ITS2 sequence of Poria was not successfully obtained and the herbal material was identified as authentic by macroscopical and morphological identification. The GenBank accession numbers of each DNA barcode sequence are listed in**Table 1** . Certain DNA barcode regions of some herbal materials were not successfully obtained due to the fail of PCR amplification. Their relevant reference sequences of these species were downloaded from GenBank for subsequent assembly accuracy analysis (**Table 1**).

A total of 39.52 Gb of raw data were obtained from the six samples*via* shotgun sequencing, including 131,786,682 paired-end sequencing reads. An average of 6.59 Gb raw data was obtained for each sample. A total of 3,391,155 paired-end sequencing reads were enriched for the ITS2, *psbA-trnH* , *matK* , and *rbcL*

regions after removing low-quality sequences. The detailed sequencing results from all samples were shown in **Supplementary Table 3** . There were 460, 10158, 75, and 127 contigs belonging to ITS2,*psbA-trnH* , *matK* , and *rbcL* were obtained by combing the assemble results of MEGAHIT and MetaSPAdes. Then, there were 139, 49, 44, and 45 sequences of ITS2, *psbA-trnH* , *matK* , and*rbcL* DNA barcoding regions were yielded after annotation and removal of the chimeras. After clustering the obtained DNA barcode sequences according to 100% sequence similarity, the ITS2 region yielded a total of 53 OTUs, with an average length of 209 bp and an average GC content of 55.8%. The number of OTUs obtained for the chloroplast region was significantly lower than that of the ITS2 sequences, with 14, 13 and, 11 for the *psbA-trnH* , *matK* , and *rbcL* regions, respectively. In addition, the GC content of*psbA-trnH* , *matK* , and *rbcL* sequences was also lower than that of the ITS2 sequences (**Table 2**).

3.2. The evaluation of the shotgun metabarcoding assembly accuracy and species detected in mock samples

There were 27, 10 12, and 10 reliable OTUs of ITS2, *psbA-trnH* ,*matK* , and *rbcL* were obtained in mock samples (HSZY159 and HSZY171). All the ingredients mixed in HSZY159 can be successfully re-detected, whereas nine ingredients (except for *C. lacryma-jobi*) and the positive control, the roots of *P. quinquefolius* , could be detected in HSZY171 (**Table 3**).

To confirm the reliability of OTUs obtained from shotgun sequencing, this study compared the sequence differences between these OTUs and the corresponding reference barcodes obtained from Sanger sequencing**(Table 1)** . For ITS2, the OTUs of seven species were exactly the same with their reference barcodes; two, two, and three types of ITS2 OTUs have been obtained for *W. villosa* , *L. purpureus* , and*N. nucifera* . Except for OTUs consistent with the three species' reference barcodes, there 1-6 different bases between other OTUs and reference barcodes. Furthermore, the proportions of reads belonging to the OTUs same as the reference barcodes of *W. villosa* , *L. purpureus* , and *N. nucifera* were 37.3%, 47.9%, and 50.0%, respectively. For the*psbA-trnH* region, OTUs of *N. nucifera* , *W. villosa* ,*C. lacryma-jobi* , *G. uralensis* , *D. oppositifolia* , and *Panax ginseng* were identical to their reference barcodes,*psbA-trnH* OTUs of *L. purpureus* and *Platycodon grandiflorus* showed eight and three bases difference with their reference barcodes. For the *matK* region, OTUs of *N. nucifera* , *G. uralensis* , and *Panax ginseng* were completely consistent with their reference barcodes. Three types of OTUs had been obtained for both *Platycodon grandiflorus* and *W. villosa* , and there were 0-3 different bases comparing to their reference barcodes. Finally, *rbcL* OTUs of seven species were identical to their reference barcodes. Similar to the situation in *matK* two additional OTUs with two and three different bases has been observed for *W. villosa* in addition to the OTU same as its reference barcode. In addition, all the OTUs of ITS2, *psbA-trnH* , *matK* , and*rbcL* for the positive control, *Panax quiquefolium* , were exactly the same as its reference barcodes.

3.3. Ingredients analysis of SLBZS pharmaceutical samples*via* shotgun metabarcoding

There were 84 reliable OTUs were obtained from the shotgun sequencing data of pharmaceutical samples. Among them, OTUs for ITS2, *psbA-trnH* , *matK* and *rbcL*were 46, 14, 13, and 11 were, respectively. There were 50,797 reads have been mapped to ITS2 OTUs, which far exceeded the number of reads mapped to *psbA-trnH* , *matK* and *rbcL* .

For ten labeled ingredients, there were detected ten, ten, nine, and seven ingredients have been detected in A03, HSZY152, HSZY191, and HSZY192, respectively (**Figure 2-3, Supplementary Figure 1-2**). Detailed reads of the prescription ingredients in the pharmaceutical samples based on four barcodes were shown in **Supplementary Table 4-7** . When using the ITS2 region for ingredient detection, it was found that the sequences of *W. cocos* , *N. nucifera* , *L. purpureus* , *W. villosa* , and *Glycyrrhiza sp.* were detected in the pharmaceutical samples. Among them, *Glycyrrhiza sp.*represents one of the three original species recorded in the Chinese Pharmacopoeia "Glycyrrhizae Radix et Rhizoma" item**(Supplementary Table 1)** , because there were two types of OTUs identified as *Glycyrrhiza sp.* , one of which was identified as*G. uralensis* and the other one as *G. glabra* or *G. inflata* . The ITS2 sequences *of A. macrocephala* ,*Platycodon grandiflorus* , *C. lacryma-jobi* , and *Panax ginseng* have been detected in three of the four samples. The

sequence of *D. oppositifolia* was not detected in all pharmaceutical samples based on the ITS2 region. Species detection based on the*psbA-trnH* sequences revealed that sequences of *D. oppositifolia* , *N. nucifera* , *L. purpureus* , *W. villosa* , and *Panax* sp. existed in all pharmaceutical samples. The *psbA-trnH* sequences of*Glycyrrhiza sp.* and *Platycodon grandiflorus* were detected in A03, HSZY152, and HSZY191, but not in HSZY192. In contrast, the*psbA-trnH* sequence of *C. lacryma-jobi* was only detected in HSZY192. The *psbA-trnH* sequence of *A. macrocephala* was not detected in all pharmaceutical samples. The identification results based on other two chloroplast sequences *matK* and *rbcL* were consistent with the identification results of the *psbA-trnH*region, except for *D. oppositifolia* . *D. oppositifolia* was not detected in one of pharmaceutical samples based on the *matK*region, while it was detected in all pharmaceutical samples based on the*rbcL* and *psbA-trnH* regions.

At the same time, sequences of some unlabeled ingredients were also detected. For example, the ITS2 sequence of *Panax quinquefolius*was detected in HSZY192. Based on the ITS2 sequence, 30 fungal-related OTUs were obtained from two mock samples and three of the four pharmaceutical samples, belonging to 14 families and 18 genera**(Supplementary Figure 4-5, and Supplementary Table 8)** . However, the overall number of fungal reads detected in this study was relatively small comparing to reads belonging to labeled ingredients, accounting for only 0.44% of the total reads of the ITS2 region.

3.4. The Validation of adulterant ingredient, *Panax quinquefolius*, found in pharmaceutical samples based on SNP-detection method

One OTU has been assigned to be *P. quinquefolius* in the pharmaceutical sample HSZY192 by the shotgun metabarcoding approach. The number of paired-end reads for this OTU accounted for 12.6% of the total number of ITS2 reads in this sample. Based on the analysis of the experimental and published ITS2 sequences of *P. ginseng* and*P. quinquefolius* , two stable SNPs (32 bp and 43 bp) were found and can be used to distinguish each other (Chen et al., 2013). A combination of C and T at positions 32 and 43 indicated there was*P. ginseng* , while a combination of T and C indicated there was*P. quinquefolius* . In order to confirm the existence of adulterant ingredients, *P. quinquefolius* , in HSZY192, we designed the taxon-specific primers of the ITS2 region for the two species. The trace data for PCR products of the mixed powder indicated that when the*P. ginseng* : *P. quinquefolius* content ratios were 1:9, 3:7, 5:5, and 7:3, different degrees of double peaks could be detected at the SNP positions (**Figure 4**). The results showed that the method can verify the existence of *P. quinquefolius* in the mixed powder of *P. ginseng* and *P. quinquefolius* . Subsequently, the PCR products ofpharmaceutical samples were sequenced and analyzed, and no double peaks were observed at the SNP positions. The combination of C and T was found at positions 32 and 43 from A03, HSZY152, and HSZY191, indicated that only the legal ingredient, *P. ginseng* , was detected in these samples. However, the combination of T and C was found at positions 32 and 43 from HSZY192, indicated that this sample does contain adulterant ingredient,*P. quinquefolius* , but not the labeled ingredients, *P. ginseng* **(Supplementary Figure 3)** . This result was consistent with that observed by the shotgun metabarcoding approach.

4. Discussion

4.1. Feasibility and advantages of shotgun metabarcoding for the authentication of labeled ingredients in SLBZS

Shotgun sequencing technology has been successfully applied to species detection of complex components in areas such as biodiversity assessment (Crampton-Platt et al., 2016; Warden et al., 2016), animal diet analysis (Srivathsan et al., 2015), and food composition identification (Haiminen et al., 2019). This study used high-throughput sequencing data of total DNA of traditional herbal patent medicine to test whether multiple ingredients could be detected in a single analysis. The results showed that, ten and nine prescription ingredients were successfully detected in lab-made samples using shotgun metabarcoding. *C. lacryma-jobi*was not detected in one of the mock samples. We guessed that because*C. lacryma-jobi* was used as a medicinal ingredient after being roasted, and so the DNA might be degraded severely after high-temperature treatment, which made it unsuccessful to obtain these commonly used barcode regions. It might also be due to the different DNA content, inhibitory substances contained in different herbal materials and the difficulty of

DNA extraction, resulting in the difference in the number of reads obtained. In the mock sample HSZY159, only the ITS2 sequence of *C. lacryma-jobi* was obtained, and the number of reads was only 12, and the sequence depth was only 4.89. In the initial data analysis of this study, it was found that *D. oppositifolia* could be analyzed by chloroplast regions, but the relevant assembly sequence was not obtained in the ITS2 region of the nuclear genome. This study and several other studies had shown that the success rate of ITS2 sequences amplification of *D. oppositifolia* was very low (Meng et al., 2020). This leads to the lack of its reference sequence in the database, and the relevant reads couldn't be mapped. To obtain reliable reference sequences, the genomic sequences associated with*Dioscorea sp.* were searched in the SRA public database of NCBI, and the ITS2 region was obtained through the assembly. Finally,*D. oppositifolia* was successfully detected in two mock samples based on the ITS2 region. In addition, Poria, as a fungal medicinal material, is the sclerotium of *W. cocos* , in which there is no chloroplast genes. Its ITS2 sequence was not successfully obtained by amplification using universal primers (Chen et al., 2010). However, the ITS2 sequence of *W. cocos* was successfully found in both mock and pharmaceutical samples *via* shotgun metabarcoding in this study, and further verified that shotgun metabarcoding could overcome the limitation of amplicon-based methods (Liu et al., 2021a). The positive control of *P. quinquefolius* added to the mock sample was also successfully detected. The result proved the feasibility and sensitivity of this method. The consistency analysis of the assembled sequences and the reference sequences showed that the method has high accuracy. In addition, due to the problem of uneven mixing when preparing a PCR-free library, there were some differences in sequencing datasets of different samples (**Supplementary Table 3**). Nevertheless, it had little effect on the results, especially for the labeled ingredients. For example, only 3.45 Gb of shotgun sequencing data has been generated for HSZY159, but all the labeled ingredients were detected. The similar results has been reported by study of Qingguo Wan (Liu et al., 2021b), which might due to the result of the abundant copies of chloroplast and ITS2 sequences in the plant cells (Liu et al., 2021b). Therefore, this study provided a reliable and operable method for the identification of biological ingredients of traditional herbal patent medicine.

4.2. Unexpected species detected *via* shotgun metabarcoding

This study also found several unexpected species in both mock and pharmaceutical samples, e.g.*P. quinquefolius* , as a common adulterant of *P. ginseng* , was detected in HSZY192. Then the ITS2 barcode of *P. quinquefolius* was used as the reference sequence, and the shotgun sequencing reads of pharmaceutical samples were used for short reads mapping using CodonCode Aligner by employing bowtie2 program. The proportion of reads belonging to *P. ginseng*at the SNP positions in ITS2 were calculated (**Supplementary Table 9**), and the results showed that the reads of *P. ginseng*were dominant in A03, HSZY152 and HSZY191, while the reads of *P. quinquefolius* were dominant in HSZY192. Then, this result was verified by using SNPs positions in the ITS2 region of *P. ginseng* and*P. quinquefolius* , and confirmed that *P. quinquefolius* did exist in HSZY192. The analysis results of the two methods were consistent, which proves that shotgun metabarcoding technology can accurately detect adulterated ingredients of the prescription ingredients. Because *P. ginseng* and *P. quinquefolius* were derived from closely related plants (The national pharmacopoeia committee of China, 2020b), their morphological, tissue structure, and chemical composition are very similar (Yap et al., 2008). If they were made into decoction pieces, powdered or traditional herbal patent medicines, it was even more difficult to distinguish each other. In addition, the fungi detected in this study mainly include*Fusarium* , *Cystofilobasidium* , *Aspergillus* , and*Tausonia* at the genus level. These fungal contaminations may occur in multiple stages such as planting, storage, and transportation (Pitt et al., 2013). There are currently more than 20 kinds of herbal materials that need to be tested for mycotoxins in the current Chinese Pharmacopoeia, including Nelumbinis Semen and Coicis Semen, which are the prescription ingredients of SLBZS in this study (The national pharmacopoeia committee of China, 2020a; The national pharmacopoeia committee of China, 2020b). However, the total number of fungal reads detected in this study was very small, accounting for only 0.44% of the total reads in the ITS2 region. In another traditional herbal patent medicine study, the ITS2 sequence of fungi belonging to*Aspergillus* was detected in Longdan Xiegan Wan (Xin et al., 2018a). But this study did not compare the reads number of*Aspergillus* and prescription ingredients, and it was likely to overestimate the number of fungi presented in LDXGW.

4.3. Combined use of multiple barcodes could get more accurate and complete analysis results

This study selected multiple barcodes for species identification. Among them, the nuclear genome ITS2 region showed higher authentication efficiency at the species level (Yao et al., 2020). Chloroplast gene sequences including *psbA-trnH* , *matK* , and *rbcL* have been widely used in the field of plant identification (Group et al., 2009). However, there were certain limitations in the identification of some chloroplast sequences at the species level (Chen et al., 2013; Liu et al., 2014). For example, the *psbA-trnH* , *matK* , and*rbcL* sequences of *P. ginsen* and *P. quinquefolius*had no mutation sites, so the BLAST results matched to the two species simultaneously. Therefore, it is necessary to combine the ITS2 region to accurately distinguish the two species. In this study, the ITS2 assembly sequence of *A. macrocephalala* , one of the labeled ingredients, could be obtained in all samples, but its chloroplast *psbA-trnH* ,*matK* , and *rbcL* sequences were not obtained. To analyze the reasons for this phenomenon, we downloaded the chloroplast genome sequence of *A. macrocephala* from NCBI as Query sequence and found that some samples can obtain incomplete *matK* sequence after assembly **(Supplementary Figure 6)** . It might be the reason that traditional herbal patent medicine has been concocted and the DNA is degraded seriously. It may also be due to uneven sequencing of the genome, resulting in low coverage of some regions. At the same time, the nuclear genome exists in all parts of the plant. But the underground parts of some plants such as Atractylodis Macrocephalae Rhizomamay contain fewer chloroplasts (Xiong & Yan, 2016). However, for *D. oppositifolia* , its ITS2 sequence was not obtained in any of the four pharmaceutical samples. Subsequently, the four pharmaceutical samples were checked by blast to observe if they contain the reads of the*D. oppositifolia* ITS2 sequence. The results showed that there were zero, six, one, and nine reads in A03, HSZY152, HSZY191, and HSZY192, but the sequencing depth was relatively low, and ITS2 sequences could not be obtained directly. However, in combination with chloroplast regions, *D. oppositifolia* was further be detected in all pharmaceutical samples. Therefore, the combined use of multiple barcodes could guarantee the high accuracy analytical results (Arulandhu et al., 2017; Group et al., 2009)

5. Acknowledgements

This work was supported by Beijing 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, Herbgenomics Team of the Youth PI (principle investigator) Science and Technology Innovation Team Project of Chengde Medical University.

6. References

Arulandhu, A. J., Staats, M., Hagelaar, R., Voorhuijzen, M. M., Prins, T. W., Scholtens, I., . . . Kok, E. (2017). Development and validation of a multi-locus DNA metabarcoding method to identify endangered species in complex samples. *Gigascience, 6* (10), 1-18. https://doi.org/10.1093/gigascience/gix080

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics, 10* , 421. https://doi.org/10.1186/1471-2105-10-421

Chen, S., Pang, X., Song, J., Shi, L., Yao, H., Han, J., & Leon, C. (2014a). A renaissance in herbal medicine identification: from morphology to DNA. *Biotechnol Adv, 32* (7), 1237-1244. https://doi.org/10.1016/j.biotechadv.2014.07.004

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., . . . Leon, C. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS One, 5* (1), e8613. https://doi.org/10.1371/journal.pone.0008613

Chen, X., Liao, B., Song, J., Pang, X., Han, J., & Chen, S. (2013). A fast SNP identification and analysis of intraspecific variation in the medicinal Panax species based on DNA barcoding. *Gene, 530* (1), 39-43. https://doi.org/10.1016/j.gene.2013.07.097

Chen, X., & Xue, J. (2014b). Effect Analysis of Shenlingbaishusan in the Immune Function of Patients after the Gynecological Abdominal Surgery. *Modern Diagnosis & Treatment, 25* (007), 1502-1503.

Crampton-Platt, A., Yu, D. W., Zhou, X., & Vogler, A. P. (2016). Mitochondrial metagenomics: letting the genes out of the bottle. *Gigascience, 5* (1), 15. https://doi.org/10.1186/s13742-016-0120-y

Etherington, G. J., Ramirez-Gonzalez, R. H., & 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. https://doi.org/10.1093/bioinformatics/btv178

Feng, J., Dai, W., Zhang, C., Chen, H., Chen, Z., Chen, Y., . . . Zhou, Y. (2020). Shen-ling-bai-zhu-san ameliorates inflammation and lung injury by increasing the gut microbiota in the murine model of Streptococcus pneumonia-induced pneumonia. *BMC complementary medicine and therapies, 20* (1), 159. https://doi.org/10.1186/s12906-020-02958-9

Gauthier, M., Konecny-Dupre, L., Nguyen, A., Elbrecht, V., Datry, T., Douady, C., & Lefebure, T. (2020). Enhancing DNA metabarcoding performance and applicability with bait capture enrichment and DNA from conservative ethanol. *Mol Ecol Resour, 20* (1), 79-96. https://doi.org/10.1111/1755-0998.13088

Group, C. P. W., Hollingsworth, P. M., Forrest, L. L., Spouge, J. L., Hajibabaei, M., Ratnasingham, S., . . . Erickson, D. L. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences, 106* (31), 12794-12797. https://doi.org/10.1073/pnas.0905845106

Haiminen, N., Edlund, S., Chambliss, D., Kunitomi, M., Weimer, B. C., Ganesan, B., . . . Beck, K. L. (2019). Food authentication from shotgun sequencing reads with an application on high protein powders. *NPJ science of food, 3* , 24. https://doi.org/10.1038/s41538-019-0056-6

Huson, D. H., Beier, S., Flade, I., Gorska, A., El-Hadidi, M., Mitra, S., . . . Tappu, R. (2016). MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Computational Biology, 12* (6), e1004957. https://doi.org/10.1371/journal.pcbi.1004957

Jia, J., Xu, Z., Xin, T., Shi, L., & Song, J. (2017). Quality Control of the Traditional Patent Medicine Yimu Wan Based on SMRT Sequencing and DNA Barcoding. *Frontiers in Plant Science, 8* , 926. https://doi.org/10.3389/fpls.2017.00926

Li, J., Wang, M., Shi, Y., Cui, C., Jia, T., & Wang, Y. (2019). Simultaneous Determination of Six Constituents in Shenling Baizhu Pulvis and Establishment of Finger- prints by UPLC. *China Pharmacist, 22* (02), 214-218.

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

Liu, J., Shi, L., Han, J., Li, G., Lu, H., Hou, J., . . . Downie, S. R. (2014). Identification of species in the angiosperm family Apiaceae using DNA barcodes. *Mol Ecol Resour, 14* (6), 1231-1238. https://doi.org/10.1111/1755-0998.12262

Liu, J., Shi, L., Song, J., Sun, W., Han, J., Liu, X., . . . Chen, S. (2017). BOKP: A DNA Barcode Reference Library for Monitoring Herbal Drugs in the Korean Pharmacopeia. *Front Pharmacol, 8* , 931. https://doi.org/10.3389/fphar.2017.00931

Liu, J., Shi, M., Zhao, Q., Kong, W., Mu, W., Xie, H., . . . Shi, L. (2021b). Precise Species Detection in Traditional Herbal Patent Medicine, Qingguo Wan, Using Shotgun Metabarcoding. *Front Pharmacol, 12* , 607210. https://doi.org/10.3389/fphar.2021.607210

Meng, X., Meng, X., Zhou, G., Wang, B., Xu, N., Li, Y., & Meng, F. (2020). Application of DNA Barcode in Identification of D. oppositifolia and Its Adulterants. *Liaoning journal of traditional Chinese medicine, 47* (04), 163-166.

Pinol, J., Mir, G., Gomez-Polo, P., & Agusti, N. (2015). Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Mol Ecol Resour, 15* (4), 819-830. https://doi.org/10.1111/1755-0998.12355

Pitt, J. I., Taniwaki, M. H., & Cole, M. B. (2013). Mycotoxin production in major crops as influenced by growing, harvesting, storage and processing, with emphasis on the achievement of Food Safety Objectives. *Food Control, 32* (1), 205-215. https://doi.org/10.1016/j.foodcont.2012.11.023

Ratnasingham, S., & Hebert, P. D. N. (2007). bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes, 7* (3), 355-364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Ripp, F., Krombholz, C. F., Liu, Y., Weber, M., Schafer, A., Schmidt, B., . . . Hankeln, T. (2014). All-Food-Seq (AFS): a quantifiable screen for species in biological samples by deep DNA sequencing. *BMC Genomics, 15* (1), 639. https://doi.org/10.1186/1471-2164-15-639

Shi, K., Qu, L., Lin, X., Xie, Y., Tu, J., Liu, X., . . . Liu, Y. (2019). Deep-Fried Atractylodis Rhizoma Protects against Spleen Deficiency-Induced Diarrhea through Regulating Intestinal Inflammatory Response and Gut Microbiota. *International Journal of Molecular Sciences, 21* (1), 124. https://doi.org/10.3390/ijms21010124

Shi, L., Liu, J., Wei, M., Xie, L., & Song, J. (2018). DNA metabarcoding identification of prescription ingredients in traditional medicine Ruyi Jinhuang San. *Scientia Sinica Vitae, 48* (04), 490-497.

The national pharmacopoeia committee of China. (2020a).*Pharmacopoeia of the People's Republic of China Part IV* . Beijing: China Medical Science Press.

The national pharmacopoeia committee of China. (2020b).*Pharmacopoeia of the People's Republic of China Part I* . Beijing: China Medical Science Press.

Srivathsan, A., Sha, J. C. M., Vogler, A. P., & 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. https://doi.org/10.1111/1755-0998.12302

van der Heyde, M., Bunce, M., Wardell-Johnson, G., Fernandes, K., White, N. E., & Nevill, P. (2020). Testing multiple substrates for terrestrial biodiversity monitoring using environmental DNA metabarcoding. *Mol Ecol Resour, 20* (3). https://doi.org/10.1111/1755-0998.13148

Vangay, P., Johnson, A. J., Ward, T. L., Al-Ghalith, G. A., Shields-Cutler, R. R., Hillmann, B. M., . . . Knights, D. (2018). US Immigration Westernizes the Human Gut Microbiome. *Cell, 175* (4), 962-972.e910. https://doi.org/10.1016/j.cell.2018.10.029

Warden, J. G., Casaburi, G., Omelon, C. R., Bennett, P. C., Breecker, D. O., & Foster, J. S. (2016). Characterization of Microbial Mat Microbiomes in the Modern Thrombolite Ecosystem of Lake Clifton, Western Australia Using Shotgun Metagenomics. *Frontiers in Microbiology, 7* , 1064. https://doi.org/10.3389/fmicb.2016.01064

Wu, T. H., Chen, I. C., & Chen, L. C. (2010). Antacid effects of Chinese herbal prescriptions assessed by a modified artificial stomach model. *World journal of gastroenterology, 16* (35), 4455-4459. https://doi.org/10.3748/wjg.v16.i35.4455

Xin, T., Su, C., Lin, Y., Wang, S., Xu, Z., & Song, J. (2018a). Precise species detection of traditional Chinese patent medicine by shotgun metagenomic sequencing. *Phytomedicine, 47* , 40-47. https://doi.org/10.1016/j.phymed.2018.04.048

Xin, T., Xu, Z., Jia, J., Leon, C., Hu, S., Lin, Y., . . . Newmaster, S. G. (2018b). Biomonitoring for traditional herbal medicinal products using DNA metabarcoding and single molecule, real-time sequencing.*Acta Pharmaceutica Sinica B, 8* (3), 488-497. https://doi.org/10.1016/j.apsb.2017.10.001

Xiong, Y., & Yan, Z. (2016). *Pharmaceutical Botany* . Beijing: People's Medical Publishing House

Yang, Q., Xu, Y., Feng, G., Hu, C., Zhang, Y., Cheng, S., . . . Gong, X. (2014a). p38 MAPK signal pathway involved in anti-inflammatory effect of Chaihu-Shugan-San and Shen-ling-bai-zhu-San on hepatocyte in non-alcoholic steatohepatitis rats. *African journal of traditional, complementary, and alternative medicines : AJTCAM, 11* (1), 213-221. https://doi.org/10.4314/ajtcam.v11i1.34

Yang, Q., Xu, Y., Liu, Y., Liang, Y., Feng, G., Zhang, Y., . . . Li, Y. (2014b). Effects of Chaihu-Shugan-San and Shen-Ling-Bai-Zhu-San on p38 MAPK Pathway in Kupffer Cells of Nonalcoholic Steatohepatitis.*Evidence-based complementary and alternative medicine : eCAM, 2014* , 671013. https://doi.org/10.1155/2014/671013

Yao, Q., Zhu, X., Han, M., Chen, C., Li, W., Bai, H., & Ning, K. (2020). Decoding herbal materials of representative TCM preparations with the multi-barcoding approach. *bioRxiv* , 2020.2006.2029.177188. https://doi.org/10.1101/2020.06.29.177188

Yap, K. Y.-L., Sui, Y., & Chu, S. (2008). The reliability of traditional authentication - A case of ginseng misfit. *Food Chemistry, 107* (1), 570-575. https://doi.org/10.1016/j.foodchem.2007.07.055

Zhang, S., Lin, L., Liu, W., Zou, B., Cai, Y., Liu, D., . . . Xie, Z. (2019). Shen-Ling-Bai-Zhu-San alleviates functional dyspepsia in rats and modulates the composition of the gut microbiota. *Nutrition research, 71* , 89-99. https://doi.org/10.1016/j.nutres.2019.10.001

7. Data Accessibility

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 SAMN22814863, SAMN22814864, SAMN22814865, SAMN22814866, SAMN22814867, and SAMN22814868. And the SRA accession numbers for the above five BioSample specimens are SRR16702785, SRR16702784, SRR16702776, SRR16702775, SRR16702774, SRR16702773, 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 MZ491209, MZ540179, MZ556835, MZ540195, MZ540186, MZ491221, MZ540178, MZ491214, MZ540184, MZ540177, MZ540199, MZ491215, MZ540185, MZ540200, MZ491212, MZ540198, MZ540193, MZ491219, MZ540190, MZ540204, MZ491211, MZ540181, MZ556837, MZ540197, MT102865, MT994329, MW000341, and MW000333.

8. Author Contributions

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

Tables and Figures

Table 1. The GenBank accession numbers of the ten herbal materials in SLBZS and the positive control, Panacis Quinquefolii Radix.

Herb medicinal material	ITS2	*psbA-trnH*	*matK*	*rbcL*
Ginseng Radix et Rhizoma	MZ491209	MZ540179	MZ556835	MZ540195
Poria	EJ501579	/	/	/
Dioscoreae Rhizoma	—	MZ540186	—	GQ436671
Atractylodis Macrocephalae Rhizoma	MZ491221	MZ540178	—	MN185070.1
Nelumbinis Semen	MZ491214	MZ540184	MZ540177	MZ540199
Lablab Semen Album	MZ491215	MZ540185	—	MZ540200

Herb medicinal material	ITS2	*psbA-trnH*	*matK*	*rbcL*
Amomi Fructus	MZ491212	KJ151870.1	KJ151842.1	MZ540198
Coicis Semen	KC181920	MZ540193	—	GQ436384.1
Glycyrrhizae Radix et Rhizoma	MZ491219	MZ540190	AB280741	MZ540204
Platycodonis Radix	MZ491211	MZ540181	MZ556837	MZ540197
Panacis Quinquefolii Radix	MT102865	MT994329	MW000341	MW000333

Note: "—" Indicates that the sequence was not successfully obtained. "/" indicates that the species does not contain chloroplasts and does not have chloroplast sequences. The reliable reference sequences for the accession numbers EJ501579, KC181920, KJ151870.1, KJ151842.1, AB280741, GQ436671, MN185070.1, and GQ436384.1 were downloaded from GenBank.

Table 2. Analysis results of shotgun sequencing data for ITS2,*psbA-trnH* , *matK* , and *rbcL*

	ITS2	*psbA-trnH*	*matK*	*rbcL*
Number of contigs	460	10158	75	127
Number of DNA barcodes	139	49	44	45
number of OTUs	53	14	13	11
Average length (bp)	209	418.9	872.1	703
Average GC content (%)	55.8	31.1	32.3	42.5

Table 3. Species detection of the ITS2, *psbA-trnH* ,*matK* , and *rbcL* DNA barcoding regions obtained *via*shotgun metabarcoding in two mock SLBZS samples

Species	HSZY159	HSZY159	HSZY159	HSZY159	HSZY171	HSZY171	HSZY171	HSZY
	ITS2	*psbA-trnH*	*matK*	*rbcL*	ITS2	*psbA-trnH*	*matK*	*rbcL*
Panax ginseng	[?]	[?]	[?]	[?]	[?]	[?]	[?]	[?]
Wolfiporia cocos	[?]	/	/	/	[?]	/	/	/
Dioscorea oppositifolia	[?]	[?]	[?]	[?]	[?]	[?]	[?]	[?]
Atractylodes macrocephala	[?]	—	—	—	[?]	—	—	—
Nelumbo nucifera	[?]	—	—	[?]	[?]	[?]	[?]	[?]
Lablab purpureus	[?]	[?]	[?]	[?]	[?]	[?]	[?]	[?]
Wurfbainia villosa	[?]	[?]	[?]	[?]	[?]	[?]	[?]	[?]
Coix lacryma-jobi	[?]	—	—	—	—	—	—	—
Glycyrrhiza uralensis	[?]	[?]	[?]	[?]	[?]	[?]	[?]	[?]
Platycodon grandiflorus	[?]	[?]	[?]	[?]	[?]	[?]	[?]	[?]
Panax quinquefolius	/	/	/	/	[?]	[?]	[?]	[?]

Note: "[?]" shows that the corresponding assembly sequence of this species was obtained; "—" indicates that the corresponding assembly sequence of this species cannot be obtained; "/" indicates that the ingredient (*P. quinquefolius*) has not been added, or the species (*W. cocos*) does not contain chloroplasts and does not have chloroplast sequences.

Figure

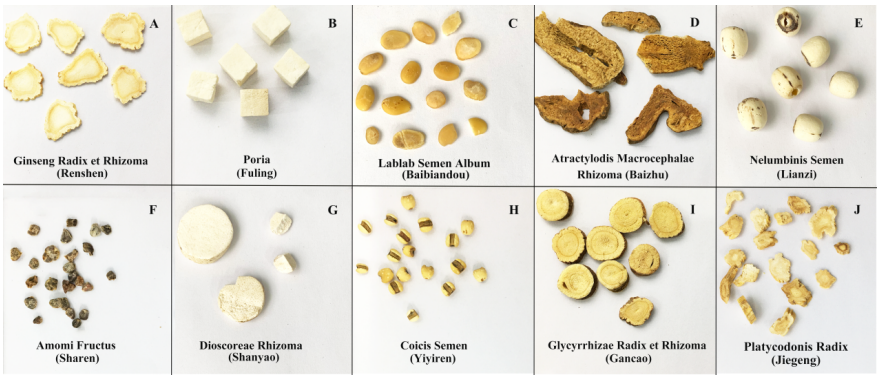

Figure 1. Morphological characteristics of ten herbal materials contained in the prescription of SLBZS. (A) Ginseng Radix et Rhizoma, (B) Poria, (C) Lablab Semen Album (Baibiandou, the ripe seed of*Lablab purpureus*), (D) Atractylodis Macrocephalae Rhizoma, (E) Nelumbinis Semen (Lianzi, the well-ripe seed of *Nelumbo nucifera*), (F) Amomi Fructus (Sharen, The ripe fruit of*Wurfbainia villosa*), (G) Dioscoreae Rhizoma (Shanyao, the rhizome or the dried steamed rhizome (rhizophore) of *Dioscorea oppositifolia*), (H) Coicis Semen, (I) Glycyrrhizae Radix et Rhizoma, and (J) Platycodonis Radix.

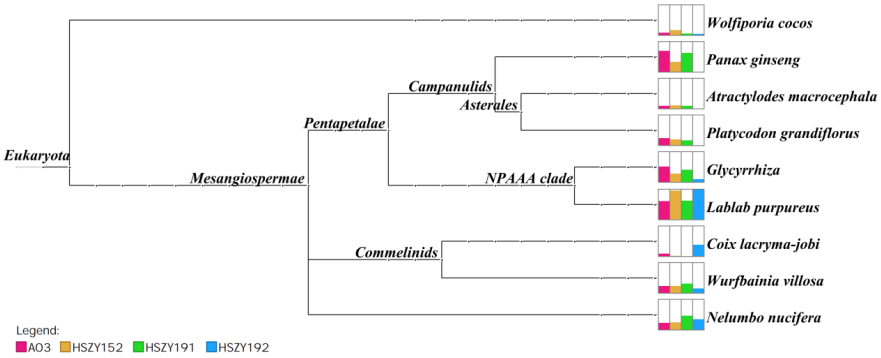

Figure 2. Taxonomic analysis of prescription ingredients in four samples detected *via* reads belonging to the ITS2 region. Each taxonomic node is drawn as a bar chart indicating the number of reads assigned to the taxon for each sample.

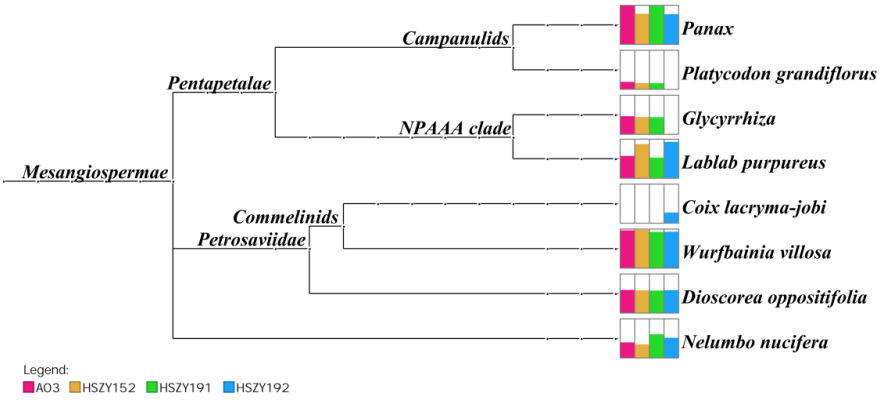

Figure 3. The taxonomic analysis of prescription ingredients in four samples detected by reads belonging to

the *psbA-trnH* region. Each taxonomic node is drawn as a bar chart indicating the number of reads assigned to the taxon for each sample.

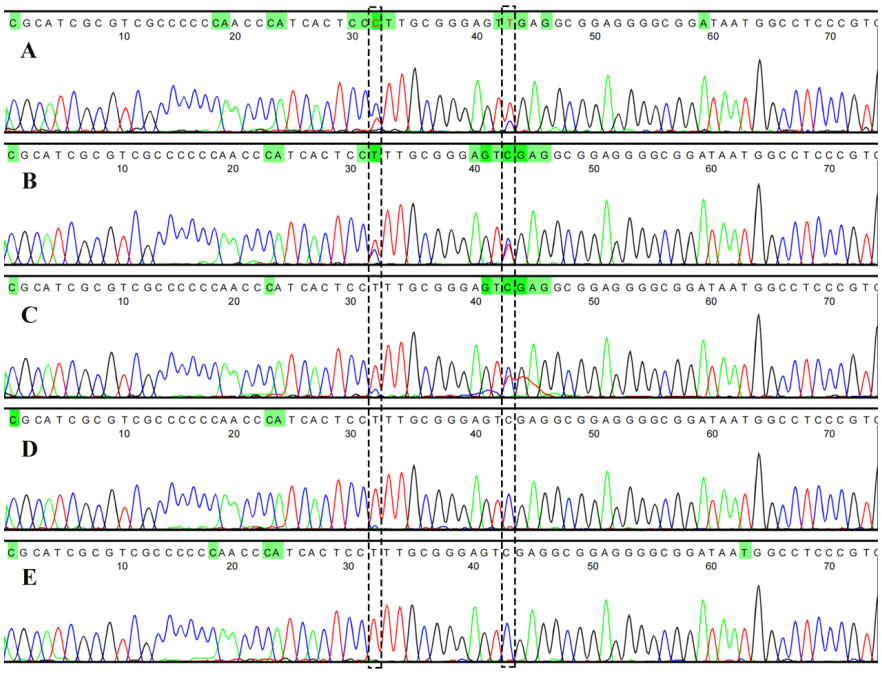

Figure 4. Trace data from PCR products by mixing powder of different proportions. A: *P. ginseng* and *P. quinquefolius*in the ratio of 9:1. B: *P. ginseng* and *P. quinquefolius* in the ratio of 7:3. C: *P. ginseng* and *P. quinquefolius* in the ratio of 5:5. D: *P. ginseng* and *P. quinquefolius* in the ratio of 3:7. E: *P. ginseng* and *P. quinquefolius* in the ratio of 1:9.

Supplemental Information

Supplementary Figure 1. Taxonomic analysis of prescription ingredients in the four samples detected *via* reads belonging to the *matK* region. Each taxonomic node is drawn as a bar chart, indicating the number of reads assigned to the taxon for each sample.

Supplementary Figure 2. Taxonomic analysis of prescription ingredients in the four samples detected *via* reads belonging to the *rbcL* region. Each taxonomic node is drawn as a bar chart indicating the number of reads assigned to the taxon for each sample.

Supplementary Figure 3. Trace data from PCR products by four pharmaceutical samples.

Supplementary Figure 4. Distribution of the fungi for each sample at the genus level.

Supplementary Figure 5. Heat map of the fungi genus identified in six samples analysed by shotgun metabarcoding. Yellow and blue colors indicate present and absent fungi genus, respectively.

Supplementary Figure 6. Mapping results of *matK* region in the HSZY171 sample with the whole genome of *A. macrocephala* as the reference sequence. (a) is the whole chloroplast genome sequence of*A. macrocephala* . (b) and (c) are the assembly sequences. (d) is the *matK* standard sequence of *A. macrocephala* .

Supplementary Table 1 Medicinal materials collected for made mock SLBZS samples.

Supplementary Table 2 Proportion of the five ingredients listed in the prescription of SLBZS.

Supplementary Table 3 Shotgun sequencing results of the two mock and four SLBZS pharmaceutical samples.

Supplementary Table 4. Reads number of the prescription ingredients in the four pharmaceutical samples based on the ITS2 sequences.

Supplementary Table 5. Reads number of the prescription ingredients in the four pharmaceutical samples based on the*psbA-trnH* sequences.

Supplementary Table 6. Reads number of the prescription ingredients in the four pharmaceutical samples based on the *matK*sequences.

Supplementary Table 7. Reads number of the prescription ingredients in the four pharmaceutical samples based on the *rbcL*sequences.

Supplementary Table 8. Reads number of the fungi in the six samples based on the ITS2 sequences.

Supplementary Table 9. Proportion of *P. ginseng* reads at the SNP positions in ITS2 in four pharmaceutical samples.