



A Comprehensive Quality Evaluation System for Complex Herbal Medicine Using PacBio Sequencing, PCR-Denaturing Gradient Gel Electrophoresis, and Several Chemical Approaches

Xiasheng Zheng^{1,2,3†}, Peng Zhang^{1,4†}, Baosheng Liao¹, Jing Li⁵, Xingyun Liu², Yuhua Shi¹, Jinle Cheng², Zhitian Lai², Jiang Xu^{1*} and Shilin Chen^{1*}

OPEN ACCESS

Edited by:

Fabio Marroni,
University of Udine, Italy

Reviewed by:

Hugo J. De Boer,
University of Oslo, Norway
Ai-dong Sun,
Beijing Forestry University, China

Ning Yu contributed to the review of
Ai-dong Sun

*Correspondence:

Jiang Xu
jxu@icmm.ac.cn
Shilin Chen
slchen@icmm.ac.cn

†These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Technical Advances in Plant Science,
a section of the journal
Frontiers in Plant Science

Received: 23 May 2017

Accepted: 29 August 2017

Published: 13 September 2017

Citation:

Zheng X, Zhang P, Liao B, Li J,
Liu X, Shi Y, Cheng J, Lai Z, Xu J
and Chen S (2017) A Comprehensive
Quality Evaluation System
for Complex Herbal Medicine Using
PacBio Sequencing, PCR-Denaturing
Gradient Gel Electrophoresis,
and Several Chemical Approaches.
Front. Plant Sci. 8:1578.
doi: 10.3389/fpls.2017.01578

¹ Key Laboratory of Beijing for Identification and Safety Evaluation of Chinese Medicine, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China, ² Key Laboratory of Technologies and Applications of Ultrafine Granular Powder of Herbal Medicine, State Administration of Traditional Chinese Medicine, Zhongshan Zhongzhi Pharmaceutical Group Limited, Zhongshan, China, ³ Guangdong Provincial Key Laboratory of New Drug Development and Research of Chinese Medicine, Guangzhou University of Chinese Medicine, Guangzhou, China, ⁴ School of Chinese Materia Medica, Beijing University of Chinese Medicine, Beijing, China, ⁵ Traditional Chinese Medicine Gynecology Laboratory in Lingnan Medical Research Center, Guangzhou University of Chinese Medicine, Guangzhou, China

Herbal medicine is a major component of complementary and alternative medicine, contributing significantly to the health of many people and communities. Quality control of herbal medicine is crucial to ensure that it is safe and sound for use. Here, we investigated a comprehensive quality evaluation system for a classic herbal medicine, Danggui Buxue Formula, by applying genetic-based and analytical chemistry approaches to authenticate and evaluate the quality of its samples. For authenticity, we successfully applied two novel technologies, third-generation sequencing and PCR-DGGE (denaturing gradient gel electrophoresis), to analyze the ingredient composition of the tested samples. For quality evaluation, we used high performance liquid chromatography assays to determine the content of chemical markers to help estimate the dosage relationship between its two raw materials, plant roots of Huangqi and Danggui. A series of surveys were then conducted against several exogenous contaminations, aiming to further access the efficacy and safety of the samples. In conclusion, the quality evaluation system demonstrated here can potentially address the authenticity, quality, and safety of herbal medicines, thus providing novel insight for enhancing their overall quality control.

Highlight: We established a comprehensive quality evaluation system for herbal medicine, by combining two genetic-based approaches third-generation sequencing and DGGE (denaturing gradient gel electrophoresis) with analytical chemistry approaches to achieve the authentication and quality connotation of the samples.

Keywords: quality evaluation system, herbal medicine, PacBio sequencing, PCR-DGGE, HPLC, Sanger sequencing, safety issues, Danggui Buxue Formula

INTRODUCTION

Complementary and alternative treatments for illnesses forms an important part of the modern medical system, which contributes to the healthcare and wellbeing of many human populations. Complementary and alternative medicines are often derived from plant, animal, and mineral materials, among which herbal medicines stand out as the form most widely used (WHO-WIPO-WTO BOOK¹). However, quality issues concerning herbal medicines such as adulterants from related and unrelated species (Heubl, 2010; Wang et al., 2015), poor quality resulting from improper harvesting times, and exogenous contaminations, namely heavy metals (Ernst, 2002) and pesticide (Chen et al., 2012) are routinely detected, which has the potential to threaten their clinical safety (Chan, 2003; Sakurai, 2011). Therefore, suitable and reliable evaluating approaches are now urgently needed to safeguard the quality of herbal medicines.

Species identification is crucial for guaranteeing the safety of herbal medicines. Traditional approaches of authentication for herbal medicines are based upon organoleptic characteristics and physical properties (Zhang, 2002), which depend heavily on subjective judgment that often leads to misjudgment (Huxtable, 1990). In the past decades, several chromatographic and spectroscopic methods have been applied to the authentication and evaluation of herbal medicines (Liang et al., 2004), including TLC (thin layer chromatography) (Mangold, 1961), HPLC (Majors, 1972), violet-visual spectrum (VVS), and near infrared spectrum (NIR) (Huck, 2015). However, these approaches remain non-specific, because the chemical profiles of herbal medicines are so similar to closely related plant species and could also be affected by physiological and storage conditions (Dhami and Mishra, 2015). Combining of HPLC with other, more advanced detectors, including mass spectrum (MS) (Horning et al., 1977) and nuclear magnetic resonance (NMR), offers a more accurate and specific solution than the above mentioned “non-specific” ones (Lindon et al., 2000). In recent years, HPLC has been serving as a powerful tool for determining the content of target chemical(s) and identifying the chemical profile of herbal medicines via a fingerprint analysis (Xie et al., 2006), rather than for identifying species *per se*.

At the same time, molecular approaches have become increasingly popular for authenticating herbal medicines and include both PCR-based and sequencing-based methods. The former has techniques, such as random amplified polymorphic DNA (RAPD) (Williams et al., 1990), restriction fragment length polymorphism (RFLP) (Botstein et al., 1980) and their variants [e.g., amplified fragment length polymorphism (AFLP; Han et al., 2007), inter-simple sequence repeat (ISSR; Choi et al., 2011)], quantitative PCR (qPCR) (Holland et al., 1991), loop-mediated isothermal amplification (LAMP) (Notomi et al., 2000;

Zhao et al., 2016) and high-resolution melting (HRM) (Wittwer et al., 2003), which are sensitive and fast. The sequencing-based approach involves DNA barcoding (Hebert et al., 2003) and high throughput sequencing (HTS) (Kircher and Kelso, 2010), which when coupled, offer detailed sequence information for authentication. DNA barcoding exhibits great potential for its high inter-specific and intra-specific discrimination ability. However, its application was often previously limited to samples consisting of a single species, and thus precluded complex mixtures such as a medicinal preparations, largely due to the inherent limitations of Sanger sequencing. Fortunately, with the available applicable markers ITS2, *trnL*, and 16s rRNA, HTS has been successfully used to determine the plant composition of herbal preparations at the species level (Coghlan et al., 2012, 2015; Cheng et al., 2014; Hawkins et al., 2015; Ivanova et al., 2016; Raclariu et al., 2017), demonstrating its advantages and reliability in the authentication of complex mixture samples, even if the contaminating species occur at a very low abundance. Compared with short-read sequencing technologies, such as the Ion Torrent (Margulies et al., 2005) and Illumina platforms (Kircher and Kelso, 2010), long-reads sequencing technologies, namely the single-molecule real-time (SMRT) sequencing used by Pacific Bioscience (PacBio), can generate longer reads of several kilobases with a unique circular template; this allows each template to be sequenced multiple times (Eid et al., 2009) to generate a CCS. Because of this, PacBio sequencing can be used for large structural and transcriptomic research projects (Goodwin et al., 2016). It also facilitates the direct sequencing of short DNA templates of metabarcoding.

Polymerase chain reaction-denaturing gradient gel electrophoresis (PCR-DGGE) has served as a powerful tool for detecting gene mutations (Forrest and Cotton, 1990). DNA separation in DGGE is based on the electrophoresis mobility of a partially melted DNA molecule in polyacrylamide gels containing a linearly increasing gradient of denaturants (Fischer and Lerman, 1983). Sequence variation within the melting domains of the DNA fragments causes migration differences in the denaturing gradient. PCR products containing a variety of amplicons from different DNA templates can be separated in DGGE because of differences in the sequence base pairings. PCR-DGGE became a common approach to explore the genetic diversity of complex microbial populations (Muyzer et al., 1993; Fukuda et al., 2016). As of yet, there are no reports of using PCR-DGGE to analyze the species composition of one or more complex herbal medicines.

With respect to the sequencing-based approaches and PCR-DGGE, it is crucial to select one or more proper biomarker(s) for species authentication. Among the candidate biomarkers, the internal transcribed spacer 2 (ITS2) has been validated with thousands of herbal species and since recommended as a standard DNA barcode (Chen et al., 2010). In addition, the chloroplast non-coding region *psbA-trnH* has been demonstrated to be reliable for species authentication of flowering plants (Kress and Erickson, 2007; Fazekas et al., 2008); it, too, was recommended as a complementary barcode to ITS2 (Chen et al., 2010). Coupling ITS2 and *psbA-trnH* regions, a species authentication system for the DNA barcoding of herbal materials

Abbreviations: C-7-G, calycosin-7-glucoside; CCS, circular consensus sequencing; DBF, Danggui Buxue Formula; DG, Danggui; DGGE, denaturing gradient gel electrophoresis; FA, ferulic acid; HPLC, high performance liquid chromatography; HQ, Huangqi; PacBio, Pacific Biosciences; PCR, polymerase chain reaction; TGS, third-generation sequencing.

¹https://www.wto.org/english/tratop_e/trips_e/trilatweb_e/ch2d_trilat_web_13_e.htm

was established, containing 78,847 sequences of 23,262 species, which altogether covers 95% of the raw herbs in pharmacopeia of China, United States, Japan, Korea, and India (Chen et al., 2014).

Nevertheless, emphasis has been laid on analyzing the biological composition of herbal medicines, yet little attention has been paid to the quantification of their components, which we believe to be an important yet underappreciated factor in quality control. This is because the dosage of the components is responsible, at least in part, for the pharmacological effects of a given herbal medicine. Further, the presence of exogenous contaminations, including those of heavy metal, pesticide, aflatoxins and microbes, etc., represent common safety issues for herbal medicines. Although invisible, these contaminants always leave a trace amount, which demands a series of detections by high-precision instruments.

In sum, the current quality control used for herbal medicines remains unsatisfactory and must be improved by establishing a thorough and robust quality evaluation system. Only by doing so can we better understand and access the quality connotation of herbal medicines. In this study, we report on a series of quality control experiments on a classic herbal medicine, DBF, including several authentication and quality evaluation approaches, with the aim of exploring a comprehensive quality evaluation system for complex herbal medicines. DBF contains 30 g of *Astragalus Radix* (i.e., dried roots of *Astragalus membranaceus*, Chinese name Huangqi, designated as HQ) and 6 g of *Angelica sinensis Radix* (i.e., dried roots of *Angelica sinensis*, Chinese name Danggui, designated as DG), mixed in a ratio of 5:1 by weight. First recorded in the “*Neiwaishang Bianhuo Lun*,” by Li Dongyuan in AD 1247, DBF has since been used clinically in China for centuries. According to modern clinical reports, DBF is a reliable prescription frequently used for promoting blood regeneration and enhancing immunity (Cui, 2005). Here, we first used the TGS and PCR-DGGE to analyze the ingredient composition of the DBF samples based on the ITS2 and *psbA-trnH* regions, while the Sanger sequencing of cloned-insert served as a validation to the results of TGS. Next, we used HPLC to determine the sample content of the chemical markers to help estimate the actual dosage relationship between the two key herbal components. Finally, a series of surveys were carried out against several exogenous contaminants, to further evaluate the safety of the DBF samples. All the three parts mentioned above contribute to the quality control system that we recommend for herbal medicine.

MATERIALS AND METHODS

Sample Collection

Three batches each of HQ and DG were manufactured and provided as ultrafine granular powder by the Zhongshan Zhongzhi Pharmaceutical Group (Zhongshan, China). To ensure their species accuracy, all of the independent samples were identified via DNA barcoding and Sanger sequencing. These samples were then ground into fine powder. The DBF samples were formulated in the laboratory by mixing together HQ (30 g) and DG (6 g) according to the prescription dosage, with three

replicates. For the TGS sequencing, a known authentic standard sample of DG served as the positive control; it was independently processed in the same way as the other DBF samples.

DNA Extraction and PCR Amplification

The DNA of the DBF samples was extracted by using the DNasecure Plant Kit (Tiangen Biotech, DP320, China). After elution, concentrations of the sample DNA were determined by a Nanodrop 2000 spectrophotometer (Thermo Scientific, United States), and then it was stored at -20°C . Individual PCR amplifications of the ITS2 and *psbA-trnH* regions were carried out in a 100 μL PCR system with the following final concentrations: $1 \times$ Taq PCR Mix solution (Aidlab, 271447AX, China), 1 μM of each primer, and ~ 100 ng of the sample DNA templates. Tagged primers, used for the preparation of the sequencing library, consisted of universal primers with 14-bp tags attached to the 5' end. The specially designed tags included 8-bp protected bases and 6-bp group-specific tags corresponding to the specific samples. For ITS2, the universal primers P3 and E4 (see **Supplementary Table S1**) were used with the following cycling conditions: 94°C for 5 min, then 40 cycles of (94°C for 30 s, 56°C for 30 s, 72°C for 45 s), and 72°C for 10 min. For *psbA-trnH*, the primers PA and TH were used with the following PCR conditions: 95°C for 4 min, then 35 cycles of (94°C for 30 s, 55°C for 1 min, 72°C for 1 min), and 72°C for 10 min. Double-distilled water served as the negative control, while the authentic DG sample served as the positive control. Electrophoresis of these PCR products was performed on 1% agarose gel. The positive PCR products were then purified by Agencourt[®] AMPure[®] XP beads (Beckman Coulter, 15824300, United States) by $0.8 \times$ volume, then washed by 70% ethanol twice and dissolved by double-distilled water. Concentrations of the purified PCR amplicons were determined by a Nanodrop 2000 spectrophotometer (Thermo Scientific, United States). Before the real PCR amplification, the primers used in this study were submitted to do the *in silico* test (Ye et al., 2012) on NCBI² using default parameters. Based on the thousands of BLAST hits, the results (see **Supplementary Table S2**) demonstrated that these primers can be used to amplify the target regions, both ITS2 and *psbA-trnH*, from a good number of plant species.

PacBio Circular Consensus Sequencing and Data Analysis

Purified PCR amplicons of the ITS2 and *psbA-trnH* regions of the DBF samples were mixed by volume to arrive at an equal concentration per sample. The sequencing library was constructed using this mixture with the SMRTbell[™] Template Prep Kit 1.0; then bound with V2 primers, by using the DNA/polymerase Binding Kit P6 V2 and P6-DNA polymerase; and transferred to a 96-well PCR plate for the real-time sequencing by using the C4 reagents with a PacBio RS II instrument. The resulting bas.h5 file was used for the subsequent data analysis. We used SMRT Analysis software (v4.0, Pacific Biosciences) to extract the CCS-fastq sequences from the bas.h5

²<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>

files. Then, the CCS reads with more than 6 (≥ 7) passes were filtered for the downstream analysis. Resulting reads belonging to different samples were grouped from the filtered CCS reads according to the sample-specific tags in the primer sequences. Next, we used the CodonCode Aligner (v5.1.5.3, CodonCode Corporation) to perform the sequences alignment to yield contigs with these parameters: end-to-end alignments, min. percent identity = 95.0, min. overlap length = 200, min. score = 150. Finally, the contigs and unassembled reads of each sample were used to perform *BLASTn* searches against two databases, the DNA Barcoding System for Identifying Herbal Medicine³ and the GenBank nucleotide NR database⁴, by using a max score ≥ 400 and Ident. $\geq 90\%$ (to define *BLAST* results as effective). These two databases were chosen because they contain most of the sequenced Chinese herbs and their adulterants (taxa species of both *Astragalus* and *Angelica* genus with ITS2 and *psbA-trnH* regions recorded in these two databases were listed in **Supplementary Table S3**). The PacBio CCS procedure was shown in **Figure 1A**.

Sanger Sequencing of Cloned Inserts

We performed a validation for the PacBio sequencing via the Sanger sequencing of cloned inserts. The vector pMD 19-T (Takara, 6013, China) was used to perform the T-cloning by ligating to the purified PCR amplicons, according to the manufacturer's manual. Subsequently, the whole reaction system was transferred into a competent cell of *Escherichia coli* (TransGen, CD201-02, China), while an empty vector was transferred as the negative control. The transgenic cell culture was spread onto a LB solid culture plate containing 50 ng/L Ampicillin, on the surface of which 40 μ L 2% X-gal and 7 μ L 20% IPTG had been spread previously. After overnight cultivation, independent white colonies were randomly selected from every plate to perform the PCR validation with the normal primers. Positive clones were grown in liquid culture and then sequenced with universal primers by the Sanger method.

PCR-DGGE Assay

The ITS2 fragments of the DBF samples were amplified using two universal primers, S2F and S3R, with a 40 bp GC clamp attached to the 5' end of the forward primer (see **Supplementary Table S1**). HQ and DG again served as the positive control for this method. The PCR system and amplifying procedure were consistent with that already described. The ensuring PCR products were directly subjected to a DGGE assay following the procedure described by Muyzer et al. (1993). We used a polyacrylamide gel of 6% concentration, which contained a 40–60% denaturing gradient. (The 100% denaturant solution contained 40% deionized formamide and 7 mol/L of urea.). Electrophoresis was performed on the Dcode universal mutation detection system (Bio-Rad, Milan, Italy) in 1 \times TAE at 60°C, 80 V, for 12 h. After the electrophoresis, the gel was stained with EB (Biotium, Fremont, United States) for 20 min, and viewed by UV transillumination. To further

confirm the DNA sequence information of the independent bands on the gel, a PCR amplification with the normal ITS2 primers (see **Supplementary Table S1**) was carried out using a small piece of each band's gel as the template. Then, a taxonomic identification was performed using *BLASTn* (as described above). The PCR-DGGE procedure was shown in **Figure 3A**.

HPLC Analysis

For the DBF samples, 0.2 g was extracted in 20 \times volumes of 70% methanol by supersonic bath for 30 min, with HQ and DG served as the positive controls. These extractions were then filtered through 0.45 μ m thick organic membranes and directly submitted to the HPLC analysis. To do this, we used Agilent equipment (Agilent, 1200 series, United States), with a C₁₈ column (Agilent, TC-C18, 250 mm \times 4.6 mm, 5 μ m) as the solid phase. The mobile phase contained 0.2% formic acid aqueous solution (A) and chromatographic pure acetonitrile (B), and following this procedure: 0–15 min, B: 5–45%; 15–60 min, B: 45–95%. An ultraviolet detector measured the ultraviolet absorption of the elution at the wavelengths of 254 and 316 nm, against HQ and DG, respectively. Authentic standard compounds, FA and C-7-G, were dissolved in 70% methanol and pure methanol, respectively, and diluted to a concentration of 10 μ g/mL. For each sample solution and standard solution, 10 μ L was injected for the HPLC assay under the conditions already described above.

Contamination Supervision

The determination of contamination residues was carried out according to the 2015 version of the Chinese Pharmacopoeia. Specially, sulfur dioxide residue was measured by the acid–base titration method (Cat.2331 in Chapter 4). Heavy metal residue was detected by the method of inductively coupled plasma mass spectrometry (Cat.2321). Pesticide residue was determined by the gas chromatography method (Cat.2341), while for aflatoxins residue the HPLC method was used (Cat.2351). Microbial limit was calculated by the cultivation method (Cat.1105 and 1106).

RESULTS

Ingredient Species Authentication of DBF

Here, we explored two novel molecular approaches to audit biological composition of the samples. However, the two

TABLE 1 | Sequencing results and statistics of the three-step data processing.

Job metric	Original results	After CCS (201–800 bp)	Further extracted by gene tags
Number of resulting reads	22,374	21,426	12,640
Total sample bases	9,348,991	9,180,147	5,633,144
Average read length (bp)	417	428	446

³<http://www.tcbarcode.cn/en/>

⁴<https://blast.ncbi.nlm.nih.gov>

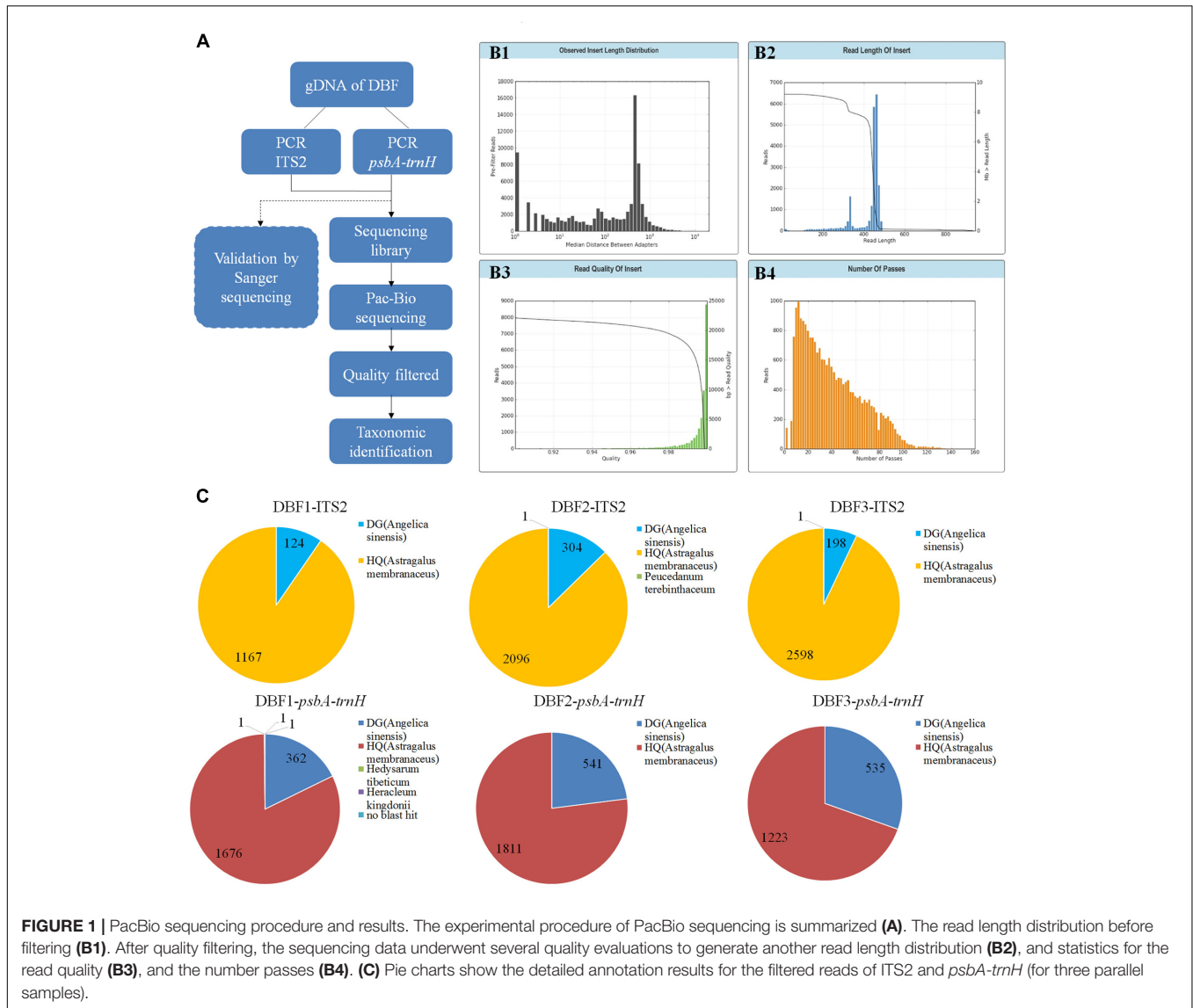


TABLE 2 | Results of the CCS and cluster analysis of the DBF samples with two biomarkers.

	ITS2			<i>psbA-trnH</i>			Total
	DBF1	DBF2	DBF3	DBF1	DBF2	DBF3	
CCS reads	1,291	2,401	2,797	2,041	2,352	1,758	12,640
Total sample bases (bp)	601,572	1,120,086	1,304,392	880,601	994,258	732,235	5,633,144
Average read length (bp)	466	467	466	431	423	417	446

approaches differed in their potential for identifying unknown ingredients in the herbal medicines.

PacBio CCS Analysis

At first, a total of 22,374 reads were obtained from the DNA library of DBF. The number was narrowed down to 21,426 reads after quality control, CCS processing, and gene length filtration (to 201–800 bp) (see B1–B4 in Figure 1. The sequencing data was deposited on the Sequence Read Archive (SRA) at

NCBI with an accession project number of PRJNA401343). Then, based on the group-specific tags introduced during the PCR amplification, the result reads for each of the six independent individual samples were grouped, yielding in 12,640 CCS reads (data generated by the three abovementioned steps are listed in Table 1) to taxonomic identification by *BLASTn*. All these resulting reads, for which the effective rate reached almost 100%, were successfully annotated and the identification of the parallel samples remained consistent. Of

the 6,489 CCS reads (amounting to 3,026,050 nucleotides) for the ITS2 fragments, there were 5,861 and 626 reads identified as HQ and DG, respectively. Likewise, of the 6,151 CCS reads (2,607,094 nucleotides) for *psbA-trnH* amplicons, 4,710 and 1,438 reads were identified as HQ and DG, respectively (Figure 1C and Table 2). The positive control showed the resulting reads of only DG (264 and 25 reads for ITS2 and *psbA-trnH*, respectively); whereas the negative control revealed neither electrophoretic band, nor sequencing library achieved. Further, a few unrelated species were detected in both biomarkers of the DBF samples, including two species of Leguminosae, two of Umbelliferae, and one of unknown (see Table 3). This result indicated that the PacBio CCS is sensitive enough to determine contaminated species that generally occur only at a very low abundance. These species were revealed by only one read each; that they occurred randomly in parallel sample indicates no biological contamination during the experimental procedures.

Sanger Sequencing of Cloned Inserts

T-A cloning and Sanger sequencing were performed to further validate the authentication result of TGS sequencing. Results showed that all sequences could be identified at the species level (Figure 2). A total of 26 and 4 ITS2 sequences were identified as HQ and DG, accounting for 86.7% and 13.3% of all positive sequences, respectively. Meanwhile, 20 and 10 *psbA-trnH* sequences were identified as HQ and DG, accounting for 66.7% and 33.3% out of all sequences, respectively.

Results of TGS sequencing and those of Sanger sequencing were very consistent, in that both HQ and DG were identified with the ITS2 and *psbA-trnH* regions. Nonetheless, the positively identified sequences of both sequencing methods gave different ratios between the two key ingredient species. This discrepancy points to a possible methodological preference in the DNA extraction and PCR amplification, which may cause difficulties in the quantitative measurement of different species' relative abundances.

PCR-DGGE Analysis

Internal transcribed spacer 2 fragments with a 40-bp GC clamp were successfully amplified from the three DBF samples and each ingredient sample. In the agarose electrophoresis, the GC-clamp ITS2 fragments of all samples had a similar migration due to the same gene size of c. 540 bp (Figure 3B). Interestingly, in the denaturing gradient gel, these amplicons of different species in the mixed sample exhibited a different migration behavior (Figure 3C). Two separate bands were observed in the DBF samples, which showed a similar migration pattern to that of their original species as positive controls. To further confirm the DNA sequences of those separated bands, a PCR amplification with the normal ITS2 primers was done by using a small piece of each band's gel as a template. The Sanger sequencing showed that Band 1 and Band 2 belonged to the ITS2 region of HQ and DG, respectively. Therefore, the PCR-DGGE offered a more visual and rapid ingredient analysis method for

TABLE 3 | Detected contaminated species in the DBF samples.

Biomarker	Sample	Number of reads	Umbelliferae		Leguminosae		Number of contaminants reads	Umbelliferae		Leguminosae		No blast hit
			DG (<i>Angelica sinensis</i>)	HQ (<i>Astragalus membranaceus</i>)	Peucedanum terebinthaceum	Meililotus officinalis		Hedysarum tibeticum	Heracleum kingdoni			
ITS2	DBF1	1,291	124	1,167	-	-	0	-	-	-	-	-
	DBF2	2,401	304	2,096	-	-	1	1	-	-	-	-
	DBF3	2,797	198	2,598	1	-	1	-	-	-	-	-
	Sum	6,489	626	5,861	2	3	0	1	1	1	1	1
<i>psbA-trnH</i>	DBF1	2,041	362	1,676	-	-	0	-	-	-	-	-
	DBF2	2,352	541	1,811	-	-	0	-	-	-	-	-
	DBF3	1,758	535	1,223	-	-	0	-	-	-	-	-
	Sum	6,151	1,438	4,710	3	3	0	1	1	1	1	1

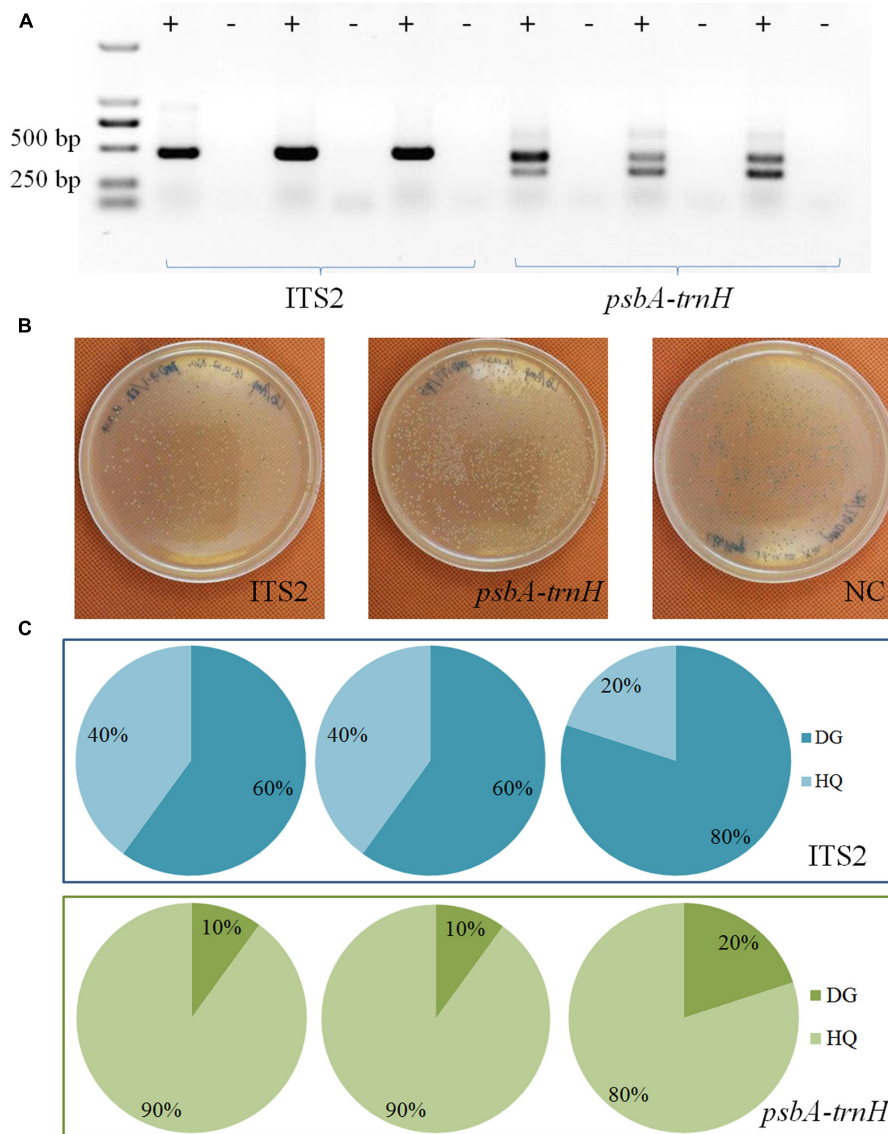


FIGURE 2 | Sanger sequencing of clone insert results. **(A)** The electrophoresis results of the PCR products. **(B)** The colonies on the plates. **(C)** The taxonomic identification results of the colonies by Sanger sequencing for the two biomarkers.

DBF than did either of the sequencing methods mentioned above.

Quality Evaluation of DBF

Determination of Chemical Markers by HPLC

Under the same gradient elution condition, DBF exhibited a similar absorbent curve to HQ at 254 nm (Figure 4A), while it was similar to DG at 316 nm (Figure 4B). Peak 1 and peak 1' were identified as the chemical markers C-7-G and FA, respectively, by comparing their retention time with that of authentic standards. The content of C-7-G in DBF was 0.0168%, but reached up to 0.0201% in HQ, whereas the FA content was 0.149% and 0.0892% in DBF and DG, respectively (as calculated by the standard curve method). To

investigate the ingredient content of raw materials, we used the peak areas of common peaks found between HQ and DBF, and DG and DBF, to calculate the relative content of the two main ingredients. Based on the multiple common peaks (Figure 4C), the results indicated that HQ accounted for 72–82% of the mixture, whereas DG made up 19–23% of it (Figure 4D). At the same time, when considering the chemical markers only, the ratio of HQ to DG content was 4:1 (Figure 4E), which was very close to the true value (5:1). The HPLC assay demonstrated that not only is it sensitive at determining the chemical marker(s), but it also provides feasibility to access a reasonable range for the ingredient dosage measurement of herbal medicines which contain raw materials.

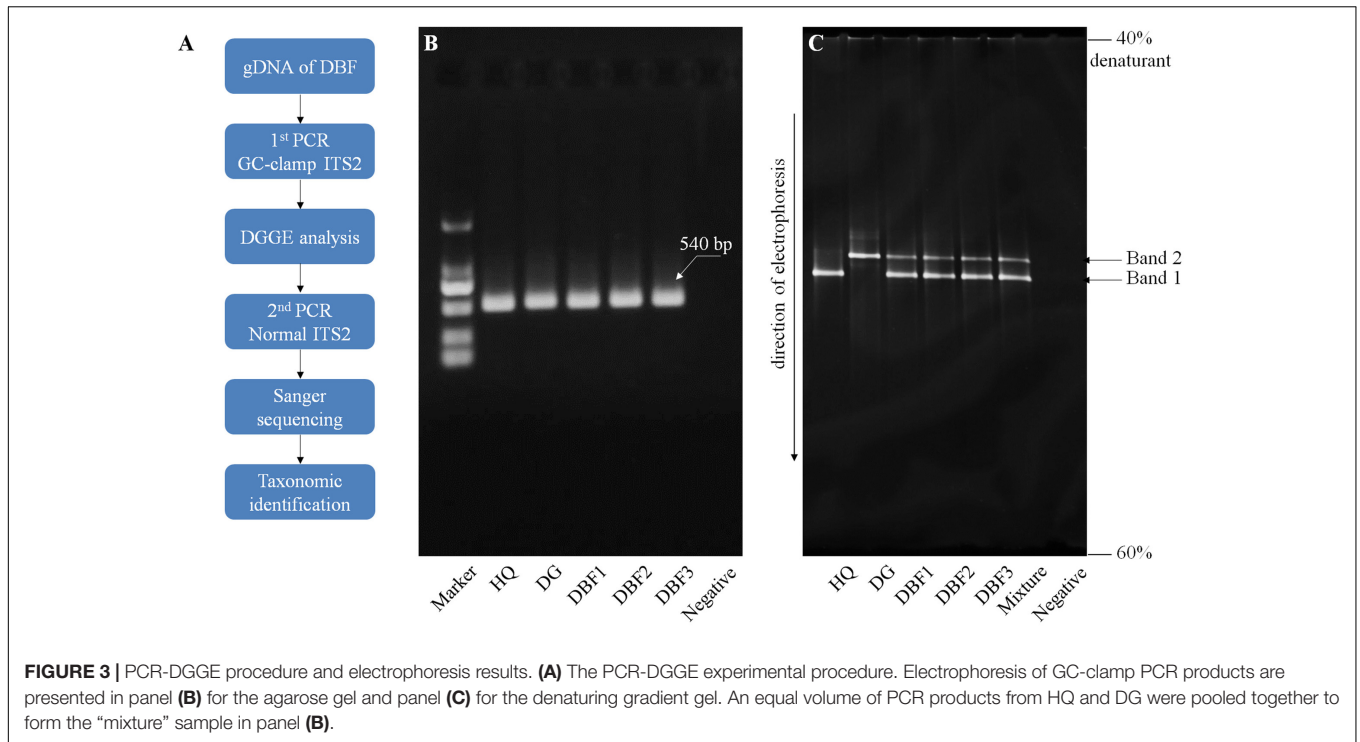


FIGURE 3 | PCR-DGGE procedure and electrophoresis results. **(A)** The PCR-DGGE experimental procedure. Electrophoresis of GC-clamp PCR products are presented in panel **(B)** for the agarose gel and panel **(C)** for the denaturing gradient gel. An equal volume of PCR products from HQ and DG were pooled together to form the “mixture” sample in panel **(B)**.

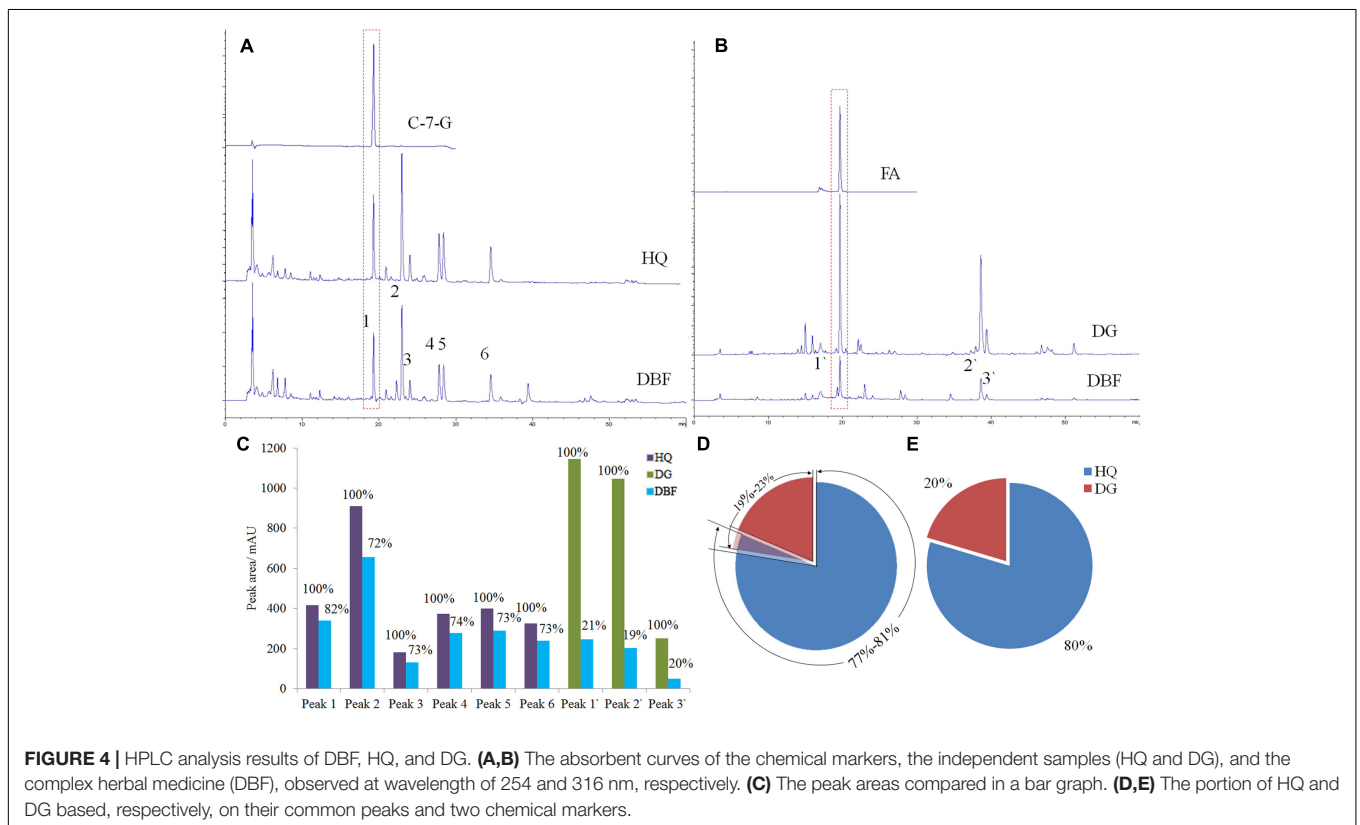


FIGURE 4 | HPLC analysis results of DBF, HQ, and DG. **(A,B)** The absorbent curves of the chemical markers, the independent samples (HQ and DG), and the complex herbal medicine (DBF), observed at wavelength of 254 and 316 nm, respectively. **(C)** The peak areas compared in a bar graph. **(D,E)** The portion of HQ and DG based, respectively, on their common peaks and two chemical markers.

Determination of DBF Contaminations

Quality control of contaminants includes those of sulfur dioxide, heavy metal, pesticide, aflatoxins, and harmful microbes. The

maximum allowable amounts of these contaminants should comply with the *Green Standards of Medicinal Plants and Preparations for Foreign Trade and Economy* and are adjusted

according to the latest version of Chinese Pharmacopeia. For the three batch samples of HQ and DG, their contamination residues (see **Supplementary Table S4** for results) were all below allowable limit, thus indicating these samples were safe for clinical use.

DISCUSSION

With the increasing popularity and demand for herbal medicines, much emphasis has fallen on their safety and quality control. Molecular identification contributes to the biological analysis of complex herbal medicines, thus forming an essential part of “herbal genomics,” which may deepen our understanding of the quality of herbal medicines (Chen et al., 2015). For investigating genetic diversity, Sanger sequencing is a tested and common approach to use (Chen et al., 2015). Yet, it turns out to be quite labor-intensive, time-consuming, and expensive when applied to the species identification of complex herbal medicines. Obviously the situation worsens as the number of ingredients in the herbal medicine increases. By contrast, HTS technology provides a much greater coverage of more samples in a shorter time period, and has better accuracy and repeatability (Jia et al., 2017). Further, introducing indexed tags enables the sequencing of multiple samples in parallel, thus reducing the analyzing time and cost of HTS (Parameswaran et al., 2007). The platform used in the present study, PacBio RS II, is capable of generating long reads and producing a greater sequencing throughput (Goodwin et al., 2016). Yet, a characteristic shortcoming of PacBio is the unsatisfactory sequencing error rate for single-pass sequencing (Schadt et al., 2010; Carneiro et al., 2012). This shortcoming, however, could be overcome by a sequencing strategy of multiple readings on the template DNA. Moreover, the high coverage allows for the differentiation of errors and SNPs (Goodwin et al., 2016). Building on this, we tested PCR-DGGE, a common tool for exploring the genetic diversity of complex microbial populations, to ascertain the species composition of a complex herbal medicine (DBF). Despite encouraging results, the feasibility of this method must be validated through more experiments. Between these two genetic approaches, the PCR-DGGE is suitable for typical labs, while PacBio CCS might be practical in institutions with extensive infrastructure, because it incurs a high cost in both equipment and sequencing, along with a rather sophisticated bioinformatics analyses.

A variety of approaches were developed for matching a query sequence against a reference database. These approaches are classified into four categories, including (1) tree-based methods, e.g., neighbor joining, maximum likelihood, *ATIM* and *SAP*; (2) similarity-based methods, e.g., nearest-neighbor (*NN*), *BLAST*, *TaxI*, *BRONX* and *jMOTU*; (3) character-based diagnostics, e.g., *BLOG* and *DNA-BAR*; and (4) statistical methods. Several studies were found reporting the comparison of these different analysis methods (Ross et al., 2008; Austerlitz et al., 2009; Virgilio et al., 2010; van Velzen et al., 2012). Yet, the results demonstrated that none of these methods can

consistently outperform the others for all data sets (Fan et al., 2014). The other methods, including *BLOG*, *DNA-BAR*, and *NN*, did not outperform significantly greater in identification success rate than *BLAST* (van Velzen et al., 2012). Further, most of the methods have difficulty in coping with large data sets except for *BLAST* (Fan et al., 2014). At the same time, both data bases used in our study are large data sets. The DNA barcoding system for identifying herbal medicine contained 78,847 sequences of 23,262 species, covering 95% of the raw herbs in pharmacopeia of China, United States, Japan, Korea, and India. Not to mention the larger one of GenBank non-redundant nucleotide database. Based on the above, we chose *BLAST* for species identification instead of other methods.

In this study, several contaminated species were detected for the commercial samples, while that no contaminants were found for the positive control sample. This indicated that those contaminated species were possibly introduced during the manufacturing process, rather than the experimental analyzing process. Those contaminated species, which share the same taxa family with both ingredients of DBF samples, are not commonly used medicinal herbs. Therefore, possible explanation for these detected contamination can be misidentification of raw material or unintentional introduction of pollutants during collection of raw material. Moreover, those contaminated species do not exhibit known toxicity, side effect and/or negative interaction, thus raising no safety risks to the usage of DBF samples in clinic. Nevertheless, monitoring the biological composition of herbal medicine is crucial for its safety.

Proposals to combine genetic approaches for species identification and chemical approaches for compound(s) determination have been put forth to better assess the quality of traditional medicines (Calixto, 2000). Quality evaluation of herbal medicines are based on analytical chemistry approaches for determining target constituents or chemical markers (Liang et al., 2009; Li et al., 2011). But it should be noted that safety issues such as contamination from endogenous or exogenous sources, are matter of concern ahead of the efficacy issues. In this study, we investigated a comprehensive quality evaluation system for herbal medicines, one that is concerned with not only robust species identification and chemical analysis, but also the oversight of several potential safety hazards. We hope that this study stimulates new thinking in how to better evaluate the quality of herbal medicines, which should help guarantee their safe and effective clinical use.

AUTHOR CONTRIBUTIONS

XZ and PZ contributed to all the researches activities associated with this work. XZ was responsible for writing the manuscript. JL and XL carried out the microscopic and HPLC assays, and also analyzed the data. BL and YS helped with the data analysis of the PacBio and Sanger sequencing. JC and ZL provided the herbal samples and were responsible

for determining the extent of contamination in them. JX and SC conceived the study, participated in its design and coordination, and helped to draft the manuscript. All authors read and approved the final manuscript.

FUNDING

We sincerely acknowledge the generous financial grants from the Postdoctoral Science Foundation (No. 2016M602597) and the National Natural Science Foundation of China (Nos. 81403053 and 81603248).

REFERENCES

- Austerlitz, F., David, O., Schaeffer, B., Bleakley, K., Olteanu, M., Leblois, R., et al. (2009). DNA barcode analysis: a comparison of phylogenetic and statistical classification methods. *BMC Bioinformatics* 10(Suppl. 14):S10. doi: 10.1186/1471-2105-10-S14-S10
- Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32, 314–331.
- Calixto, J. B. (2000). Efficacy, safety, quality control, marketing and regulatory guidelines for herbal medicines (phytotherapeutic agents). *Braz. J. Med. Biol. Res.* 33, 179–189. doi: 10.1590/S0100-879X2000000200004
- Carneiro, M. O., Russ, C., Ross, M. G., Gabriel, S. B., Nusbaum, C., and DePristo, M. A. (2012). Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC Genomics* 13:375. doi: 10.1186/1471-2164-13-375
- Chan, K. (2003). Some aspects of toxic contaminants in herbal medicines. *Chemosphere* 52, 1361–1371. doi: 10.1016/S0045-6535(03)00471-5
- Chen, L., Song, F., Liu, Z., Zheng, Z., Xing, J., and Liu, S. (2012). Multi-residue method for fast determination of pesticide residues in plants used in traditional Chinese medicine by ultra-high-performance liquid chromatography coupled to tandem mass spectrometry. *J. Chromatogr. A* 1225, 132–140. doi: 10.1016/j.chroma.2011.12.071
- 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, 1237–1244. doi: 10.1016/j.biotechadv.2014.07.004
- Chen, S., Song, J., Sun, C., Xu, J., Zhu, Y., Rob, V., et al. (2015). Herbal genomics: examining the biology of traditional medicines. *Science* 347(Suppl. 6219), S27–S29.
- 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:e8613. doi: 10.1371/journal.pone.0008613
- Cheng, X., Su, X., Chen, X., Zhao, H., Bo, C., Xu, J., et al. (2014). Biological ingredient analysis of traditional Chinese medicine preparation based on high-throughput sequencing: the story for Liuwei Dihuang Wan. *Sci. Rep.* 4:5147. doi: 10.1038/srep05147
- Choi, H. I., Kim, N. H., Kim, J. H., Choi, B. S., Ahn, I. O., Lee, J. S., et al. (2011). Development of reproducible EST-derived SSR markers and assessment of genetic diversity in *Panax ginseng* cultivars and related species. *J. Ginseng Res.* 35, 399–412. doi: 10.5142/jgr.2011.35.4.399
- Coghlan, M. L., Haile, J., Houston, J., Murray, D. C., White, N. E., Moolhuijzen, P., et al. (2012). Deep sequencing of plant and animal DNA contained within traditional Chinese medicines reveals legality issues and health safety concerns. *PLOS Genet.* 8:e1002657. doi: 10.1371/journal.pgen.1002657
- 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
- Cui, X. (2005). *The Research of Curative Effect and Mechanism on Hematopoiesis Regulation by Dangguibuxuetang Formula*. Jinan: Shandong University of Chinese Medicine.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.01578/full#supplementary-material>

TABLE S1 | Primers used in this study.

TABLE S2 | Results of primers specificity testing via *in silico* approach.

TABLE S3 | Taxa species of both *Astragalus* and *Angelica* genus with ITS2 and *psbA-trnH* regions recorded in the two databases used in this study.

TABLE S4 | Determination results of contaminants.

- Dhami, N., and Mishra, A. D. (2015). Phytochemical variation: how to resolve the quality controversies of herbal medicinal products? *J. Herb. Med.* 5, 118–127. doi: 10.1016/j.hermed.2015.04.002
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., et al. (2009). Real-time DNA sequencing from single polymerase molecules. *Science* 323, 133–138. doi: 10.1126/science.1162986
- Ernst, E. (2002). Toxic heavy metals and undeclared drugs in Asian herbal medicines. *Trends Pharmacol. Sci.* 23, 136–139. doi: 10.1016/S0165-6147(00)01972-6
- Fan, L., Hui, J. H., Yu, Z. G., and Chu, K. H. (2014). VIP Barcoding: composition vector-based software for rapid species identification based on DNA barcoding. *Mol. Ecol. Resour.* 14, 871–881. doi: 10.1111/1755-0998.12235
- Fazekas, A. J., Burgess, K. S., Kesanakurti, P. R., Graham, S. W., Newmaster, S. G., Husband, B. C., et al. (2008). Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLOS ONE* 3:e2802. doi: 10.1371/journal.pone.0002802
- Fischer, S. G., and Lerman, L. S. (1983). DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory. *Proc. Natl. Acad. Sci. U.S.A.* 80, 1579–1583. doi: 10.1073/pnas.80.6.1579
- Forrest, S., and Cotton, R. G. (1990). Methods of detection of single base substitutions in clinical genetic practice. *Mol. Biol. Med.* 7, 451–459.
- Fukuda, K., Ogawa, M., Taniguchi, H., and Saito, M. (2016). Molecular approaches to studying microbial communities: targeting the 16S ribosomal RNA gene. *J. UOEH* 38, 223–232. doi: 10.7888/juoeh.38.223
- Goodwin, S., McPherson, J. D., and McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 17, 333–351. doi: 10.1038/nrg.2016.49
- Han, J., Zhang, W., Cao, H., Chen, S., and Wang, Y. (2007). Genetic diversity and biogeography of the traditional Chinese medicine, *Gardenia jasminoides*, based on AFLP markers. *Biochem. Syst. Ecol.* 35, 138–145. doi: 10.1016/j.bse.2006.05.021
- Hawkins, J., de Vere, N., Griffith, A., Ford, C. R., Allainguillaume, J., Hegarty, M. J., et al. (2015). Using DNA metabarcoding to identify the floral composition of honey: a new tool for investigating honey bee foraging preferences. *PLOS ONE* 10:e0134735. doi: 10.1371/journal.pone.0134735
- Hebert, P. D., Cywinska, A., Ball, S. L., and deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proc. Biol. Sci.* 270, 313–321. doi: 10.1098/rspb.2002.2218
- Heubl, G. (2010). New aspects of DNA-based authentication of Chinese medicinal plants by molecular biological techniques. *Planta Med.* 76, 1963–1974. doi: 10.1055/s-0030-1250519
- Holland, P. M., Abramson, R. D., Watson, R., and Gelfand, D. H. (1991). Detection of specific polymerase chain reaction product by utilizing the 5′–3′ exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 88, 7276–7280. doi: 10.1073/pnas.88.16.7276
- Horning, E. C., Carroll, D. I., Dzidic, I., Haegle, K. D., Lin, S., Oertli, C. U., et al. (1977). Development and use of analytical systems based on mass spectrometry. *Clin. Chem.* 23, 13–21.

- Huck, C. W. (2015). Advances of infrared spectroscopy in natural product research. *Phytochem. Lett.* 11, 384–393. doi: 10.1039/C6GC01482H
- Huxtable, R. J. (1990). The harmful potential of herbal and other plant products. *Drug Saf.* 5(Suppl. 1), 126–136. doi: 10.2165/00002018-199000051-00020
- Ivanova, N. V., Kuzmina, M. L., Braukmann, T. W., Borisenko, A. V., and Zakharov, E. V. (2016). Authentication of herbal supplements using next-generation sequencing. *PLOS ONE* 11:e0156426. doi: 10.1371/journal.pone.0156426
- 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
- Kircher, M., and Kelso, J. (2010). High-throughput DNA sequencing—concepts and limitations. *Bioessays* 32, 524–536. doi: 10.1002/bies.200900181
- Kress, W. J., and Erickson, D. L. (2007). A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLOS ONE* 2:e508. doi: 10.1371/journal.pone.0000508
- Li, S. P., Zhao, J., and Yang, B. (2011). Strategies for quality control of Chinese medicines. *J. Pharm. Biomed. Anal.* 55, 802–809. doi: 10.1016/j.jpba.2010.12.011
- Liang, X. M., Jin, Y., Wang, Y. P., Jin, G. W., Fu, Q., and Xiao, Y. S. (2009). Qualitative and quantitative analysis in quality control of traditional Chinese medicines. *J. Chromatogr. A* 1216, 2033–2044. doi: 10.1016/j.chroma.2008.07.026
- Liang, Y. Z., Xie, P., and Chan, K. (2004). Quality control of herbal medicines. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 812, 53–70. doi: 10.1016/S1570-0232(04)00676-2
- Lindon, J. C., Nicholson, J. K., and Wilson, I. D. (2000). Directly coupled HPLC-NMR and HPLC-NMR-MS in pharmaceutical research and development. *J. Chromatogr. B Biomed. Sci. Appl.* 748, 233–258. doi: 10.1016/S0378-4347(00)00320-0
- Majors, R. E. (1972). High-performance liquid chromatography on small particle silica gel. *Anal. Chem.* 44, 1722–1726. doi: 10.1021/ac60319a007
- Mangold, H. K. (1961). Thin-layer chromatography of lipids. *J. Am. Oil Chem. Soc.* 38, 708–727. doi: 10.1007/BF02633061
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380. doi: 10.1038/nature03959
- Muyzer, G., de Waal, E. C., and Uitterlinden, A. G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* 59, 695–700.
- Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., et al. (2000). Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res.* 28, E63. doi: 10.1093/nar/28.12.e63
- Parameswaran, P., Jalili, R., Tao, L., Shokralla, S., Gharizadeh, B., Ronaghi, M., et al. (2007). A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Res.* 35:e130. doi: 10.1093/nar/gkm760
- Raclariu, A. C., Paltinean, R., Vlase, L., Labarre, A., Manzanilla, V., Ichim, M. C., et al. (2017). Comparative authentication of *Hypericum perforatum* herbal products using DNA metabarcoding. TLC and HPLC-MS. *Sci. Rep.* 7, 1291. doi: 10.1038/s41598-017-01389-w
- Ross, H. A., Murugan, S., and Li, W. L. (2008). Testing the reliability of genetic methods of species identification via simulation. *Syst. Biol.* 57, 216–230. doi: 10.1080/10635150802032990
- Sakurai, M. (2011). Perspective: herbal dangers. *Nature* 480:S97. doi: 10.1038/480S97a
- Schadt, E. E., Turner, S., and Kasarskis, A. (2010). A window into third-generation sequencing. *Hum. Mol. Genet.* 19, R227–R240. doi: 10.1093/hmg/ddq416
- van Velzen, R., Weitschek, E., Felici, G., and Bakker, F. T. (2012). DNA barcoding of recently diverged species: relative performance of matching methods. *PLOS ONE* 7:e30490. doi: 10.1371/journal.pone.0030490
- Virgilio, M., Bäckeljau, T., Nevado, B., and De Meyer, M. (2010). Comparative performances of DNA barcoding across insect orders. *BMC Bioinformatics* 11:206. doi: 10.1186/1471-2105-11-206
- Wang, L., Kong, W., Yang, M., Han, J., and Chen, S. (2015). Safety issues and new rapid detection methods in traditional Chinese medicinal materials. *Acta Pharm. Sin. B* 5, 38–46. doi: 10.1016/j.apsb.2014.12.005
- Williams, J. G., Kubelik, A. R., Livak, K. J., Rafalski, J. A., and Tingey, S. V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* 18, 6531–6535. doi: 10.1093/nar/18.22.6531
- Wittwer, C. T., Reed, G. H., Gundry, C. N., Vanderstee, J. G., and Pryor, R. J. (2003). High-resolution genotyping by amplicon melting analysis using LCGreen. *Clin. Chem.* 49, 853–860. doi: 10.1373/49.6.853
- Xie, P., Chen, S., Liang, Y. Z., Wang, X., Tian, R., and Upton, R. (2006). Chromatographic fingerprint analysis—a rational approach for quality assessment of traditional Chinese herbal medicine. *J. Chromatogr. A* 1112, 171–180. doi: 10.1016/j.chroma.2005.12.091
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13:134. doi: 10.1186/1471-2105-13-134
- Zhang, G. (2002). *Identification of Chinese Materia Medica*, 1st Edn. Beijing: China Science Publishing.
- Zhao, M., Shi, Y., Wu, L., Guo, L., Liu, W., Xiong, C., et al. (2016). Rapid authentication of the precious herb saffron by loop-mediated isothermal amplification (LAMP) based on internal transcribed spacer 2 (ITS2) sequence. *Sci. Rep.* 6:25370. doi: 10.1038/srep25370

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

Copyright © 2017 Zheng, Zhang, Liao, Li, Liu, Shi, Cheng, Lai, Xu and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.