



## APPLICATION NOTE

# DNA Extraction Protocol for Biological Ingredient Analysis of Liuwei Dihuang Wan



Xinwei Cheng<sup>1,3</sup>, Xiaohua Chen<sup>2</sup>, Xiaoquan Su<sup>2</sup>, Huanxin Zhao<sup>1</sup>, Maozhen Han<sup>2</sup>, Cunpei Bo<sup>2</sup>, Jian Xu<sup>2</sup>, Hong Bai<sup>1,\*</sup>, Kang Ning<sup>2,\*</sup>

<sup>1</sup> Institute of Materia Medica, Shandong Academy of Medical Sciences, Jinan 250062, China

<sup>2</sup> Shandong Key Laboratory of Energy Genetics, CAS Key Laboratory of Biofuels and Bioinformatics Group of Single Cell Center, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao 266101, China

<sup>3</sup> School of Medicine and Life Sciences, University of Jinan, Shandong Academy of Medical Sciences, Jinan 250062, China

Received 25 February 2014; revised 29 March 2014; accepted 31 March 2014

Available online 14 May 2014

Handled by Jun Yu

## KEYWORDS

DNA extraction;  
Traditional Chinese medicine preparation;  
Biological ingredients;  
Liuwei Dihuang Wan

**Abstract** Traditional Chinese medicine (TCM) preparations are widely used for healthcare and clinical practice. So far, the methods commonly used for quality evaluation of TCM preparations mainly focused on chemical ingredients. The biological ingredient analysis of TCM preparations is also important because TCM preparations usually contain both plant and animal ingredients, which often include some mis-identified herbal materials, adulterants or even some biological contaminants. For biological ingredient analysis, the efficiency of DNA extraction is an important factor which might affect the accuracy and reliability of identification. The component complexity in TCM preparations is high, and DNA might be destroyed or degraded in different degrees after a series of processing procedures. Therefore, it is necessary to establish an effective protocol for DNA extraction from TCM preparations. In this study, we chose a classical TCM preparation, Liuwei Dihuang Wan (LDW), as an example to develop a TCM-specific DNA extraction method. An optimized cetyl trimethyl ammonium bromide (CTAB) method (TCM-CTAB) and three commonly-used extraction kits were tested for extraction of DNA from LDW samples. Experimental results indicated that DNA with the highest purity and concentration was obtained by using TCM-CTAB. To further evaluate the different extraction methods, amplification of the second internal transcribed spacer (ITS2) and the chloroplast genome *trnL* intron was carried out.

\* Corresponding authors.

E-mail: [baihong@gmail.com](mailto:baihong@gmail.com) (Bai H), [ningkang@qibebt.ac.cn](mailto:ningkang@qibebt.ac.cn) (Ning K).

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.



Production and hosting by Elsevier

The results have shown that PCR amplification was successful only with template of DNA extracted by using TCM-CTAB. Moreover, we performed high-throughput 454 sequencing using DNA extracted by TCM-CTAB. Data analysis showed that 3–4 out of 6 prescribed species were detected from LDW samples, while up to 5 contaminating species were detected, suggesting TCM-CTAB method could facilitate follow-up DNA-based examination of TCM preparations.

## Introduction

Traditional Chinese medicine (TCM) preparations are widely used for healthcare and clinical practice in China and many other Asian countries. They are usually made from medicinal plants, animal materials and minerals, containing hundreds of compounds [1,2]. So far, the commonly used methods for quality evaluation of TCM preparations have mainly focused on chemical ingredient analysis by various chromatographic and spectroscopic methods [3,4]. However, these targeted approaches can only measure the chemicals of interest, but often fail to identify the contaminating biological ingredients. Mis-identified herbal materials, adulterants or even some biological contaminants could be potentially included during the complex collection and manufacture procedures. Therefore, in addition to the analysis of chemical ingredients, the analysis of biological ingredients in TCM preparations is also important.

In recent years, DNA fingerprinting and sequencing technologies have been applied to the field of TCM quality evaluation [5–7]. Some studies on molecular authentication of TCM preparations have been reported recently [8–10]. In combination with the sensitive high-throughput sequencing (HTS) technology [11], all biological constituents, even those with trace amount, in TCM preparations could be potentially identified, including both prescribed and contaminating species [12]. DNA of different ingredients present in the preparations could be amplified and sequenced, whose abundance would be proportionally represented within the final DNA extract [13]. Therefore, the efficiency of DNA extraction is critical, which might affect the accuracy and reliability of ingredient identification.

The sequencing-based method has been used in a number of herb identification instances, and methods for extracting DNA from fresh plants have been intensively investigated [14–17]. However, there is still lack of studies exploring efficient protocols for DNA extraction from TCM preparations. On the one hand, compared to fresh herbal plants, TCM preparations are composed of various decoction pieces and some excipients, thus extraction methods with universal applicability would be desired. On the other hand, after a series of processing procedures such as drying and stewing, DNA would be destroyed or degraded to different extents. Therefore, an improved DNA extraction method should be developed to extract DNA with high sensitivities and accuracies for the identification of both prescribed and contaminating species in TCM preparations. As a result, the following amplification, sequencing and data analysis could reliably reflect the biological ingredients in TCM preparations.

In this study, we chose a classical TCM preparation, Liuwei Dihuang Wan (LDW), as an example in order to develop a TCM-specific DNA extraction method. LDW comprises six traditional herbs including *Rehmannia glutinosa* Libosch., *Cornus officinalis* Sieb. et Zucc., *Paeonia suffruticosa* Andr.,

*Dioscorea opposita* Thunb., *Poria cocos* (Schw.) Wolf and *Alisma orientalis* (Sam.) Juzep. Among them, *R. glutinosa* and *C. officinalis* are steam processed while the others are raw materials. All herbal materials are crushed to powder, mixed and made into pills together with honey or water. With each of its components having distinct properties (processed vs. unprocessed, plant vs. fungi), LDW could serve as a representative TCM preparation for the assessment of DNA extraction method.

## Results and discussion

### Assessment of DNA extraction methods for LDW samples

We used LDW purchased in Qingdao, China (QD) and Singapore (SG), respectively, as examples for DNA extraction and selected four DNA extraction methods for performance comparison. These included three commonly-used commercial DNA extraction kits: OMEGA E.Z.N.A. HP Plant DNA Kit (OMEGA Kit), MOBIO PowerSoil DNA Isolation Kit (MOBIO Kit) and Chinese Herbal Medicine Kit (Henan Huier Nano Technology, Huier Kit), as well as our newly-developed TCM-cetyl trimethyl ammonium bromide (CTAB) method. OMEGA Kit is designed for fresh, dry or frozen specimen; MOBIO Kit is used for environmental samples; whereas Huier Kit is a universal extraction kit for Chinese medicinal materials. The TCM-CTAB method was reported for the first time in this work. TCM-CTAB was optimized from the CTAB method [18], which is a manual method widely used for DNA extraction from different types of samples. We combined the CTAB method with sodium dodecyl sulfate (SDS) method [19] to remove polysaccharides effectively [20]. In addition, the lysis duration was prolonged to increase DNA yield, whereas phenol–chloroform–isoamyl alcohol extraction was performed twice to remove any possible remaining proteins and chloroform–isoamyl alcohol extraction was performed to remove residual phenol [21], which might interfere with PCR amplification.

The yield and quality of DNA extracted were determined spectrometrically. It was shown that the purity of DNA extracted by the TCM-CTAB method was high for both QD and SG samples, with an  $A_{260}/A_{280}$  ratio in the optimal range (1.6–1.8) (Table 1). However,  $A_{260}/A_{280}$  ratio of DNA extracted by OMEGA, MOBIO and Huier kits was lower than 1.6, suggesting possible contaminants like polysaccharides, polyphenols, etc. In addition, different DNA yield was achieved when using four extraction protocols. With the TCM-CTAB method, the DNA concentrations of samples QD and SG were  $589.0 \pm 61.2$  and  $691.0 \pm 54.2$  ng/ $\mu$ L, respectively, whereas the extraction rates were  $70.7 \pm 7.3$  and  $82.9 \pm 6.3$   $\mu$ g/g, respectively. Compared to the results of the other three kits (Table 1), it was obvious that the best performance in terms of both DNA concentration and extraction rate was achieved using TCM-CTAB. Furthermore, gel

**Table 1** Purity and concentration of DNA extracted from LDW samples with different extraction methods

| Extraction method | Sample | $A_{260}/A_{280}$ ratio | DNA concentration (ng/ $\mu$ L) | Extraction rate ( $\mu$ g/g) |
|-------------------|--------|-------------------------|---------------------------------|------------------------------|
| MOBIO             | QD     | 1.27 $\pm$ 0.08         | 30.0 $\pm$ 6.3                  | 3.6 $\pm$ 0.8                |
|                   | SG     | 1.14 $\pm$ 0.03         | 40.5 $\pm$ 11.7                 | 4.86 $\pm$ 1.4               |
| OMEGA             | QD     | 1.06 $\pm$ 0.12         | 47.4 $\pm$ 13.6                 | 28.4 $\pm$ 8.1               |
|                   | SG     | 0.79 $\pm$ 0.13         | 60.5 $\pm$ 22.3                 | 36.3 $\pm$ 13.4              |
| Huier             | QD     | 0.82 $\pm$ 0.04         | 109.0 $\pm$ 13.9                | 32.7 $\pm$ 4.2               |
|                   | SG     | 1.43 $\pm$ 0.01         | 145.1 $\pm$ 32.7                | 43.5 $\pm$ 9.8               |
| TCM-CTAB          | QD     | 1.67 $\pm$ 0.15         | 589.0 $\pm$ 61.2                | 70.7 $\pm$ 7.3               |
|                   | SG     | 1.73 $\pm$ 0.012        | 691.0 $\pm$ 54.2                | 82.9 $\pm$ 6.3               |

Note: Data were obtained based on 3 replicates for each setting.

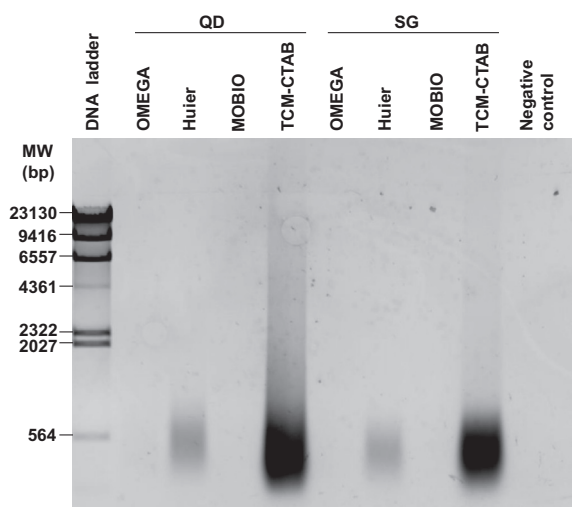
electrophoresis analysis (Figure 1) showed that obvious bands were only detected for DNA extracted using TCM-CTAB and Huier Kit. Moreover, the band intensity of DNA extracted using TCM-CTAB was higher than that using Huier Kit, although both exhibited dispersal pattern. All these data indicated that the highest extraction efficiency and quality of DNA were obtained using the TCM-CTAB method. Commercial DNA extraction kits usually have defined reagents and protocols, making them difficult to modify according to the specific requirements of varied TCM preparations. On the contrary, when using the TCM-CTAB method, we can easily change the concentration and composition of extraction buffer. On top of that, the TCM-CTAB method is also cost effective, compared to other DNA extraction kits tested in this study.

There were many successful trials of genome extraction for fresh herbs, but few studies were reported for TCM preparations. While some ingredients in TCM preparations might remain unprocessed before mixing with others (such as *P. suffruticosa* in LDW), some other plant or animal tissues in TCM preparations (such as *R. glutinosa* in LDW) might be completely or partially destroyed during the processing procedures, which would lead to degraded genomes in various degrees. As a result, the amount and quality of DNA retained in TCM preparations would be much lower than those in the fresh

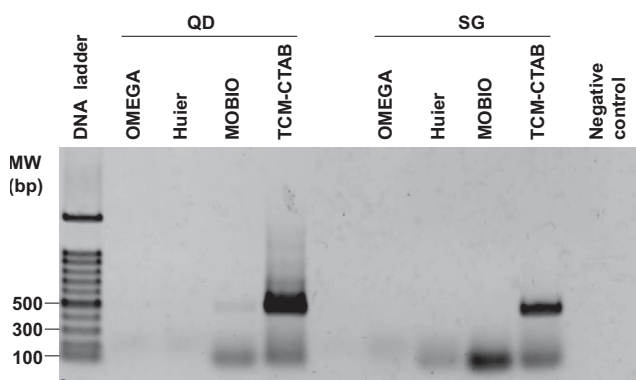
plants. Therefore, thorough lysis of TCM preparations to enrich DNA with high quality is very important. In addition, the complex components in TCM preparations, including many kinds of secondary metabolites such as phenolics and polysaccharides, might have negative effects on DNA purification [22,23]. Besides, different excipients were also used in TCM preparations. All these factors would reduce DNA purity, which might affect the following amplification and sequencing.

#### DNA amplification of LDW samples

Since serious DNA degradation was observed with gel electrophoresis, PCR amplification of the second internal transcribed spacer (ITS2) [24] and the chloroplast genome *trnL* (intron) with universal primers [25] was performed to evaluate the integrity of the extracted DNA samples. ITS2 (~500 bp) has been used as a standard molecular marker to identify medicinal plants for its high inter-specific and intra-specific discrimination power [26,27] while *trnL* (p-loop) (~200 bp) is a short fragment that can be easily amplified in heavily degraded DNA samples [28]. Therefore, ITS2 and *trnL* were chosen as the biomarkers for species discrimination of LDW. PCR without DNA template was served as a negative control. As shown in Figure 2, no band was revealed in the negative control, indicating no contamination from environment and reagents. Furthermore, when DNA extracted with the TCM-CTAB method was used as template, strong bands for ITS2 and *trnL* were

**Figure 1** DNA extraction from LDW samples with different extraction methods

PCR without DNA template was served as the negative control, which was provided at the right-most column. The DNA ladder is  $\lambda$ DNA /HindIII.

**Figure 2** PCR amplification of ITS2 using DNA extracted by different methods

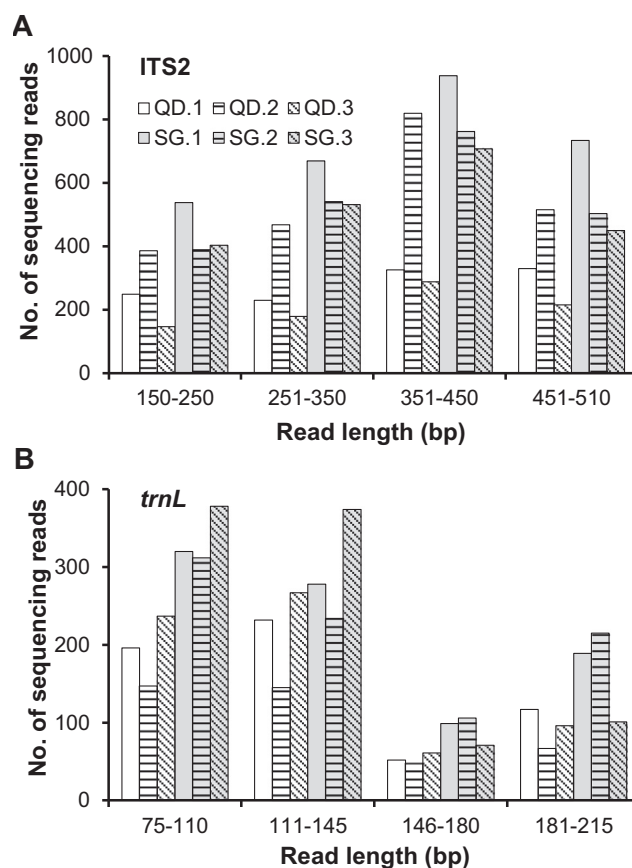
PCR without DNA template was served as the negative control, which was provided at the right-most column. The DNA ladder is 100 bp DNA ladder.

observed at 500 bp (Figure 2) and 200 bp for both the QD and SG samples. Using ITS2 as the biomarker, a weak band was detected with DNA extracted by Huier Kit as the template for QD samples but not for SG samples, whereas no bands were detected for any samples when using DNA isolated by OMEGA or MOBIO Kit as the template. Therefore, DNA extracted using the TCM-CTAB method was readily amplified compared to the other three kits. Given the highest quality and concentration of DNA recovered, as well as the satisfactory PCR results, the TCM-CTAB method was the most effective method for the DNA isolation of LDW among the four DNA extraction methods tested. We thus chose TCM-CTAB for the following sequencing experiments.

### High-throughput sequencing of LDW samples

The ITS2 and *trnL* amplification products were further subjected to high-throughput sequencing, so as to examine the actual biological ingredients of LDW samples. The high-throughput sequencing was performed by a 454 GS-Titanium sequencer with default setting, and an analysis of sequencing data was carried out for both QD and SG samples. After stringent filtering process for quality control (see Materials and methods), we obtained 4151 ITS2 and 2677 *trnL* reads (1384 ITS2 reads and 892 *trnL* reads per sample on average) for QD samples, while 7162 ITS2 and 1665 *trnL* reads were obtained for SG samples (2387 ITS2 reads and 555 *trnL* reads per sample on average) (Table 2). The length of ITS2 and *trnL* sequencing reads from QD samples varied from 151 bp to 493 bp (354 bp on average) and 75 bp to 208 bp (126 bp on average). Similar results were also obtained for SG samples. The read length was 150–502 bp (352 bp on average) for ITS2 and 75–211 bp (125 bp on average) for *trnL* (Table S1). The length distribution of the sequencing reads for all LDW samples was shown in Figure 3. Some variations were observed between samples for both ITS2 and *trnL* groups. As we primarily focused on the evaluation of DNA extraction methods for TCM preparations in this study, the details of biological and technical effects would be scrutinized in the future.

Sequencing reads were searched against the NCBI database by BLAST for species identification. As listed in Table 3 for ITS2 and Table 4 for *trnL*, *P. suffruticosa*, *D. opposita* and *A. orientalis* were detected from both QD and SG samples by both biomarkers, while *C. officinalis* (by *trnL*) and *R. glutinosa* (by ITS2 and *trnL*) were only detected in SG samples. We failed to identify *P. cocos* from any samples by either ITS2 or *trnL*. Failure to identify *P. cocos* using *trnL* can be easily explained by the fact that as a fungus, *P. cocos* does not possess chloroplast. However it is surprising that *P. cocos* was not detected using ITS2. We suspected that the universal ITS2 primers may not work for *P. cocos*. To test this possibility, multiple sequence alignment was performed for ITS2 reference sequences (from NCBI) for all 6 species in LDW around the



**Figure 3** The length distribution of sequencing reads for LDW samples

The length distributions for ITS2 and *trnL* are shown in panel A and panel B, respectively.

region where the forward primer S2F is located. It was shown that ITS2 S2F primer sequence was identical to the sequences of ITS2 from 5 plant species but not *P. cocos*. 4-bp mismatch existed between the primer sequence and the corresponding ITS2 sequence of *P. cocos* (Figure S1), which may underlie the failure to identify *P. cocos* by using universal ITS2 primers. Instead, our ongoing efforts indicated that *P. cocos* can be detected with a primer set specifically designed according to the *P. cocos* ITS2 sequences (unpublished data). These data suggested that ITS2 universal primers have limited applicability and caution should be taken when using these primers for the identification of some species. Moreover, *R. glutinosa* and *C. officinalis* were only identified from one SG sample. *R. glutinosa* and *C. officinalis* in LDW samples are processed by steaming and stewing, and these harsh processing procedures might lead to severe DNA degradation or loss. In addition, it is also possible that there might be differences between

**Table 2** Number of sequencing reads from LDW samples after quality control

| Sequencing library | QD   |      |     |       |         | SG   |      |      |       |         |
|--------------------|------|------|-----|-------|---------|------|------|------|-------|---------|
|                    | 1    | 2    | 3   | Total | Average | 1    | 2    | 3    | Total | Average |
| ITS2               | 1135 | 2188 | 828 | 4151  | 1384    | 2879 | 2193 | 2090 | 7162  | 2387    |
| <i>trnL</i>        | 886  | 867  | 924 | 2677  | 892     | 597  | 407  | 661  | 1665  | 555     |

**Table 3 Identification of biological ingredients in LDW samples based on ITS2 sequencing data**

| Ingredient                  | QD |   |   | SG |   |   |
|-----------------------------|----|---|---|----|---|---|
|                             | 1  | 2 | 3 | 1  | 2 | 3 |
| <i>Alisma orientalis</i>    | ✓  | ✓ | ✓ | ✓  | ✓ | ✓ |
| <i>Dioscorea opposita</i>   |    | ✓ |   |    |   | ✓ |
| <i>Paeonia suffruticosa</i> | ✓  | ✓ | ✓ | ✓  | ✓ | ✓ |
| <i>Rehmannia glutinosa</i>  |    |   |   | ✓  |   |   |
| <b>Apiaceae</b>             |    |   |   |    |   |   |
| – <i>Saposhnikovia</i>      |    |   |   | ✓  |   |   |
| <b>Asteraceae</b>           |    |   |   |    |   |   |
| – <i>Taraxacum</i>          |    |   |   |    |   | ✓ |
| <b>Fabaceae</b>             |    |   |   |    |   |   |
| – <i>Vigna</i>              |    |   |   | ✓  |   |   |
| <b>Polygonaceae</b>         |    |   |   |    |   |   |
| – <i>Persicaria</i>         |    |   |   |    |   | ✓ |
| – <i>Rumex</i>              |    |   |   |    |   | ✓ |

Note: Sequencing data were generated using the DNA extracted with TCM-CTAB method as template for ITS2 amplification. The words in bold represented the family-level taxonomical terms of contaminating species.

**Table 4 Identification of biological ingredients in LDW samples based on *trnL* sequencing data**

| Ingredient                  | QD |   |   | SG |   |   |
|-----------------------------|----|---|---|----|---|---|
|                             | 1  | 2 | 3 | 1  | 2 | 3 |
| <i>Alisma orientalis</i>    | ✓  | ✓ | ✓ | ✓  |   | ✓ |
| <i>Cornus officinalis</i>   |    |   |   |    | ✓ |   |
| <i>Dioscorea opposita</i>   | ✓  | ✓ | ✓ | ✓  | ✓ | ✓ |
| <i>Paeonia suffruticosa</i> | ✓  | ✓ | ✓ | ✓  | ✓ | ✓ |
| <i>Rehmannia glutinosa</i>  |    |   |   |    |   | ✓ |

Note: Sequencing data were generated using the DNA extracted with TCM-CTAB method as template for *trnL* amplification.

SG and QD samples in the source of medicinal plants or processing procedures. Without a clearly-defined standard for the processing procedures, DNA preservation from herbal materials and TCM preparations would be varied, which may lead to variations in detection.

Other than the prescribed species, 5 other plant genera including *Persicaria*, *Rumex*, *Vigna*, *Saposhnikovia* and *Taraxacum* were detected from SG samples, suggesting possible biological contaminations. We speculated that these plant species might be incorporated unintentionally during collection, processing or manufacturing steps. Biological contaminations would lower the efficacy of the TCM preparations or might cause adverse effects. Therefore, strict manufacturing process would be of great importance for the safety of TCM preparations.

## Conclusion

In conclusion, we tested four different DNA extraction methods for LDW samples and indicated that our newly-developed DNA extraction protocol TCM-CTAB can be used to extract DNA efficiently for the following detection of both prescribed and potential contaminating biological species in TCM preparations. Compared to three commonly-used DNA extraction kits (OMEGA Kit, MOBIO Kit and Huier Kit), the TCM-CTAB method can work more competently when tested on

LDW samples. By using the TCM-CTAB method, the DNA was recovered with desirable yield and purity. Then by means of PCR amplification and high-throughput sequencing technology, 4 prescribed species and 5 contaminating species were identified from LDW samples based on the ITS2 sequencing analysis, and 5 prescribed species were identified based on the *trnL* sequencing analysis. Therefore, the TCM-CTAB method can serve as an efficient DNA extraction procedure to achieve relatively complete identification of biological ingredients in TCM preparations.

## Materials and methods

### Sample collection

LDW was produced by Beijing Tong Ren Tang Co. Two sets of LDW were collected, with one purchased in Qingdao, China (QD) in April 2012 and the other purchased in Singapore (SG) in June 2012.

### DNA extraction and quantification

We extracted DNA from the two LDW samples (each with three replicates) using four protocols, including TCM-CTAB and three commercial DNA extraction kits, OMEGA E.Z.N.A. HP Plant DNA Kit (OMEGA Bio-Tek, Georgia,

USA), MOBIO PowerSoil DNA Isolation Kit (MOBIO, San Diego, USA) and Chinese Herbal Medicine Kit (Henan Huier Nano Technology, Henan, China). DNA was extracted from 50–250 mg LDW per sample as instructed by the manufacturers except TCM-CTAB and the resulting DNA was dissolved in 30  $\mu$ L TE buffer.

The TCM-CTAB method is an optimized CTAB method specialized for the extraction of DNA from biological ingredients of TCM samples. The procedures of the TCM-CTAB method were described as follows. Each sample (2.5 g) was homogenized with 5 mL buffer containing 0.1 M Tris-HCl (pH 8.0) and 20 mM EDTA (pH 8.0) with mortar and pestle. Afterward, 1 mL of the homogenate was transferred into 50 mL tubes and diluted with 4 mL of CTAB extraction buffer containing 100 mM Tris-HCl (pH 8.0), 20 mM EDTA (pH 8.0), 2% CTAB and 1.4 M NaCl, supplemented with 500  $\mu$ L 10% SDS, 10  $\mu$ L proteinase K (10 mg/mL) and 100  $\mu$ L  $\beta$ -mercaptoethanol. The mixture was incubated at 65 °C for 3 h with occasional vortexing. Polysaccharides and proteins were removed by extracting twice with phenol:chloroform:isoamyl-alcohol (25:24:1) and the residue phenol was removed by extracting once using chloroform:isoamyl-alcohol (24:1). The supernatant was then incubated at -20 °C with 0.6-fold of cold isopropanol for 1 h and then centrifuged at 12,500 rpm for 10 min. The resulting precipitate was rinsed with 1 mL 70% ethanol twice and then dissolved in 30  $\mu$ L TE buffer.

2  $\mu$ L of each DNA sample obtained with the four aforementioned methods was electrophoresed on 1% agarose gel and their concentration was quantified using NanoVue UV visible spectrophotometer.

### PCR amplification and sequencing of ITS2/*trnL*

Standard PCR amplification was performed with 50–200 ng DNA as template and the following primers (Sangon). The primers are S2F (5'-ATG CGA TAC TTG GTG TGA AT-3') and S3R (5'-GAC GCT TCT CCA GAC TAC AAT-3') with 7 bp multiplex identifier (MID) tags for ITS2 [24], and *trnL c* (5'-CGA AAT CGG TAG ACG CTA CG-3') and *trnL h* (5'-CCA TTG AGT CTC TGC ACC TAT C-3') with 7 bp MID tags for *trnL* [25]. The PCR products were electrophoresed on 2% agarose gel and purified with QIAquick Gel Extraction kit (QIAGEN). ITS2 and *trnL* amplicons were sequenced by a Roche 454 GS FLX Titanium sequencer with default parameters (XLR70 sequence kit, 2  $\times$  200 cycles).

### Sequencing data analysis

There were totally 25,137 ITS2 and 20,661 *trnL* raw sequencing reads obtained (Table S2) and a relatively strict quality control process was applied in order to eliminate false negative data as much as possible. We used Mothur software package [29,30] to discard sequencing reads that were shorter than 150 bp in ITS2 dataset and 75 bp in *trnL* datasets. Sequences with an average quality score < 20 in each 5-bp window rolling along the whole read were also discarded. Moreover, the sequencing reads containing primer mismatches, uncorrectable barcodes, ambiguous bases, or homopolymer runs in excess of 8 bp were also removed from both datasets. We also filtered the *trnL* sequences for which the possible corresponding species is evidenced (matched) by only 1 read, and ITS2 sequences

for which the possible corresponding species is evidenced by 3 or less reads, in order to avoid random sequence alignment bias.

Afterward, BLASTN [31] searches were performed (*E* value cutoff:  $10^{-10}$ ) for all query sequences against the NCBI database for identification. According to the BLASTN results, the majority of the top hits (96.57% of the matches for ITS2 reads and 97.01% of the matches for *trnL* reads) showed high identity (>98%) with the reference databases. Thus, we believe that top hits of BLASTN were adequate for identification of query sequences.

The 454 sequencing data for 6 LDW samples were deposited to NCBI SRA database with accession number SRR1049940.

### Authors' contributions

HB and KN conceived of and proposed the idea. XC, XC, HB and KN designed the study. XC, MH and CB performed the experiments. XC, XS and KN analyzed the sequencing data. XC, HB and KN drafted the manuscript. XC, XS, XC, HZ, JX, HB and KN edited and proof-read the manuscript. All authors read and approved the final manuscript.

### Competing interests

The authors declare that they have no competing interests.

### Acknowledgments

This work is partially supported by the grants from the National Natural Science Foundation of China (Grant Nos. 30870572, 61303161 and 61103167) and National High-tech R&D Program (863 Program) funded by Ministry of Science and Technology of China (Grant Nos. 2012AA02A707 and 2014AA021502) and the Key Laboratory for Rare and Uncommon Diseases of Shandong Province, China.

### Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.gpb.2014.03.002>.

### References

- [1] Li Y, Qu H, Cheng Y. Identification of major constituents in the traditional Chinese medicine "QI-SHEN-YI-QI" dropping pill by high-performance liquid chromatography coupled with diode array detection-electrospray ionization tandem mass spectrometry. *J Pharm Biomed Anal* 2008;47:407–12.
- [2] Li Y, Liu L, Cheng Y. Identification of components of traditional Chinese medicine preparation "Jing-Zhi-Guan-Xin" troche by medium pressure LC and HPLC-DAD-MS. *Chromatographia* 2007;65:749–55.
- [3] Li S, Zhao J, Yang B. Strategies for quality control of Chinese medicines. *J Pharm Biomed Anal* 2011;55:802–9.

- [4] Jing J, Parekh HS, Wei M, Ren W, Chen S. Advances in analytical technologies to evaluate the quality of traditional Chinese medicines. *TrAC Trends Anal Chem* 2013;44:39–45.
- [5] Joshi K, Chavan P, Warude D, Patwardhan B. Molecular markers in herbal drug technology. *Curr Sci* 2004;87:159–65.
- [6] Hon CC, Chow YC, Zeng FY, Leung FCC. Genetic authentication of ginseng and other traditional Chinese medicine. *Acta Pharmacol Sin* 2003;24:841–6.
- [7] Sucher NJ, Hennell JR, Carles MC. Genomic and transcriptomic profiling: tools for the quality production of plant-based medicines. In: Chandra S, Lata H, Varma A, editors. *Biotechnology for Medicinal Plants*. Berlin Heidelberg: Springer; 2013. p. 439–55.
- [8] Cao M, Wang J, Yao L, Xie S, Du J, Zhao X. Authentication of animal signatures in traditional Chinese medicine of Lingyang Qingfei Wan using routine molecular diagnostic assays. *Mol Biol Rep* 2014;41:2485–91.
- [9] Cui Z, Jiang C, Li M, Chen M, Zhou L, Yuan Y. Molecular identification of raw materials from Lianqiao Baidu Wan. *Acta Pharm Sin* 2013;48:590–6.
- [10] Newmaster SG, Grguric M, Shanmughanandhan D, Ramalingam S, Ragupathy S. DNA barcoding detects contamination and substitution in North American herbal products. *BMC Med* 2013;11:222.
- [11] Shokralla S, Spall JL, Gibson JF, Hajibabaei M. Next-generation sequencing technologies for environmental DNA research. *Mol Ecol* 2012;21:1794–805.
- [12] Coghlan ML, Haile J, Houston J, Murray DC, White NE, Moolhuijzen P, et al. Deep sequencing of plant and animal DNA contained within traditional Chinese medicines reveals legality issues and health safety concerns. *PLoS Genet* 2012;8:436–46.
- [13] Binladen J, Gilbert MTP, Bollback JP, Panitz F, Bendixen C, Nielsen R, et al. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One* 2007;2:e197.
- [14] Pirttilä AM, Hirsikorpi M, Kämäräinen T, Jaakola L, Hohtola A. DNA isolation methods for medicinal and aromatic plants. *Plant Mol Biol Rep* 2001;19:273.
- [15] Khan S, Qureshi MI, Kamaluddin TA, Abdin M. Protocol for isolation of genomic DNA from dry and fresh roots of medicinal plants suitable for RAPD and restriction digestion. *Afr J Biotechnol* 2007;6:175–8.
- [16] Azmat MA, Khan IA, Cheema HMN, Rajwana IA, Khan AS, Khan AA. Extraction of DNA suitable for PCR applications from mature leaves of *Mangifera indica* L. *J Zhejiang Univ Sci B* 2012;13:239–43.
- [17] Shepherd LD, McLay TGB. Two micro-scale protocols for the isolation of DNA from polysaccharide-rich plant tissue. *J Plant Res* 2011;124:311–4.
- [18] Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res* 1980;8:4321–5.
- [19] Möller EM, Bahnweg G, Sandermann H, Geiger HH. A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies, and infected plant tissues. *Nucleic Acids Res* 1992;20:6115.
- [20] Kim WK, Mauthe W, Hausner G, Klassen GR. Isolation of high molecular weight DNA and double-stranded RNAs from fungi. *Can J Bot* 1990;68:1898–902.
- [21] Zidani S, Ferchichi A, Chaieb M. Genomic DNA extraction method from pearl millet (*Pennisetum glaucum*) leaves. *Afr J Biotechnol* 2005;4:862–6.
- [22] Chaudhary K, Mathur N, Choudhary P, Pillai U. Protocol for isolation of genomic DNA from dry and fresh leaves of *Vigna* species suitable for RAPD and restriction digestion. *Adv Biol Res* 2008;2:5–6.
- [23] Porebski S, Bailey LG, Baum BR. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol Biol Rep* 1997;15:8–15.
- [24] Chen S, Yao H, Han J, Liu C, Song J, Shi L, et al. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS One* 2010;5:e8613.
- [25] Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, et al. Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding. *Nucleic Acids Res* 2007;35:e14.
- [26] Yao H, Song J, Liu C, Luo K, Han J, Li Y, et al. Use of ITS2 region as the universal DNA barcode for plants and animals. *PLoS One* 2010;5:e13102.
- [27] Gao T, Yao H, Song J, Liu C, Zhu Y, Ma X, et al. Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. *J Ethnopharmacol* 2010;130:116–21.
- [28] Valentini A, Miquel C, Nawaz MA, Bellemain EVA, Coissac E, Pompanon F, et al. New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the *trnL* approach. *Mol Ecol Resour* 2009;9:51–60.
- [29] Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009;75:7537–41.
- [30] Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS One* 2011;6:e27310.
- [31] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–402.