



BOKP: A DNA Barcode Reference Library for Monitoring Herbal Drugs in the Korean Pharmacopeia

Jinxin Liu^{1,2†}, Linchun Shi^{1,3†}, Jingyuan Song¹, Wei Sun³, Jianping Han¹, Xia Liu³, Dianyun Hou¹, Hui Yao¹, Mingyue Li¹ and Shilin Chen^{1,3*}

¹ Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, ² Hebei Key Laboratory of Study and Exploitation of Chinese Medicine, Chengde Medical College, Chengde, China, ³ Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China

OPEN ACCESS

Edited by:

Rudolf Bauer,
University of Graz, Austria

Reviewed by:

Michael Wink,
Universität Heidelberg, Germany
Angelica Cibrian-Jaramillo,
Centro de Investigación y de Estudios
Avanzados del Instituto Politécnico
Nacional (CINVESTAV-IPN), Mexico

*Correspondence:

Shilin Chen
sichen@implad.ac.cn

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Ethnopharmacology,
a section of the journal
Frontiers in Pharmacology

Received: 17 June 2017

Accepted: 11 December 2017

Published: 19 December 2017

Citation:

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

Herbal drug authentication is an important task in traditional medicine; however, it is challenged by the limitations of traditional authentication methods and the lack of trained experts. DNA barcoding is conspicuous in almost all areas of the biological sciences and has already been added to the British pharmacopeia and Chinese pharmacopeia for routine herbal drug authentication. However, DNA barcoding for the Korean pharmacopeia still requires significant improvements. Here, we present a DNA barcode reference library for herbal drugs in the Korean pharmacopeia and developed a species identification engine named KP-IDE to facilitate the adoption of this DNA reference library for the herbal drug authentication. Using taxonomy records, specimen records, sequence records, and reference records, KP-IDE can identify an unknown specimen. Currently, there are 6,777 taxonomy records, 1,054 specimen records, 30,744 sequence records (ITS2 and *psbA-trnH*) and 285 reference records. Moreover, 27 herbal drug materials were collected from the Seoul Yangnyeongsi herbal medicine market to give an example for real herbal drugs authentications. Our study demonstrates the prospects of the DNA barcode reference library for the Korean pharmacopeia and provides future directions for the use of DNA barcoding for authenticating herbal drugs listed in other modern pharmacopeias.

Keywords: DNA barcoding, reference library, identification engine, herbal drugs, Korean pharmacopeia

INTRODUCTION

Traditional Korean medicine (TKM), also referred to as traditional oriental medicine or Eastern medicine, originated during prehistoric times and has been widely used in Korea for thousands of years (Kim et al., 2005). TKM has been deeply influenced by traditional Chinese medicine since the periods of Baekje, Silla, and Goguryeo. TKM flourished during the period of Joseon in which the three Korean medical classics Hyangyak Jipseongbang, Uibang Yuchwi, and Donggeui Bogam were completed and published. Subsequently, TKM gradually began to accumulate distinctive features based on the distinctive environmental, cultural, social, and political situations in Korea (Cha et al., 2007). At the beginning of the nineteenth century, TKM adopted Chinese medical theories (such as yin yang and the five phases) and introduced a new frontier with its Sasang typology, which classified individuals into the Tae-Yang, So-Yang, Tae-Eum, and So-Eum types. The Sasang typology adopts the philosophy that an individual patient's medical treatment should be based on

the patient's biopsychological characteristics and his or her response to herbal drugs (Chae et al., 2012). The framework of the TKM system has been shaped by the Medicine for Citizens Act (1951) and paralleled Western medicine (Huang and Shin, 2012; Han S. Y. et al., 2016).

Herbal drugs are the foundation of TKM in medicinal practice, and the efficacy of TKM mostly depends on the quality of the herbal drugs (Choi et al., 2002). The Korean pharmacopeia is a statute for the standardization and quality control of herbal drugs. The latest edition of the Korean pharmacopeia was published in 2014 and includes 159 herbal drug monographs (MFDS, 2014). These herbal drugs are partially derived from natural materials, such as flowers, leaves, barks, fruits, seeds, stems, and roots. Herbal drugs on the pharmaceutical medicine market are usually used as cut herbal drugs (small pieces or small blocks) or powdered herbal drugs (coarse, medium, fine, or very fine powder). Since 1991, the Korean government has imposed strict regulations on Korean traditional herbal drugs. The World Health Organization (WHO) also provided guidelines for guaranteeing the safety and efficacy of traditional herbal medicines in 1991 (World Health Organization, 1991). However, the authentication of herbal drugs using traditional identification methods is very difficult because these herbals drugs often lack the key morphological diagnostic characteristics that are essential for species identification (Chen et al., 2014). Moreover, the number of herbal drugs and their preparations have greatly increased during the past several decades, and the domestic supply of herbal drugs is insufficient; therefore, a large portion of herbal drugs must be imported from China, Japan, and other countries. The increase in international herbal drug imports have led to an urgent demand for new powerful authentication methods that can achieve a rapid, accurate, and even automated identification of the herbal drugs.

DNA barcoding technology, which uses short standardized genetic markers to identify individual species, has significantly impacted taxonomic research (Savolainen et al., 2005; Miller, 2007; Valentini et al., 2009) and powerfully contributed to the regulatory authentication of medicinal plants, herbal drugs and certain herbal drug preparations. The challenging of plant DNA barcoding was to select a suitable marker or marker combination for discriminate a large range of plant taxa (Group et al., 2011; Hollingsworth et al., 2011) and several markers have been proposed successively (Kress et al., 2005; Kress and Erickson, 2007; Lahaye et al., 2008; Group et al., 2009). However, for medicinal plants, a specially designated plant group, Chen et al. proposed that nuclear internal transcribed spacer 2 (ITS2) and the chloroplast *psbA-trnH* intergenic region could serve as standard DNA barcodes for identifying medicinal plants and their closely related species (Chen et al., 2010). Similar results were obtained in subsequent experiments related to other plant groups and their closely related species (Group et al., 2011; Pang et al., 2011). In 2012, Coghlan et al. used high-throughput sequencing and DNA barcoding reference databases to detect the presence of toxic and endangered organic ingredients in highly processed herbal products presented in the form of powders, tablets, and capsules. Most of these toxic

and endangered organic ingredients were rarely declared by the product manufacturer (Coghlan et al., 2012). In 2013, Newmaster et al. demonstrated the presence of considerable herbal product substitutions and contaminations in most of the tested herbal products using a standard reference material herbal barcode library. Thus, the herbal industry should employ DNA barcoding to authenticate the raw materials used in the manufacturing of herbal products (Newmaster et al., 2013). In 2016, Han et al. indicated that ~4.2% of the raw herbal materials were adulterants in the 1,260 samples collected from pharmaceutical medicine markets (Han J. et al., 2016). The use of DNA barcoding technology could lead to a significant increase in the detection of herbal substitutions and contaminations, particularly after the establishment of a reference DNA barcode database (Chen et al., 2014).

MATERIALS AND METHODS

Specimen Collection

Our sampling strategy was as follows: for herbal drugs derived from a typical original plant and other species of the same genus, we collected the typical original plant and another plant that is often used instead of the original plant. For herbal drugs derived from a typical original plant and its varieties, we collected only the typical original plant. For the remaining herbal drugs derived from unambiguous species, we collected their specific origins according to the Korean pharmacopeia monographs. Both leaf samples and medicinal materials were collected from the original plants and pharmaceutical medicine markets, respectively.

To provide a good representation of the intraspecific genetic variability and the true situation in traditional herbal drug markets, we gathered 1,054 specimens to provide coverage for 208 species of 153 herbal drugs, including 530 medicinal materials and 524 original plant leaf materials. To provide an example for authenticating real herbal drugs, we collected 27 herbal drug materials from the Seoul Yangnyeongsi herbal medicine market, which is the largest herbal medicine market in Korea. The above-mentioned specimens are preserved at the Institute of Medicinal Plant Development, Chinese Academy of Medicinal Sciences. Detailed information regarding the data collection and the voucher information is provided in **Table S1** and **Table 1**.

Laboratory Protocols

Total genomic DNA was isolated from 20 to 30 mg of the leaf samples or 50~100 mg of the medicinal materials using a Hi-DNAsecure Plant Kit (Tiangen Biotech Co., Beijing, China). The PCR amplification was performed in a 25 μ L reaction mixture that contained 2 μ L genomic DNA, 12.5 μ L PCR MasterMix (Aidlab Biotechnologies Co., Beijing, China), 8.5 μ L ddH₂O, and 1 μ L each of the forward and reverse primers (2.5 μ M, synthesized by Sangon Co., China). The primers and the reaction conditions followed previously published methods for ITS2 and *psbA-trnH* (Chen et al., 2010; Group et al., 2011). The PCR products were purified using the QIAquick PCR purification kit (Tiangen Biotech, Beijing, China), and the bidirectional

TABLE 1 | Herbal drugs collected from Seoul Yangnyeongsi herbal medicine market and their identification result using KP-IDE software.

Specimen ID	Herbal Drug Name in Label	Medicinal Part	Identify Result
KSY001	Acanthopanax Root Bark	Bark	<i>Eleutherococcus sessiliflorus</i>
KSY002	Achyranthes Root	Root	<i>Achyranthes bidentata</i>
KSY003	Angelica Gigas Root	Root	<i>Angelica gigas</i>
KSY004	Apricot Kernel	Seed	<i>Prunus armeniaca</i>
KSY005	Aralia Continentalis Root	Root	<i>Aralia continentalis</i>
KSY006	Cimicifuga Rhizome	Rhizome	<i>Cimicifuga dahurica</i>
KSY007	Lithospermum Root	Root	<i>Lithospermum erythrorhizon</i>
KSY008	Citrus Unshiu Immature Peel	Pericarp	<i>Citrus reticulata</i>
KSY009	Citrus Unshiu Peel	Pericarp	<i>Citrus reticulata</i>
KSY010	Cnidium Rhizome	Rhizome	<i>Ligusticum sinense</i>
KSY011	Coptis Rhizome	Rhizome	<i>Coptis chinensis</i>
KSY012	Dioscorea Rhizome	Rhizome	<i>Dioscorea polystachya</i>
KSY013	Jujube	Fruit	<i>Zizyphus jujube</i>
KSY014	Kalopanax Bark	Bark	<i>Kalopanax pictus</i>
KSY015	Liriope Tuber	Tuber	<i>Ophiopogon japonicus</i>
KSY016	Lithospermum Root	Root	<i>Lithospermum erythrorhizon</i>
KSY017	Magnolia Bark	Bark	<i>Magnolia officinalis</i>
KSY018	Mentha Herb	Aerial Part	<i>Mentha arvensis</i> var. <i>piperascens</i>
KSY019	Ostericum Root	Root	<i>Notopterygium incisum</i>
KSY020	Perilla Leaf	Leaf and Twig	<i>Perilla frutescens</i>
KSY021	Perilla Leaf	Leaf and Twig	<i>Perilla frutescens</i>
KSY022	Polygonatum Rhizome	Rhizome	<i>Polygonatum sibiricum</i>
KSY023	Polygonatum Rhizome	Rhizome	<i>Polygonatum sibiricum</i>
KSY024	Poncirus Immature Fruit	Fruit	<i>Citrus trifoliata</i>
KSY025	Prunella Spike	Spike	<i>Prunella vulgaris</i>
KSY026	Rubus Fruit	Fruit	<i>Rubus crataegifolius</i>
KSY027	Dioscorea Rhizome	Rhizome	<i>Dioscorea polystachya</i>

sequencing was accomplished with an ABI 3730XL sequencer using the original amplification primer.

Data Analysis

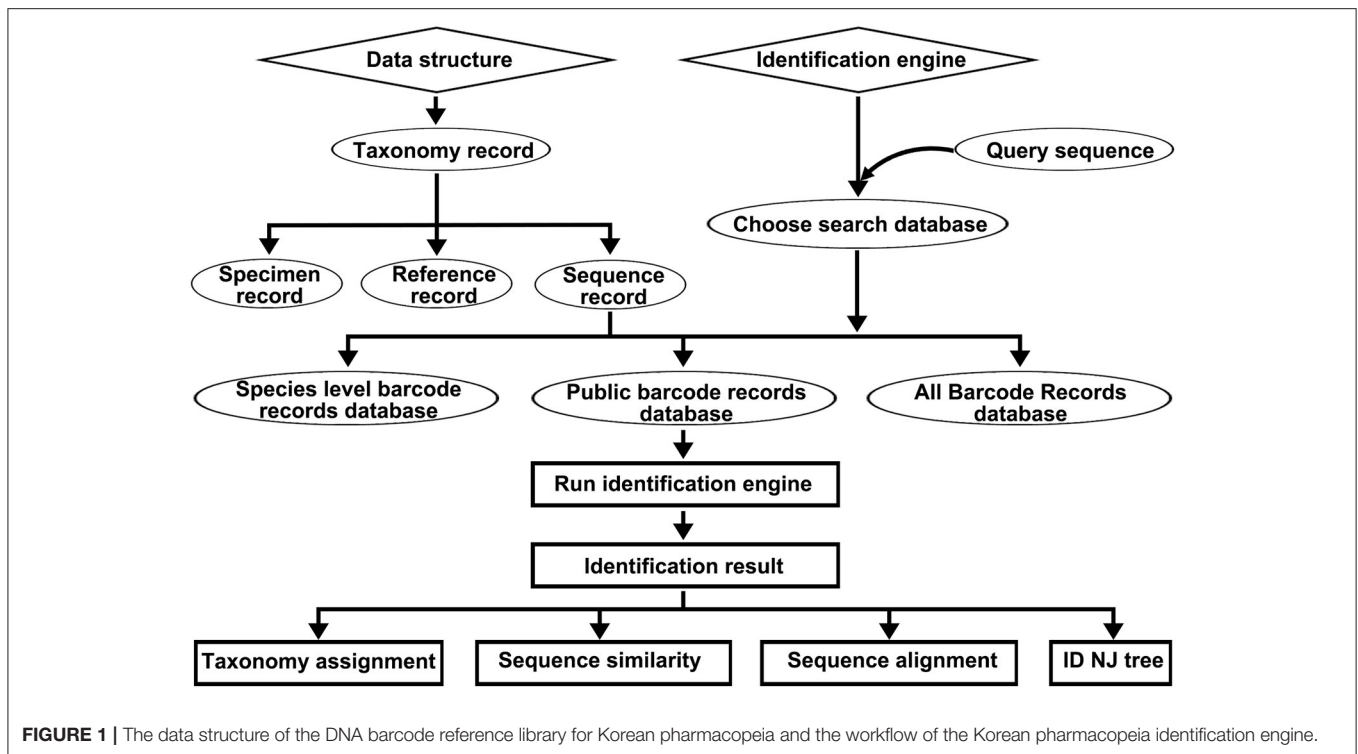
Data analysis was accomplished with a general pipeline: First of all, DNA barcodes were assembled and preprocessed. Secondly, DNA barcodes, together with specimens and references, were mapped on to species levels and formed the DNA barcode reference library of Korean pharmacopeia. Thirdly, the Korean pharmacopeia identification engine (KP-IDE) were developed and optimized. Finally, an independent dataset was applied to validate the reference library and identification engine. Common methods and data analysis techniques were as follow. High-quality bidirectional sequences (contigs) were assembled using CodonCode Aligner 6.0.1 (CodonCode Corporation, Massachusetts, USA). The delimitation and identification of the ITS2 and *psbA-trnH* core regions were achieved using HMMER 3.0 (<http://hmmer.org>). The sequence alignment analysis was accomplished using Muscle 3.8.1 (Edgar, 2004) and was restricted to the sequence after the annotation. The sequence divergences for the DNA barcode ITS2 and *psbA-trnH* region were calculated using PAUP* 4.0 (Swofford, 2002) and MEGA7 (Kumar et al., 2016) using the Kimura-2-Parameter (K2P) distance metric. Neighbor-joining (NJ) trees were constructed based on the

Muscle alignment results according to the K2P distances using MEGA7. Unknown specimens were identified using the (KP-IDE, **Data Sheet 1**) and herbal drug barcode libraries. KP-IDE was developed using Python 2.7 and Biopython 1.68, and the NCBI BLAST software was used as a similarity search engine to collect the nearest neighbors from the DNA barcode reference library.

RESULTS

Data Structure and Species Identification Engine

The DNA barcode reference library of the Korean pharmacopeia contains the following four types of records: taxonomy records, specimen records, sequence records, and reference records (**Figure 1**). The taxonomy records include a collection of standard nomenclature and classification repositories of the other three types of records. The taxonomy records include organism scientific names, synonyms, and taxonomic lineages for each of the other record types. All specimen records, sequence records, and reference records are closely linked to the original nomenclature of the herbal drugs, and all records must map onto the origin's taxonomy entry at or below the species level. Currently, there are 6,777 taxonomy



records (**Data Sheet 2**). The specimen records assemble various collateral data, particularly taxonomic information regarding specific herbal drug specimen, such as the locations in which the specimen were captured and other useful information. Each specimen record is linked to a companion sequence record. Currently, there are 1,054 specimen records included in this library (**Data Sheet 3**). The sequence records are a collection of DNA barcodes related to Korean herbal drugs and their closely related species. Currently, there are 30,744 deposited sequences, including 22049 ITS2 (**Data Sheet 4**) sequences and 8695 *psbA-trnH* sequences (**Data Sheet 5**). The reference records store all the bibliographies, including the publicly available DNA barcodes for Korean herbal drugs, and currently, there are 285 records (**Data Sheet 6**).

The above-mentioned types of records were embedded in the (KP-IDE) to facilitate the use of this reference library for identifying unknown specimens (**Figure 1**). The KP-IDE was written in Python and can operate very well on a Unix-like operating system. Unknown specimens were identified by pasting their barcode sequence into a fasta format file. The query sequence must satisfy the following criteria: appropriate length (ITS2 ≥ 150 bp; *psbA-trnH* ≥ 100 bp), confusion bases $\leq 1\%$ and no invalid bases. The query file was transferred to KP-IDE by the parameter “-q” or “-query.” Then, a BLAST search tool was used to collect the nearest neighbors from the user specified reference sequence record database or the default species level barcode record database (SBD). The parameter used for the transmission of the database is “-d” or “-database.” Currently, the species level barcode record database contains 5,986 sequences for 1,006 species, including 3770 ITS2 sequences

and 2216 *psbA-trnH* sequences. The other two databases are the public record barcodes database (PBD) and all record barcodes on BOKP database (ABD). The public record barcodes database is a collection of records from published articles related to Korean herbal drugs and currently include 12,045 sequences from 2,983 species, including 8006 ITS2 sequences and 4039 *psbA-trnH* sequences. The final type of database consists of all record barcodes related to Korean herbal drugs that are available on BOKP. Currently, there are 30744 sequences from 6,528 species, including 22049 ITS2 sequences and 8695 *psbA-trnH* sequences. When the identification engine retrieved a result, the KP-IDE provided the taxonomic assignment (**Figure 2A**), sequence similarity (**Figure 2B**) between unknown and top nearest reference sequences, NJ Tree (**Figure 2C**) showing top nearest reference sequences, and sequence alignment (**Figure 2D**) between unknown and top nearest reference sequences.

Standard Nomenclature of the Origins and the Differences in the Origins between the Korean and Chinese Pharmacopeias

All information, such as the specimen records, sequence records, and reference records, is closely linked by the organism nomenclature and must map onto the origin's taxonomy. Each taxonomy entry in this library includes an authority name, a primary name (i.e., the origin's name in the Korean pharmacopeia) and any number of synonyms in the Chinese pharmacopeia, the Flora of China and the NCBI taxonomy databases (only maintain the scientific name). Of the 232 herbal

drugs in the Korean pharmacopeia had more materials from the origin than those in the Chinese pharmacopeia (expansion, 11%), and 13 herbal drugs had less materials from the origin than those in the Chinese pharmacopeia (contraction, 8%). Compared with those in the Chinese pharmacopeia, seven herbal drugs in the Korean pharmacopeia had alternative origin materials (alternation, 4%). Finally, the origins of three herbal drugs were divided into two different herbal drug categories in the Chinese pharmacopeia. Epimedium Herb is the aerial part of *Epimedium koreanum*, *E. brevicornum*, *E. pubescens*, *E. wushanense*, and *E. sagittatum*, whereas *E. wushanense* is the origin of an independent herbal drug, Epimedium wushanense Herb, in the Chinese pharmacopeia. Liriope Tuber is the tuber of *Liriope platyphylla* and *Ophiopogon japonicas*, while *Ophiopogon japonicas* is the origin of another herbal drug, Ophiopogon Tuber, in the Chinese pharmacopeia. Phellodendron Bark is the bark of *Phellodendron amurense* and *P. chinense*; however, the two species are the origin of Phellodendron Amurense Bark and Phellodendron Chinense Bark, respectively, in the Chinese pharmacopeia. Detailed information regarding the differences in the origins of the herbal drugs in the Korean pharmacopeia and the Chinese pharmacopeia is provided in **Table S3**.

Assessment of the Sequence Variation in the Species Level Barcode Records Database and Real Herbal Drug Authentication

The following six parameters were used to characterize the inter-specific divergence and intraspecific divergence for herbal drug origins and their closely related species in the species level barcode record database: (i) average inter-specific distance, (ii) theta prime, (iii) average minimum inter-specific distance, (iv) average intraspecific distance, (v) theta, and (vi) average coalescent depth. The six parameters used for the ITS2 sequences are as follows: 0.0484 ± 0.0468 , 0.0637 ± 0.0506 , 0.0237 ± 0.0364 , 0.0018 ± 0.0035 , 0.0029 ± 0.0042 , 0.0049 ± 0.0065 ; the six parameters used for the *psbA-trnH* sequences are as follows: 0.0245 ± 0.0549 , 0.0535 ± 0.0992 , 0.0107 ± 0.0386 , 0.0023 ± 0.0276 , 0.0037 ± 0.0237 , 0.0066 ± 0.0437 . The distributions of average inter-specific distance, the minimum inter-specific distance, average intra-specific distance, and coalescent depth for each species are shown in **Figure 3A** (ITS2) and **Figure 3B** (*psbA-trnH*).

Twenty-seven herbal drug materials of 23 herbal drugs were collected from the Seoul Yangnyeongsi herbal medicine market to give an example for identifying real market samples. High-quality DNA was easily extracted from all 27 samples; 22 samples were amplified ITS2 barcodes, whereas the other five samples were amplified *psbA-trnH* barcodes. When performing the authentication using KP-IDE, 26 samples were matched with their labels. The herbal drug Rubus Fruit was an exception and was matched to the genus species *Rubus crataegifolius* rather than to its official origin *Rubus coreanus*. The identification results are shown in **Table 1**.

DISCUSSION

Standard Reference Barcodes Coupled with a Flexible Identification Engine Can Facilitate the Routine Authentication Related to Health and Safety Concerns

The DNA barcoding method built by Paul Hebert still has a major impact on species identification decades after its development (Hebert et al., 2003). The DNA barcode reference library has been shown to be a reliable resource that is essential for the identification of unknown specimens related to public health and safety concerns, such as food and herbal drugs (Stoeckle et al., 2011; Galimberti et al., 2013). The consortium for the barcode of life has initiated several campaigns for large groups of animals and provided an informatics workbench for the acquisition, storage, analysis, and publication of DNA barcode records (Ratnasingham and Hebert, 2007). The fish barcode of life (Fish-Bol, <http://www.fishbol.org>) may be the most influential campaign and has been widely employed to identify commercial seafood products (Ward et al., 2009). After more than 10 years of development, Fish-Bol assembled a standardized reference DNA sequence library for more than 8,000 fish species, which covered a vast majority of the world's most important commercial species (Becker et al., 2011). Subsequently, Hanner et al. procured 254 seafood samples from numerous retail establishments located in five Canadian metropolitan areas and demonstrated that 41% of the samples were mislabeled (Hanner et al., 2011). Wong et al. indicated that 25% of the 91 market seafood samples were potentially mislabeled in North America (Wong and Hanner, 2008). DNA barcoding has been demonstrated to be a powerful tool for seafood authentication and was adopted by the U.S. Food and Drug Administration to identify seafood products (Deeds et al., 2014).

Although there are several published DNA barcode reference libraries for herbal drugs, there is a lack of an informatics workbench, such as the BOLD system, which would allow the establishment of projects and private data repositories. The limitation of a private data repository and sharing hamper the ability of data increment and the flexible application of herbal drug authentication. In 2010, the first medicinal materials DNA barcode database (MMDBD, www.cuhk.edu.hk/icm/mmdbd.htm) was published by Lou et al. (2010). The latest update of the MMDBD was completed in 2014, and 51,375 sequences from 1,661 species were deposited. In 2014, Chen et al. established a preliminary system for the DNA barcoding of herbal materials (www.tcmbarcodes.cn), which contained 78,847 sequences belonging to 23,262 species (Chen et al., 2014). In 2016, Vassou et al. created a reference DNA barcode library for the authentication of Ayurvedic medicines (Vassou et al., 2016). This barcode library assembled 374 medicinal plants and revealed that only 79% of raw drugs were authentic. In 2017, Chen et al. sequenced ~95% of the species recorded in the JP and constructed an online DNA barcode identification system (<http://www.jpbarcode.com>; Chen et al., 2017). Here, we provide an open source identification

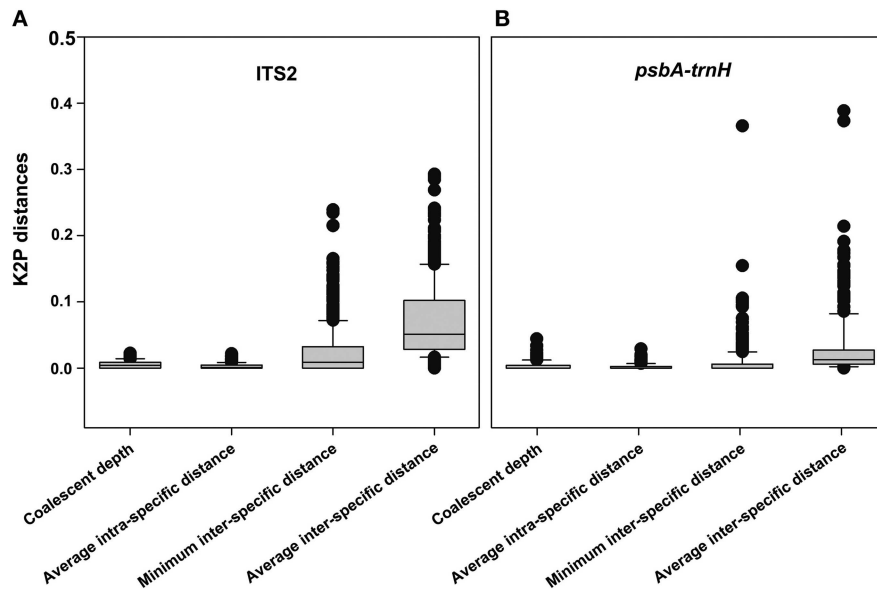


FIGURE 3 | The distributions of average inter-specific distance, the minimum inter-specific distance, average intra-specific distance and coalescent depth for each species. **(A)** ITS2; **(B)** *psbA-trnH*.

engine that permits researchers interested in herbal drug authentication to use DNA barcoding to study, change, and distribute this software to anyone for any purpose. Users who are not familiar with bioinformatics may use the parameter “-d” followed with “SDD” or “selfDefinded” to use their private dataset.

The Established DNA Barcode Reference Library Could Greatly Improve the Stability, Accuracy and Reliability of Korean Herbal Drug Authentication

Korean herbal drugs following the Korean pharmacopeia quality standards are generally recognized as safe and efficacious. However, the inappropriate use of herbal drugs due to misidentification poses considerable health risks. To combat the health risks associated with herbal drug misidentification, various methods for herbal drugs identification have been developed worldwide, primarily including morphological, microscopic, and physicochemical identification (Choi et al., 2002; Kim et al., 2010; Sahoo et al., 2010). These techniques are useful in herbal drug authentication, but there are also some limitations. The accuracy of the morphological and microscopic identification results faces significant challenges due to a dwindling professional workforce (Miller, 2007), and the authenticity of the physicochemical identification results is compromised as closely related species may share similar plant ingredients (Chen et al., 2014). For example, the identification method for *Akebia* stem is described in the Korean pharmacopeia as follows: weigh 0.5 g of pulverized *Akebia* stem, add 10 mL of water, boil, allow to cool and shake vigorously, and a lasting fine foam is produced. This soap-like foaming phenomenon is produced by saponins that are present

in the entire *Akebia* genus, Araliaceae and Liliaceae plants and not only in the diagnostic component of the origin plant *Akebia quinata*. Moreover, several other herbal drugs in the Korean pharmacopeia share these foam properties, such as *Anemarrhena* rhizome, *Aster* root and rhizome, and *Codonopsis pilosula* root.

DNA barcoding technology has been formally used for medicinal plant identification since 2010 (Chen et al., 2010) and has been shown to be a valuable addition to not only traditional identification methods but also practical field applications for the identification of medicinal plants and authentication of herbal drugs (Han J. et al., 2016). For example, the herbal drug *Akebia* stem has long been confused with *Caulis aristolochiae manshuriensis* in Asia and Europe. *Akebia* stem is the stem of *A. quinata* (Lardizabalaceae) whereas *Caulis aristolochiae manshuriensis* is the stem of *Aristolochia manshuriensis* (Aristolochiaceae). *A. manshuriensis* and other species of the same genus contain aristolochic acid (AA), which causes severe nephropathy. Although herbal drugs containing AA were rigorously administered in nearly all countries, the incidence of AA-contamination remains a worldwide problem (DeBelle et al., 2008). To avoid the health risks due to AAs, Wu et al. (2015) provided a DNA barcoding-based authentication system that can efficiently and reliably distinguish Aristolochiaceae from non-Aristolochiaceae materials. The DNA barcode reference library established here contains nearly all herbal drugs listed in the Korean pharmacopeia, and the herbal drug authentication test using real market samples collected from the Seoul Yangnyeongsi herbal medicine market further validated the value of the BOKP in the routine authentication of herbal drugs.

DNA Barcode Reference Library for Herbal Drugs in the Korean Pharmacopeia Has Room for Further Improvement

BOKP is the primary system used for the authentication of herbal drugs in the Korean pharmacopeia. Notably, the assembly of incorrectly identified sequences, poor quality sequences or unauthorized sequences threatens to exasperate the herbal drug authentication crisis. Moreover, the validation of data records is a time-consuming and labor intensive process. Thus, an automatic data pre-processing software should be developed to minimize manual system maintenance and updating. As reported elsewhere the marker of *psbA-trnH* has several shortcomings, such as inversions, rps19 insertions, insertions/deletions, AT-rich direct repeats, and 2-3 tandem repeats (Pang et al., 2012). For example, a core ATGAAAAC/GTTTTTCAT inversion exists among the origins of the herbal drug *Polygonatum* rhizome. A core AACAAAAC/GTTTTGTT inversion exists among the origins of the herbal drug *Coptis* rhizome. These inversions may lead to a substantial overestimation of the intra-/inter-specific distances and pose a challenge for species identification using both the blast and distance identification method (Whitlock et al., 2010). Moreover, the more frequent occurrence of the AT-rich direct repeats and tandem repeats in the origins of herbal drugs can lead to problems in obtaining bidirectional sequences and accurate identification results using blast-like methods (Group et al., 2009). Thus, a rigorously preprocessed protocol should be implemented automatically.

Economic development and increased international trade are leading to higher rates of inaccurate herbal drug identification, which risks patient safety and herbal efficacy. The failure in identification is undoubtedly due to nomenclature confusion, i.e., the phenomenon of synonymy. For example, *Evodia* Fruit is the fruit of *Evodia rutaecarpa*, *E. rutaecarpa* var. *officinalis* or *E. rutaecarpa* var. *bodiniieri* (Rutaceae). This herb is slightly toxic due to the presence of many alkaloids and should not be overdosed or used for a long period. The authorized name for the three *Evodia* Fruit origins is integrated into *Tetradium ruticarpum* in the Flora of China and the NCBI taxonomy database. Therefore, if an unknown specimen is identified as *T. ruticarpum*, the identification engine should note the nomenclature transition to avoid confusing a Korean user. In total, 26% of the origins of Korean herbal drugs have synonyms in the flora of China, and 24% of the origins have synonyms in the NCBI taxonomy database. This synonymy is a general

phenomenon for which the taxonomic issues (synonymies, misspellings, and alternate classifications) of nomenclature for Korean herbal drug origins should be resolved before the addition of new sequence data to the DNA barcode reference library.

DATA ACCESSIBILITY

DNA sequences from this study has been submitted to GenBank under the Genbank accessions: MF095889-MF097181.

AUTHOR CONTRIBUTIONS

JL, LS, and SC designed research. JL, JS, LS, WS, JH, XL, DH, HY, and ML performed research. JL and LS analyzed data. JL, LS, and JS wrote the paper.

ACKNOWLEDGMENTS

We thank Zhigang Hu and Li Xiang for plant materials. This work was supported by CAMS Innovation Fund for Medical Sciences (CIFMS, 2016-I2M-3-016), The College Youth Talent Project of Hebei Province (BJ2016002), The Hebei Provincial Natural Science Foundation (H2017406031), National Natural Science Foundation of China (81703659), and China Postdoctoral Science Foundation (2017M610815).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2017.00931/full#supplementary-material>

Table S1 | Standard reference materials and their barcode's GenBank accession numbers.

Table S2 | The species name for origins of herbal drugs in Korean pharmacopeias and their Synonyms in the flora of China and NCBI taxonomy database.

Table S3 | The list of origin differences (expansion, contraction, and alternation) between Korean pharmacopeia and Chinese pharmacopeia.

Data Sheet 1 | The program codes for identification engine KP_IDE.

Data Sheet 2 | Taxonomy records of the DNA barcode reference library.

Data Sheet 3 | Specimen records of the DNA barcode reference library.

Data Sheet 4 | ITS2 sequence records of the DNA barcode reference library.

Data Sheet 5 | *psbA-trnH* sequence records of the DNA barcode reference library.

Data Sheet 6 | Reference records of the DNA barcode reference library.

REFERENCES

- Becker, S., Hanner, R., and Steinke, D. (2011). Five years of FISH-BOL: brief status report. *Mitochondrial DNA* 22, 3–9. doi: 10.3109/19401736.2010.535528
- Cha, W.-S., Oh, J.-H., Park, H.-J., Ahn, S.-W., Hong, S.-Y., and Kim, N.-I. (2007). Historical difference between traditional Korean medicine and traditional Chinese medicine. *Neurol. Res.* 29, 5–9. doi: 10.1179/016164107X172293
- Chae, H., Lee, S., Park, S. H., Jang, E., and Lee, S. J. (2012). Development and validation of a personality assessment instrument for traditional Korean

- medicine: sasang personality questionnaire. *Evid. Based Complement. Altern. Med.* 2012:12. doi: 10.1155/2012/657013
- 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., 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
- Chen, X., Xiang, L., Shi, L., Li, G., Yao, H., Han, J., et al. (2017). Identification of crude drugs in the Japanese pharmacopeia using a DNA barcoding system. *Sci. Rep.* 7:42325. doi: 10.1038/srep42325

- Choi, D. W., Kim, J. H., Cho, S. Y., Kim, D. H., and Chang, S. Y. (2002). Regulation and quality control of herbal drugs in Korea. *Toxicology* s181–s182, 581–586. doi: 10.1016/S0300-483X(02)00487-0
- 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
- Debelle, F. D., Vanherweghem, J.-L., and Nortier, J. L. (2008). Aristolochic acid nephropathy: a worldwide problem. *Kidney Int.* 74, 158–169. doi: 10.1038/ki.2008.129
- Deeds, J. R., Handy, S. M., Fry, F. Jr., Granade, H., Williams, J. T., Powers, M., et al. (2014). Protocol for building a reference standard sequence library for DNA-based seafood identification. *J. AOAC Int.* 97, 1626–1633. doi: 10.5740/jaoacint.14-111
- Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. doi: 10.1186/1471-2105-5-113
- Galimberti, A., De Mattia, F., Losa, A., Bruni, I., Federici, S., Casiraghi, M., et al. (2013). DNA barcoding as a new tool for food traceability. *Food Res. Int.* 50, 55–63. doi: 10.1016/j.foodres.2012.09.036
- Group, C. P. B., Li, D.-Z., Gao, L.-M., Li, H.-T., Wang, H., Ge, X.-J., et al. (2011). Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proc. Natl. Acad. Sci. U.S.A.* 108, 19641–19646. doi: 10.1073/pnas.1104551108
- Group, C. P. W., Hollingsworth, P. M., Forrest, L. L., Spouge, J. L., Hajibabaei, M., Ratnasingham, S., et al. (2009). A DNA barcode for land plants. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12794–12797. doi: 10.1073/pnas.0905845106
- Han, J., Pang, X., Liao, B., Yao, H., Song, J., and Chen, S. (2016). An authenticity survey of herbal medicines from markets in China using DNA barcoding. *Sci. Rep.* 6:18723. doi: 10.1038/srep18723
- Han, S. Y., Kim, H. Y., Lim, J. H., Cheon, J., Kwon, Y. K., Kim, H., et al. (2016). The past, present, and future of traditional medicine education in Korea. *Integr. Med. Res.* 5, 73–82. doi: 10.1016/j.imr.2016.03.003
- Hanner, R., Becker, S., Ivanova, N. V., and Steinke, D. (2011). FISH-BOL and seafood identification: geographically dispersed case studies reveal systemic market substitution across Canada. *Mitochondrial DNA* 22, 106–122. doi: 10.3109/19401736.2011.588217
- Hebert, P. D., Cywinska, A., and Ball, S. L. (2003). Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B Biol. Sci.* 270, 313–321. doi: 10.1098/rspb.2002.2218
- Hollingsworth, P. M., Graham, S. W., and Little, D. P. (2011). Choosing and using a plant DNA barcode. *PLoS ONE* 6:e19254. doi: 10.1371/journal.pone.0019254
- Huang, D. S., and Shin, H.-K. (2012). Perception of traditional medical doctors on policy issues in Korea: cooperation with western medicine, and the demand and supply of practitioners. *J. Korean Med.* 33, 17–25. Available online at: <https://www.jkom.org/journal/view.php?number=3>
- Kim, H. J., Jee, E. H., Ahn, K. S., Choi, H. S., and Jang, Y. P. (2010). Identification of marker compounds in herbal drugs on TLC with DART-MS. *Arch. Pharm. Res.* 33, 1355–1359. doi: 10.1007/s12272-010-0909-7
- Kim, Y.-S., Jun, H., Chae, Y., Park, H.-J., Kim, B. H., Chang, I.-M., et al. (2005). The practice of Korean medicine: an overview of clinical trials in acupuncture. *Evid. Based Complement. Altern. Med.* 2, 325–352. doi: 10.1093/ecam/neh102
- 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
- Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A., and Janzen, D. H. (2005). Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U.S.A.* 102, 8369–8374. doi: 10.1073/pnas.0503123102
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Lahaye, R., van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., et al. (2008). DNA barcoding the floras of biodiversity hotspots. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2923–2928. doi: 10.1073/pnas.0709936105
- Lou, S.-K., Wong, K.-L., Li, M., But, P. P., Tsui, S. K., and Shaw, P.-C. (2010). An integrated web medicinal materials DNA database: MMDDB (Medicinal Materials DNA Barcode Database). *BMC Genomics* 11:402. doi: 10.1186/1471-2164-11-402
- MFDS, K. (2014). *Korean Pharmacopoeia X (KPX): The KFDA Notification*.
- Miller, S. E. (2007). DNA barcoding and the renaissance of taxonomy. *Proc. Natl. Acad. Sci. U.S.A.* 104, 4775–4776. doi: 10.1073/pnas.0700466104
- Newmaster, S. G., Grguric, M., Shanmughanandhan, D., Ramalingam, S., and Ragupathy, S. (2013). DNA barcoding detects contamination and substitution in North American herbal products. *BMC Med.* 11:222. doi: 10.1186/1741-7015-11-222
- Pang, X., Liu, C., Shi, L., Liu, R., Liang, D., Li, H., et al. (2012). Utility of the *trnH-psbA* intergenic spacer region and its combinations as plant DNA barcodes: a meta-analysis. *PLoS ONE* 7:e48833. doi: 10.1371/journal.pone.0048833
- Pang, X., Song, J., Zhu, Y., Xu, H., Huang, L., and Chen, S. (2011). Applying plant DNA barcodes for Rosaceae species identification. *Cladistics* 27, 165–170. doi: 10.1111/j.1096-0031.2010.00328.x
- Ratnasingham, S., and Hebert, P. D. N. (2007). BOLD: the barcode of life data system (www.barcodinglife.org). *Mol. Ecol. Notes* 7, 355–364. doi: 10.1111/j.1471-8286.2007.01678.x
- Sahoo, N., Manchikanti, P., and Dey, S. (2010). Herbal drugs: standards and regulation. *Fitoterapia* 81, 462–471. doi: 10.1016/j.fitote.2010.02.001
- Savolainen, V., Cowan, R. S., Vogler, A. P., Roderick, G. K., and Lane, R. (2005). Towards writing the encyclopaedia of life: an introduction to DNA barcoding. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1805–1811. doi: 10.1098/rstb.2005.1730
- Stoeckle, M. Y., Gamble, C. C., Kirpekar, R., Young, G., Ahmed, S., and Little, D. P. (2011). Commercial teas highlight plant DNA barcode identification successes and obstacles. *Sci. Rep.* 1:42. doi: 10.1038/srep00042
- Swofford, D. (2002). *PAUP* 4.0: Phylogenetic Analysis Using Parsimony*. Sunderland, MA: Sinauer Associates.
- Valentini, A., Pompanon, F., and Taberlet, P. (2009). DNA barcoding for ecologists. *Trends Ecol. Evol.* 24, 110–117. doi: 10.1016/j.tree.2008.09.011
- Vassou, S. L., Nithaniyal, S., Raju, B., and Parani, M. (2016). Creation of reference DNA barcode library and authentication of medicinal plant raw drugs used in Ayurvedic medicine. *BMC Complement. Altern. Med.* 16:186. doi: 10.1186/s12906-016-1086-0
- Ward, R. D., Hanner, R., and Hebert, P. D. (2009). The campaign to DNA barcode all fishes, FISH-BOL. *J. Fish Biol.* 74, 329–356. doi: 10.1111/j.1095-8649.2008.02080.x
- Whitlock, B. A., Hale, A. M., and Groff, P. A. (2010). Intraspecific inversions pose a challenge for the *trnH-psbA* plant DNA barcode. *PLoS ONE* 5:e11533. doi: 10.1371/journal.pone.0011533
- World Health Organization (1991). *Guidelines for the Assessment of Herbal Medicines*. Programme Traditional Medicine.
- Wong, E. H.-K., and Hanner, R. H. (2008). DNA barcoding detects market substitution in North American seafood. *Food Res. Int.* 41, 828–837. doi: 10.1016/j.foodres.2008.07.005
- Wu, L., Sun, W., Wang, B., Zhao, H., Li, Y., Cai, S., et al. (2015). An integrated system for identifying the hidden assassins in traditional medicines containing aristolochic acids. *Sci. Rep.* 5:11318. doi: 10.1038/srep11318

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 Liu, Shi, Song, Sun, Han, Liu, Hou, Yao, Li 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.