



Development of a Specific Mini-Barcode From Plastome and its Application for Qualitative and Quantitative Identification of Processed Herbal Products Using DNA Metabarcoding Technique: A Case Study on *Senna*

OPEN ACCESS

Edited by:

Sukvinder Kaur Bhamra,
University of Kent, United Kingdom

Reviewed by:

Subhash Chandra Mandal,
Directorate of Drugs Control,
Government of West Bengal, India
Sol Cristians,
National Autonomous University of
Mexico, Mexico

*Correspondence:

Lin Miao
mmlin@tjutc.edu.cn
Xiaoxuan Tian
tian_xiaoxuan@tjutc.edu.cn

Specialty section:

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

Received: 21 July 2020

Accepted: 26 November 2020

Published: 17 December 2020

Citation:

Yu X, Tan W, Gao H, Miao L and Tian X
(2020) Development of a Specific Mini-Barcode From Plastome and its Application for Qualitative and Quantitative Identification of Processed Herbal Products Using DNA Metabarcoding Technique: A Case Study on *Senna*.
Front. Pharmacol. 11:585687.
doi: 10.3389/fphar.2020.585687

Xiaolei Yu, Wei Tan, Han Gao, Lin Miao* and Xiaoxuan Tian*

State Key Laboratory of Component-based Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin, China

Herbal products play an important role globally in the pharmaceutical and healthcare industries. However, some specific groups of herbal products are easily adulterated by confused materials on the market, which seriously reduces the products' quality. Universal conventional DNA barcodes would function poorly since the processed herbal products generally suffer from varying degrees of DNA degradation and DNA mixing during processing or manufacturing. For quality control purposes, an accurate and effective method should be provided for species identification of these herbal products. Here, we provided a strategy of developing the specific mini-barcode using *Senna* as an example, and by coupling with the metabarcoding technique, it realized the qualitative and quantitative identification of processed herbal products. The plastomes of *Senna obtusifolia* (L.) H.S. Irwin & Barneby and *Senna occidentalis* (L.) Link were newly assembled, and the hypervariable coding-regions were identified by comparing their genomes. Then, the specific mini-barcodes were developed based on the identified hypervariable regions. Finally, we applied the DNA metabarcoding technique to the developed mini-barcodes. Results showed that the lengths of plastomes of *S. obtusifolia* and *S. occidentalis* were 162,426 and 159,993 bp, respectively. Four hypervariable coding-regions *ycf1*, *rpl23*, *petL*, and *matK* were identified. Two specific mini-barcodes were successfully developed from *matK*, and the mini-barcode of primer 647F-847R was proved to be able to qualitatively and quantitatively identify these two processed *Senna* seeds. Overall, our study established a valuable way to develop the specific mini-barcode, which may provide a new idea for the quality control of processed herbal products.

Keywords: mini-barcode, plastome, metabarcoding, species identification, Herbal products, *Senna*

INTRODUCTION

Herbal products use medicinal plants as raw materials for herbal medicines, herbal extracts and dietary supplements. In recent years, there was a tremendous increase in the global demand of herbal products, making herbal products adulteration and counterfeiting a global problem (Mackey and Liang, 2013; Gromek et al., 2015; Gaudiano et al., 2016). Hence, the authentication of herbal products has become an important topic within and beyond the pharmaceutical and healthcare industries. DNA barcoding is a technique for authenticating species using a standard DNA region, aiming to provide rapid, automatable, and cost-effective methods for accurate identification at the species-level. Studies on DNA barcoding have made remarkable progress in species identification (Naem et al., 2014; Zhang et al., 2016; Viglietti et al., 2019). Initially used as an identification tool, DNA barcoding is now applied in the industrial quality assurance context to identify herbal products (Mosa et al., 2018; Amritha et al., 2020). However, DNA barcoding faces practical limitations that exclusively restrict it to identifying single ingredient herbal products. Once the plants undergo a series of extraction and processing steps resulting in DNA mixing, DNA barcoding would not be a suitable choice. Unfortunately, herbal products generally suffer from varying degrees of DNA mixing during processing or manufacturing. For a suitable solution to this case, DNA metabarcoding appeared to be an alternative choice.

DNA metabarcoding, combining the next-generation sequencing and DNA barcoding, enables simultaneous multi-taxa identification by using the total DNA extracted from complex samples containing DNA from different origins (Taberlet et al., 2012; Staats et al., 2016). It has generally been assumed that the proportion of reads obtained for a given species is proportional to the contribution of species biomass (Amend et al., 2010; Egge et al., 2013). Hence, species reads obtained from DNA metabarcoding can reflect the richness of species in the community to a certain extent, which makes DNA metabarcoding a widely used tool for biodiversity studies (Egge et al., 2013; Klindworth et al., 2013; Duke and Burton, 2020). Nowadays, DNA metabarcoding has also proved to be applicable for authentication of species diversity in herbal products, and has been used to investigate the level of discrepancy between the expected and detected plant species in the herbal product market (Cheng et al., 2014; Coghlan et al., 2015; Ivanova et al., 2016). However, although DNA metabarcoding is superior to DNA barcoding in the assessment of complex herbal mixtures, it still faces some limitations similar to DNA barcoding. For instance, the selection of barcodes has always been a challenge. Barcodes that are too short may not provide sufficient resolution for identification of multiple-taxa, e.g., P6 loop of *trnL* intron (Taberlet et al., 2007). In contrast, conventional barcodes like *matK*, *rbcl*, *trnH-psbA*, and *ITS2* are generally longer than 500 bp, which is very unfavorable for PCR amplification when facing DNA degraded samples including the processed herbal products (Gao et al., 2019). In this case, a short but informative DNA barcode could be a solution, and that is called a mini-barcode.

The DNA mini-barcode is a short DNA fragment, 100–250 bp in length, with sufficient variable sites for species identification (Little, 2014b). Due to the significantly reduced length of barcode

regions, PCR amplification success could be much improved, but the barcode resolution would thus be limited. So, it is necessary to design a specific mini-barcode for accurate species identification of close-related species. For angiosperms, it is now realistic to find a specific DNA mini-barcode by searching the whole plastome owing to the ease of next-generation sequencing (Dong et al., 2013). The plastomes of most land plants exhibit a typical quadripartite structure with stable gene content and gene order (Li et al., 2017). Owing to their characteristics of maternally inherited, multi-copy, and moderate evolutionary rate, the plastome sequences are widely used for molecular marker development (Park et al., 2018; Yu et al., 2019; Li et al., 2020). Moreover, the abundant interspecific sequence diversity makes the plastome sequence a particularly useful tool for providing high-resolution barcodes for close-related species (Jiao et al., 2019; Xia et al., 2019). The utility of mini-barcodes has been successfully demonstrated in a number of specific taxa, such as *Panax* (Dong et al., 2014), *Hypericum* (Costa et al., 2016), and *Phyllanthus* (Srirama et al., 2014).

The seed of *Senna obtusifolia*, called Juemingzi in China, is regarded as a dual-use material for food and medicine by China Food and Drug Administration. Studies on Juemingzi showed that it has various pharmaceutical properties such as hypertension regulation (Li and Guo, 2002), hepatoprotective effect (Kim et al., 2009) and eyesight improvement (Yang et al., 2012), which have made it a very popular herbal product in China and some other Asian countries. Since the wide applications in pharmaceutical and healthcare industries, the demand for Juemingzi increased rapidly, and the proportion of its adulterant increased at the same time. Among these adulterants, the most common and indistinguishable one is the seed of *S. occidentalis*. However, just like the conditions of other herbal products generally suffering from varying degrees of DNA degradation during processing, the conventional DNA barcodes for Juemingzi was difficult to amplify as well. Moreover, mixing seeds could make DNA barcoding function poorly. Therefore, it is necessary to design the specific mini-barcode and combine with the metabarcoding technique for quality control of Juemingzi.

The aims of our study included: 1) to provide a strategy of developing the specific mini-barcode using *Senna* as an example; 2) to test the feasibility of combining mini-barcodes and DNA metabarcoding techniques for qualitative and quantitative identification of processed herbal products. In this study, the plastomes of *S. obtusifolia* and *S. occidentalis* were first assembled, then the hypervariable coding-regions were further sought out by comparing their genomes. Subsequently, the length and position of suitable mini-barcodes were determined, and finally, we applied DNA metabarcoding techniques to the developed mini-barcodes.

MATERIALS AND METHODS

Plant Material

The fresh plant of *S. obtusifolia* was collected from the Medicinal Botanical Garden of Tianjin University of Traditional Chinese Medicine, Tianjin City (117.06°E,

38.96°N), China. The ungerminated seed of *S. occidentalis* was collected from Baoding City (115.33°E, 38.42°N), China. They were identified by Prof. Tianxiang Li from School of Chinese Materia Medica, Tianjin University of Traditional Chinese Medicine. The voucher species were deposited in Tianjin State Key Laboratory of Modern Chinese Medicine, Tianjin University of Traditional Chinese Medicine, and the voucher numbers were JMZ201806 (*S. obtusifolia*) and WJN201808 (*S. occidentalis*). The Extract Plant DNA kit (Sangon Biotech Co., Ltd., Shanghai, China) was used to extract the total genomic DNA from the fresh leaves of *S. obtusifolia* and the whole seed of *S. occidentalis*. In addition, four batches of processed seeds of *S. obtusifolia* and *S. occidentalis* were purchased from Anguo medicine market to construct experimental mixtures. In particular, the seeds were identified by Prof. Tianxiang Li to avoid the possibility of adulteration. DNA purity was checked using NanoPhotometer[®]spectrophotometer (IMPLEN, CA, United States). Concentrated DNA was measured using Qubit[®] DNA Assay Kit in Qubit[®] 2.0 Fluorometer (Life Technologies, CA, United States). Sequencing library was generated using Truseq Nano DNA HT Sample preparation Kit (Illumina United States) following the manufacturer's recommendations. The library was sequenced by Illumina HiSeq X Ten platform (Novogene, Nanjing, China) and 150 bp paired-end reads were generated.

Plastome *de Novo* Assembly and Annotation

Raw data of fastq format of DNA was processed by removing adapter sequences, removing reads with the ratio of N (N indicates that base information cannot be determined) greater than 10%, and removing low-quality reads in which >50% of the bases had a quality value $Q_{phred} \leq 5$. Subsequent analyses were based on the filtered high-quality sequences. The plastomes of *S. obtusifolia* and *S. occidentalis* were assembled via combination of *de novo* and reference-guided assembly approaches following the procedure described by Niu *et al.* (Niu *et al.*, 2017). The plastome of *S. tora* (NCBI accession number NC_030193) was used as a reference. Then, the finished plastomes of *S. obtusifolia* and *S. occidentalis* were annotated using GeSeq (Tillich *et al.*, 2017), coupled with manual corrections for start and stop codons. Finally, the plastomes of *S. obtusifolia* and *S. occidentalis* were visualized using OGDRAW (Lohse *et al.*, 2013).

Identification of Hypervariable Coding-Regions and Design of Primers

To find the hypervariable coding-regions, the protein-coding genes of two plastomes were respectively extracted and aligned using PhyloSuite_v1.1.15 (Zhang *et al.*, 2020). Then, the nucleotide variability (Pi) values of protein-coding genes were calculated using DnaSP version 6.11.01 software (Rozas *et al.*, 2017). The regions with high Pi values were selected as the candidate regions for mini-barcode development. The Primer Premier V6.0 was used to design primers for the mini-barcode, and the parameters were as

follows: product size between 150 and 300 bp, primer size between 18 and 30 bp, melting temperature (T_m) between 40° and 70°C, GC content between 30 and 70%. Then, the physicochemical properties of the designed primers such as hairpin structure, primer dimer, and annealing temperature were evaluated using Oligo seven software, and primers that were likely to have hairpin structures, primer dimers, or excessive annealing temperature were abandoned.

Experimental Mixtures Construction, Next-Generation Sequencing, and Analysis of Amplicon Sequence Variants (ASVs)

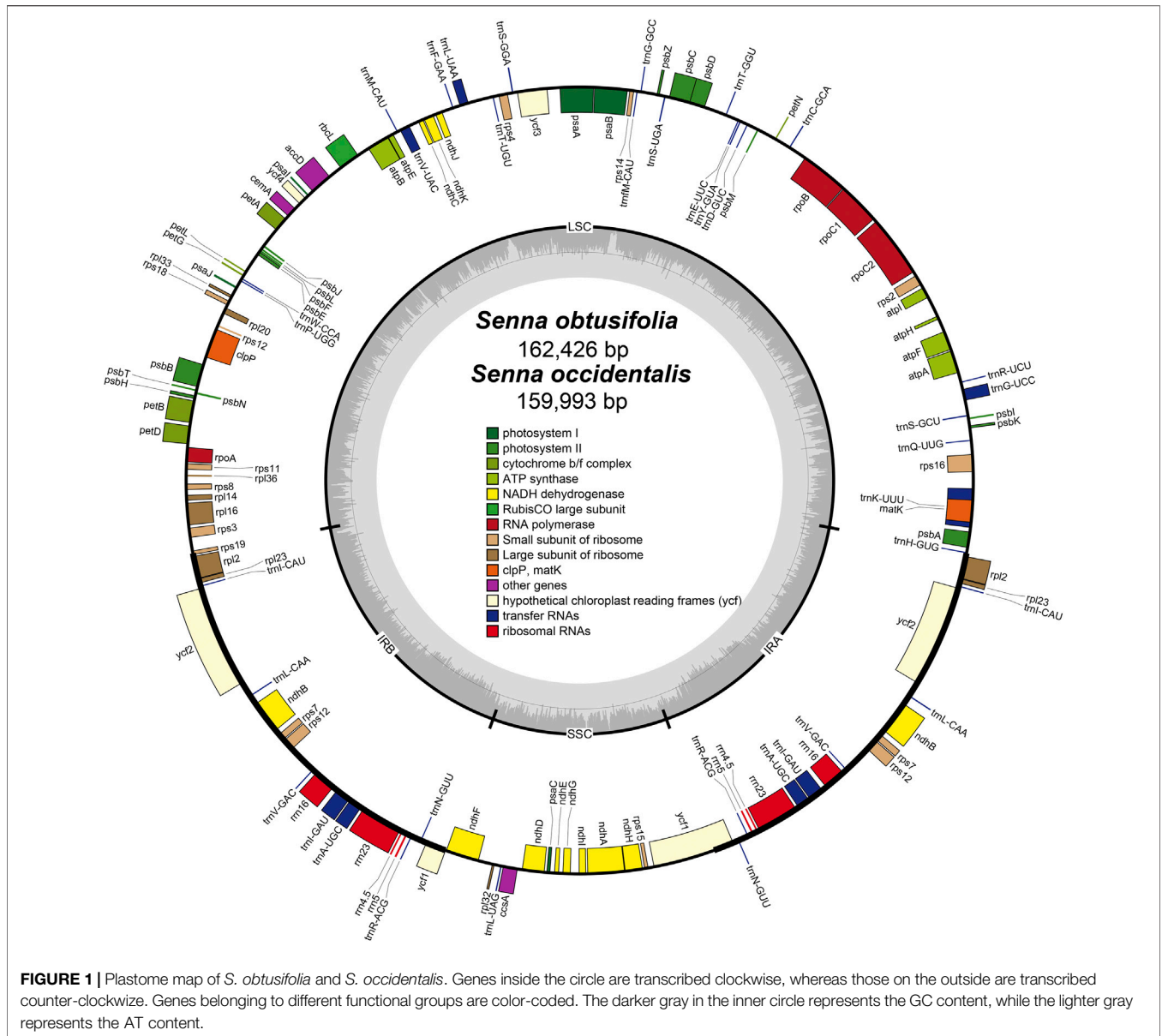
To test the qualitative and quantitative capacity of developed mini-barcodes in the processed mixture of *S. obtusifolia* seeds and *S. occidentalis* seeds, we prepared four experimental mixtures (for the biomass of each species in the four experimental mixtures, please see **Supplementary Table S1**). Each experimental mixture contained processed seeds (seeds were fully crushed by a powder mill to facilitate sample mixing) of *S. obtusifolia* and *S. occidentalis*. Then, the genomic DNA was extracted from each experimental mixture using the Extract Plant DNA kit (Sangon Biotech Co., Ltd., Shanghai, China), respectively. The target regions were amplified using two pairs of fusion primers with matching tag sequences (for detailed tag sequences, please see **Supplementary Table S2**) to ensure that tag jumps would not result in false assignments of sequences to samples (Schnell *et al.*, 2015). Then, PCR reaction was conducted in a 25 μ L reaction with 12.5 μ L of 2 \times Gflex PCR Buffer (containing 1 mM Mg^{2+} and 200 μ M dNTP), 0.5 μ L of each primer, 0.5 μ L Tks Gflex[™] (Takara Biomedical Technology Co., Ltd., Beijing, China) DNA Polymerase (1.25 units/ μ L), 2 μ L template DNA, and approximately 9 μ L ddH₂O. The PCR protocol was as follows: preheating at 94°C for 1 min, 30 cycles at 98°C for 10 s, annealing at 55°C for 15 s and elongation at 68°C for 30 s, and final extension at 68°C for 5 min. The negative controls were included in every run. Then, the PCR products were detected on 2% agarose gels. Subsequently, equimolar concentrations of PCR products were pooled to construct the library. Finally, the library was sequenced with 2 \times 150 bp reads on the Illumina HiSeq X Ten platform.

The fastq-multx (Aronesty, 2013) was used to split the generated data according to the tag sequences. Then, the primer sequences were trimmed using Cutadapt (Kechin *et al.*, 2017). To construct ASVs (sequences with 100% similarity will be assigned into each ASV), denoise and quality control were performed using DADA2 (Callahan *et al.*, 2016). Then, BlastN was used to search the sequences of ASVs from the chloroplast genomes of *S. obtusifolia* and *S. occidentalis*. Finally, the relationship between species reads and species biomass was analyzed using R package ampvis2 (Andersen *et al.*, 2018).

RESULTS

Plastome Features

The plastomes of *S. obtusifolia* and *S. occidentalis* were 162,426 and 159,993 bp in length, respectively (**Figure 1**). Either of the two plastomes displayed a typical quadripartite structure consisting of a pair of IR regions (26,791 bp in *S. obtusifolia*, 26,101 bp in *S.*



occidentalis) separated by the LSC (90,843 bp in *S. obtusifolia*, 89,322 bp in *S. occidentalis*) and SSC (26,791 bp in *S. obtusifolia*, 26,101 bp in *S. occidentalis*) regions. Both plastomes exhibited a low GC content (36.0% and 36.2% in *S. obtusifolia* and *S. occidentalis*, respectively). The two *Senna* plastomes encoded an identical set of 129 predicted functional genes, 111 of which were unique, and 18 were duplicated in the IR regions. Overall, the genome structure, gene type, gene content, and GC content of the two *Senna* plastomes were highly similar.

Selection of Hypervariable Coding-Regions and Design of Primers

To find the hypervariable coding-regions for mini-barcode development, the Pi values of the shared 77 protein-genes of two

plastomes were calculated using DnaSP. As shown in **Figure 2A**, *ycf1* had the highest Pi values (0.04363), followed by *rpl23* (0.04301), *petL* (0.03226), and *matK* (0.03206). These observed hypervariable genes were regarded as the candidate regions for mini-barcode development. Then, the primers for the mini-barcode were designed based on these selected hypervariable regions. As depicted in **Table 1**, two primer pairs were successfully designed from *matK*. However, limited by the length and variability of the barcode regions, as well as the physicochemical characteristics and conservation of the primers, no suitable primers were designed from *rpl23*, *petL* and *ycf1*. The amplicon sizes for primer 647F-847R and 478F-629R were 200 and 151 bp, respectively (**Figures 2B,C**). Both mini-barcode of primer 647F-847R and 478F-629R had seven variable sites, suggesting a great capacity for identifying the two different *Senna* species.

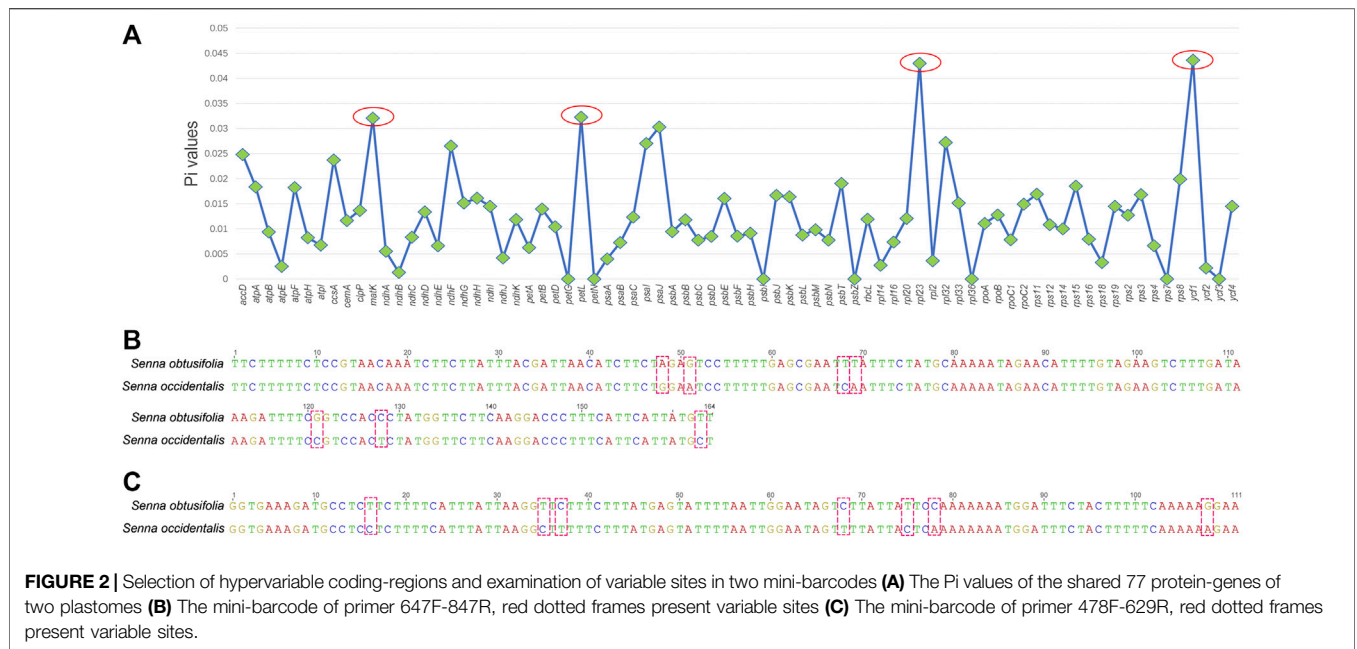


FIGURE 2 | Selection of hypervariable coding-regions and examination of variable sites in two mini-barcodes **(A)** The Pi values of the shared 77 protein-genes of two plastomes **(B)** The mini-barcode of primer 647F-847R, red dotted frames present variable sites **(C)** The mini-barcode of primer 478F-629R, red dotted frames present variable sites.

TABLE 1 | Primers developed from *matK*.

Primer name	647F-847R	478F-629R
Forward primer sequence 5' to 3	GTGAATACGAATCTATCT	GTTCAAACCCTTCGATACTG
Reverse primer sequence 5' to 3	GGATTTTCCTTGATATCT	GGAACAGGAAAAATCTTGGAA
Amplicon size (bp)	200	151
Variable sites of mini-barcode	7	7
Sequence excluding primers (bp)	164	111
GC% (For/Rev)	37.5/31.2	33.3/33.3
Tm (For/Rev)	45.2/42.4	45.7/45.5

TABLE 2 | BLAST results and sequencing reads of ASV.

Primer name	ASV	Sequencing reads of ASV in four experimental mixtures				Blast result	Identity (%)
		JM1	JM2	JM3	JM4		
647F-847R	ASV1	56,065	564,176	11,901	400,029	<i>C.</i>	100
	ASV2	57,978	31,078	321,102	348,064	<i>C. occidentalis</i>	100
478F-629R	ASV3	108,018	288,609	6,666	65,827	<i>C.</i>	100
	ASV4	27,160	2,937	79,349	3,718	<i>C. obtusifolia</i>	100

Qualitative and Quantitative Identification of Two *Senna* Species in Experimental Mixtures by DNA Metabarcoding Technique

Two designed primer pairs were used to amplify the barcode regions in four experimental mixtures. Illumina sequencing results of the amplified products showed that primer 647F-847R generated 1,790,393 reads, which were subsequently clustered into two ASVs

(Table 2). These two ASVs were identified as *S. obtusifolia* (ASV1) and *S. occidentalis* (ASV2), with 100% similarity against the corresponding barcode regions. Similarly, for primer 478F-629R, a total of 582,284 reads were generated and clustered into two ASVs, which were identified as two *Senna* species (100% similarity) (Table 2). These two primer pairs successfully amplified the target barcode regions in four experimental mixtures, indicating a qualitative capacity of two mini-barcodes in identifying these two *Senna* species.

To validate the quantitative capacity of the mini-barcodes, the relationship between species reads proportion and species biomass proportion was evaluated in four experimental mixtures (Figure 3A). For primer 478F-629R, there was a significant difference between species reads proportion and species biomass proportion in JM1 and JM4, suggesting a poor quantitative capacity of this mini-barcode in identifying two *Senna* species. For primer 647F-847R, the species reads proportion were almost identical to species biomass proportion in four experimental mixtures. Further correlation analysis (Figure 3B) showed a significant correlation of species reads proportion and species biomass proportion in four experimental mixtures ($R^2 = 0.9975$ in two *Senna* species), which indicated that the mini-barcode of primer 647F-847R demonstrated a relatively accurate quantitative ability in identifying two *Senna* species.

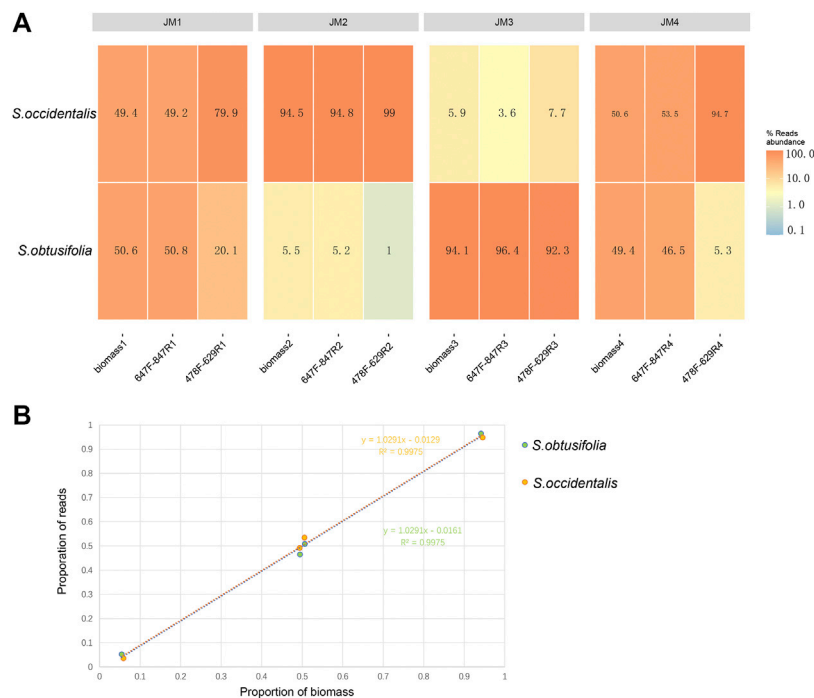


FIGURE 3 | Analysis of species reads abundance and species biomass abundance **(A)** Heat map of reads abundance of two primer pairs in four experimental mixtures **(B)** For primer 657F-858R, the correlation analysis of species reads proportion and species biomass proportion in four experimental mixtures. X-axis, proportion of biomass; Y-axis, proportion of reads.

DISCUSSION

Practically, the length of a barcode has always been an issue of concern. A barcode can be generally classified into four types according to its length, for example, micro-barcode within 100 bp (Taberlet et al., 2007), mini-barcode of 100–250 bp (Little, 2014b), standard barcode of 400–800 bp (Hebert et al., 2003), and the whole genome as a super-barcode (Chen et al., 2018). However, micro-barcodes can provide limited resolution for species identification. Although the standard barcode and super-barcode can provide strong resolution, they are difficult to be obtained from the degraded samples due to length limitation. Compared with the three types mentioned above, mini-barcode is the best choice for identifying processed and confused herbal products due to its sufficient variable sites and short in length. However, mini-barcode of high resolution is not easily found, and that is why the whole genome is indispensable for the development of mini-barcodes.

An ideal mini-barcode should meet two requirements: one is a barcode region with sufficient variable sites and short in length, and the other is a primer pair with conservation and meeting physicochemical properties. It is necessary to incorporate the physicochemical properties of the primer into consideration because the hairpin structure, primer dimer, and annealing temperature are all important factors that affect the success of primer amplification (Xie et al., 2019). In addition, insertion and deletion of nucleotides, as well as excessive length variations, are frequently observed in the intergenic region of the chloroplast

genome (Bortiri et al., 2008; Liu et al., 2019), which is not conducive to primer conservation. Moreover, the intra-specific variation in the intergenic region is much larger than that in the coding-region (Jiang et al., 2017). Any intra-specific variation may affect the quantitative capacity of the mini-barcode when there are multiple individuals of target species in the mixture. Therefore, the intergenic region was excluded when screening potential DNA barcode regions based on the chloroplast genome.

matK is a recommended gene as the candidate DNA barcode because of its high evolution rate (Lahaye et al., 2008). At the family and genus levels, *matK* provides the high support value of phylogenetic tree and the high-resolution of species discrimination in a given taxon (Ji et al., 2007; Sun et al., 2012). In some groups like *Juglans* (Stanford et al., 2000) and *Fagopyrum* (Ohsako and Ohnishi, 2000), *matK* has also played a useful role in studying intra- and interspecific phylogeny. However, a standard barcode length of *matK* is approximately 800 bp, which is not suitable for identifying severely degraded samples. So, there have been studies focused on the development of mini-barcodes from *matK* and have achieved good results in some specific taxa. For example, Costa et al. (2016) suggested that *matK* was an adequate mini-barcode region for the differentiation of two *Hypericum* species in herbal infusions. Little (2014a) developed a 166 bp mini-barcode from *matK* to authenticate *Ginkgo biloba* herbal dietary supplements. Our study developed two specific mini-barcodes and successfully applied them to identify two *Senna* species, which once again verified the feasibility of developing a specific mini-barcode from *matK*.

We recommend that *matK* can be regarded as a preferred option for specific mini-barcode development in subsequent studies.

Primer bias is a well-known factor that substantially influences results in metabarcoding studies. For instance, primer bias has been shown to skew the relative abundance of amplified DNA from experimental mixtures such as stonefly species and some terrestrial arthropods (Elbrecht and Leese, 2015; Piñol et al., 2015). Although our results showed that both primer 647F-847R and 478F-629R provided a positive relationship between species reads abundance and species biomass abundance in the experimental mixtures, which was consistent with other metabarcoding studies (Thomsen et al., 2012; Evans et al., 2016). However, the primer 478F-629R exhibited the large difference between observed reads proportion and given biomass proportion in JM1 and JM4 (i.e., biomass1 vs. 478F-629R1 and biomass4 vs. 478F-629R4). This large variability is most likely resulted from primer bias, and primer binding efficiency and inherent species-specific differences could lead to primer bias in metabarcoding studies (Duke and Burton, 2020). Although the quantitative ability of metabarcoding has always been tested due to primer bias (Elbrecht and Leese, 2015; Bista et al., 2018), our results showed that metabarcoding could provide a relatively accurate quantitative relationship between species read abundance and species abundance when there was no primer bias inference (e.g., primer 647F-847R).

The currently available quality control assessment methods for herbal products are mainly chemical methods, such as thin-layer chromatography, high-performance liquid chromatography, and mass spectrometry (Turova et al., 2018; Mukherjee, 2019). However, the chemical properties of herbal products can be affected by many factors, such as processing methods, storage conditions, and geographic location (Parveen et al., 2016), posing difficulties for proper chemical analyses and objective judgment (de Boer et al., 2015). Moreover, chemical methods may not be able to distinguish closely related species since they generally share chemical compounds (Raclariu et al., 2017), making it challenging to find representative chemical markers for authentication. Compared with chemical methods, DNA mini-barcoding can provide stable quality control assessment for herbal products due to the stability of DNA, and its short barcode region can overcome the difficulties caused by DNA degradation during processing and manufacturing of herbal products. Besides, our study has proved that specific mini-barcodes combined with the metabarcoding technique can realize the qualitative and quantitative identification of closely related species (i.e., *S. obtusifolia* and *S. occidentalis*). From an economic perspective, in our experiment, genomic DNA extraction takes about 1 h, PCR amplification takes about 2 h, and sequencing relies on the Illumina HiSeq X Ten platform, which takes about three days. The cost of each sample is about \$100, of which the main cost is library construction and data generation (about \$80). However, in the past 20 years, the cost per megabase of DNA sequencing has dropped from about \$10,000,000 to about \$0.01 (National Human Genome Research Institute, 2020), and it will continue to decline in the future. Moreover, with the introduction of portable sequencers,

such as MinION and DNBSEQ E series, the sequencing time has been greatly shortened, and even real-time sequencing can be achieved. Thus, it is conceivable that this molecular technique will become an efficient and economical method for the quality control of herbal products driven by technological development in the near future.

It has always been a concern of the herbal products market whether there is adulteration and how much adulteration since it is not uncommon that the adulterants of expensive or shortage materials are not only similar but also cheaper or easily available. A specific DNA mini-barcode combined with the metabarcoding technique can realize the qualitative and quantitative identification of samples with DNA degradation, which is suitable for processed herbal products like *S. obtusifolia* seeds since they are frequently adulterated with *S. occidentalis* seeds and generally suffer from varying degrees of DNA degradation during harvesting, storage and processing. Nevertheless, not all herbal products require the mini barcodes for quality control assessment. Whether a mini barcode or a regular barcode is needed depends largely on the degree of DNA degradation. Besides, whether specific or universal barcodes are required is also a question to be considered since the quantitative ability of barcodes is greatly interfered by primer bias. Therefore, we recommend that in follow-up studies, different types of herbal products need to be evaluated for the degree of DNA degradation and the suitability of primers to find an optimal quality control strategy.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/genbank/>, MK817504, <https://www.ncbi.nlm.nih.gov/genbank/>, MK817505.

AUTHOR CONTRIBUTIONS

XT and LM designed the study; XY and WT assembled, annotated and analyzed the plastomes; HG and XY performed the experiment; XY drafted the manuscript; XT and LM revised the manuscript.

FUNDING

This work is supported by the grants from Science and Technology Program of Tianjin (No.19ZYPTJC00060), China.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2020.585687/full#supplementary-material>.

REFERENCES

- Amend, A. S., Seifert, K. A., and Bruns, T. D. (2010). Quantifying microbial communities with 454 pyrosequencing: does read abundance count?. *Mol. Ecol.* 19 (24), 5555–5565. doi:10.1111/j.1365-294X.2010.04898.x
- Amritha, N., Bhooma, V., and Parani, M. (2020). Authentication of the market samples of Ashwagandha by DNA barcoding reveals that powders are significantly more adulterated than roots. *J. Ethnopharmacol.* 256, 112725. doi:10.1016/j.jep.2020.112725
- Andersen, K. S., Kirkegaard, R. H., Karst, S. M., and Albertsen, M. (2018). Ampvis2: an R package to analyse and visualise 16S rRNA amplicon data. Available at: <https://www.biorxiv.org/content/10.1101/299537v1.article-info>.
- Aronesty, E. (2013). Comparison of sequencing utility programs. *Open Bioinf. J.* 7, 1–8. doi:10.2174/1875036201307010001
- Bista, I., Carvalho, G. R., Tang, M., Walsh, K., Zhou, X., Hajibabaei, M., et al. (2018). Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples. *Mol. Ecol. Resour.* 8, 1020–1034. doi:10.1111/1755-0998.12888
- Bortiri, E., Coleman-Derr, D., Lazo, G. R., Anderson, O. D., and Gu, Y. Q. (2008). The complete chloroplast genome sequence of *Brachypodium distachyon*: sequence comparison and phylogenetic analysis of eight grass plastomes. *BMC Res. Notes* 1, 61. doi:10.1186/1756-0500-1-61
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., and Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13 (7), 581–583. doi:10.1038/nmeth.3869
- Chen, X., Zhou, J., Cui, Y., Wang, Y., Duan, B., and Yao, H. (2018). Identification of ligularia herbs using the complete chloroplast genome as a super-barcode. *Front. Pharmacol.* 9, 695. doi:10.3389/fphar.2018.00695
- Cheng, 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
- 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
- Costa, J., Campos, B., Amaral, J. S., Nunes, M. E., Oliveira, M. B. P. P., and Mafra, I. (2016). HRM analysis targeting ITS1 and matK loci as potential DNA mini-barcode for the authentication of *Hypericum perforatum* and *Hypericum androsaemum* in herbal infusions. *Food Contr.* 61, 105–114. doi:10.1016/j.foodcont.2015.09.035
- de Boer, H. J., Ichim, M. C., and Newmaster, S. G. (2015). DNA barcoding and pharmacovigilance of herbal medicines. *Drug Saf.* 38 (7), 611–620. doi:10.1007/s40264-015-0306-8
- Dong, W., Liu, H., Xu, C., Zuo, Y., Chen, Z., and Zhou, S. (2014). A chloroplast genomic strategy for designing taxon specific DNA mini-barcode: a case study on ginsengs. *BMC Genet.* 15, 138. doi:10.1186/s12863-014-0138-z
- Dong, W., Xu, C., Cheng, T., Lin, K., and Zhou, S. (2013). Sequencing angiosperm plastid genomes made easy: a complete set of universal primers and a case study on the phylogeny of saxifragales. *Genome Biol Evol* 5 (5), 989–997. doi:10.1093/gbe/evt063
- Duke, E. M., and Burton, R. S. (2020). Efficacy of metabarcoding for identification of fish eggs evaluated with mock communities. *Ecol. Evol.* 10 (7), 3463–3476. doi:10.1002/ece3.6144
- Egge, E., Bittner, L., Andersen, T., Audic, S., de Vargas, C., and Edvardsen, B. (2013). 454 pyrosequencing to describe microbial eukaryotic community composition, diversity and relative abundance: a test for marine haptophytes. *PLoS One* 8 (9), e74371. doi:10.1371/journal.pone.0074371
- Elbrecht, V., and Leese, F. (2015). Can DNA-based ecosystem Assessments quantify species abundance? Testing primer bias and biomass–sequence relationships with an innovative metabarcoding protocol. *PLoS One* 10 (7), e0130324. doi:10.1371/journal.pone.0130324
- Evans, N. T., Olds, B. P., Renshaw, M. A., Turner, C. R., Li, Y., Jerde, C. L., et al. (2016). Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. *Mol. Ecol. Resour.* 16 (1), 29–41. doi:10.1111/1755-0998.12433
- Gao, Z., Liu, Y., Wang, X., Wei, X., and Han, J. (2019). DNA mini-barcoding: a derived barcoding method for herbal molecular identification. *Front. Plant Sci.* 10, 987. doi:10.3389/fpls.2019.00987
- Gaudiano, M. C., Manna, L., Bartolomei, M., Rodomonte, A. L., Bertocchi, P., Antonielli, E., et al. (2016). Health risks related to illegal and on-line sale of drugs and food supplements: results of a survey on marketed products in Italy from 2011 to 2013. *Ann. Ist. Super Sanita* 52 (1), 128–132. doi:10.4415/ann_16_01_21
- Gromek, K., Drummond, N., and Simas, P. (2015). Pharmacovigilance of herbal medicines. *Int. J. Risk Saf. Med.* 27 (2), 55–65. doi:10.3233/jrs-150643
- Hebert, P. D., Cywinska, A., Ball, S. L., and deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proc. Biol. Sci.* 270 (1512), 313–321. doi:10.1098/rspb.2002.2218
- 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 (5), e0156426. doi:10.1371/journal.pone.0156426
- Ji, S. G., Pan, S. L., Wang, J., and Huo, K. K. (2007). [Phylogeny relationship and molecular identification of ten *Huperzia* species (*Huperziaceae*) based on matK gene sequences]. *Zhongguo Zhongyao Zazhi* 32 (19), 1971–1975.
- Jiang, D., Zhao, Z., Zhang, T., Zhong, W., Liu, C., Yuan, Q., et al. (2017). The chloroplast genome sequence of *scutellaria baicalensis* provides insight into intraspecific and interspecific chloroplast genome diversity in *scutellaria*. *Genes* 8 (9), 227. doi:10.3390/genes8090227
- Jiao, L., Lu, Y., He, T., Li, J., and Yin, Y. (2019). A strategy for developing high-resolution DNA barcodes for species discrimination of wood specimens using the complete chloroplast genome of three *Pterocarpus* species. *Planta* 250 (1), 95–104. doi:10.1007/s00425-019-03150-1
- Kechin, A., Boyarskikh, U., Kel, A., and Filipenko, M. (2017). cutPrimers: a new tool for accurate cutting of primers from reads of targeted next generation sequencing. *J. Comput. Biol.* 24 (11), 1138–1143. doi:10.1089/cmb.2017.0096
- Kim, D. H., Kim, S., Jung, W. Y., Park, S. J., Park, D. H., Kim, J. M., et al. (2009). The neuroprotective effects of the seeds of *Cassia obtusifolia* on transient cerebral global ischemia in mice. *Food Chem. Toxicol.* 47 (7), 1473–1479. doi:10.1016/j.fct.2009.03.028
- Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41 (1), e1. doi:10.1093/nar/gks808
- Lahaye, R., van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., et al. (2008). DNA barcoding the floras of biodiversity hotspots. *Proc. Natl. Acad. Sci. U. S. A.* 105 (8), 2923–2928. doi:10.1073/pnas.0709936105
- Li, B., Lin, F., Huang, P., Guo, W., and Zheng, Y. (2017). Complete chloroplast genome sequence of *decaisnea insignis*: genome organization, genomic resources and comparative analysis. *Sci. Rep.* 7 (1), 10073. doi:10.1038/s41598-017-10409-8
- Li, L., Jiang, Y., Liu, Y., Niu, Z., Xue, Q., Liu, W., et al. (2020). The large single-copy (LSC) region functions as a highly effective and efficient molecular marker for accurate authentication of medicinal *Dendrobium* species. *Acta Pharm. Sin. B* 10, 1989–2001. doi:10.1016/j.apsb.2020.01.012
- Li, X. E., and Guo, B. J. (2002). [Effect of protein and anthraquinone glucosides from cassia seed on serum lipid of hyperlipidemia rats]. *Zhongguo Zhongyao Zazhi* 27 (5), 374–376.
- Little, D. P. (2014a). Authentication of *Ginkgo biloba* herbal dietary supplements using DNA barcoding. *Genome* 57 (9), 513–516. doi:10.1139/gen-2014-0130
- Little, D. P. (2014b). A DNA mini-barcode for land plants. *Mol. Ecol. Resour.* 14 (3), 437–446. doi:10.1111/1755-0998.12194
- Liu, X. F., Zhu, G. F., Li, D. M., and Wang, X. J. (2019). Complete chloroplast genome sequence and phylogenetic analysis of *Spathiphyllum 'Parrish'*. *PLoS One* 14 (10), e0224038. doi:10.1371/journal.pone.0224038
- Lohse, M., Drechsel, O., Kahlau, S., and Bock, R. (2013). OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* 41, W575–W581. doi:10.1093/nar/gkt289
- Mackey, T. K., and Liang, B. A. (2013). Improving global health governance to combat counterfeit medicines: a proposal for a UNODC-WHO-Interpol trilateral mechanism. *BMC Med.* 11, 233. doi:10.1186/1741-7015-11-233

- Mosa, K. A., Soliman, S., El-Keblawy, A., Ali, M. A., Hassan, H. A., Tamim, A. A. B., et al. (2018). Using DNA barcoding to detect adulteration in different herbal plant-based products in the United Arab Emirates: proof of concept and validation. *Recent Pat. Food, Nutr. Agric.* 9 (1), 55–64. doi:10.2174/2212798410666180409101714
- Mukherjee, P. K. (2019). "Thin-layer chromatography for evaluation of herbal drugs," in *Quality control and evaluation of herbal drugs*. (Cambridge, MA: Elsevier), 329–376.
- Naem, A., Khan, A. A., Cheema, H. M., Khan, I. A., and Buerkert, A. (2014). DNA barcoding for species identification in the Palmae family. *Genet. Mol. Res.* 13 (4), 10341–10348. doi:10.4238/2014.December.4.29
- National Human Genome Research Institute (2020). DNA sequencing costs: data. Available at: <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data> (Accessed November 10, 2020).
- Niu, Z., Xue, Q., Wang, H., Xie, X., Zhu, S., Liu, W., et al. (2017). Mutational biases and GC-biased gene conversion affect GC content in the plastomes of dendrobium genus. *Int. J. Mol. Sci.* 18 (11). doi:10.3390/ijms18112307
- Ohsako, T., and Ohnishi, O. (2000). Intra- and interspecific phylogeny of wild Fagopyrum (Polygonaceae) species based on nucleotide sequences of noncoding regions in chloroplast DNA. *Am. J. Bot.* 87 (4), 573–582.
- Park, I., Yang, S., Kim, W. J., Noh, P., Lee, H. O., and Moon, B. C. (2018). The complete chloroplast genomes of six Ipomoea species and indel marker development for the discrimination of authentic pharbitidis semen (seeds of I. Nil or I. Purpurea). *Front. Plant Sci.* 9, 965. doi:10.3389/fpls.2018.00965
- Parveen, I., Gafner, S., Techen, N., Murch, S. J., and Khan, I. A. (2016). DNA barcoding for the identification of botanicals in herbal medicine and dietary supplements: strengths and limitations. *Planta Med.* 82 (14), 1225–1235. doi:10.1055/s-0042-111208
- Piñol, J., Mir, G., Gomez-Polo, P., and Agustí, N. (2015). Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Mol. Ecol. Resour.* 15 (4), 819–830. doi:10.1111/1755-0998.12355
- 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 (1), 1291. doi:10.1038/s41598-017-01389-w
- Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34 (12), 3299–3302. doi:10.1093/molbev/msx248
- Schnell, I. B., Bohmann, K., and Gilbert, M. T. (2015). Tag jumps illuminated--reducing sequence-to-sample misidentifications in metabarcoding studies. *Mol. Ecol. Resour.* 15 (6), 1289–1303. doi:10.1111/1755-0998.12402
- Srirama, R., Gurumurthy, B. R., Senthilkumar, U., Ravikanth, G., Uma Shaanker, R., and Shivanna, M. B. (2014). Are mini DNA-barcodes sufficiently informative to resolve species identities? An in silico analysis using *Phyllanthus*. *J. Genet.* 93 (3), 823–829. doi:10.1007/s12041-014-0432-6
- Staats, M., Arulandhu, A. J., Gravendeel, B., Holst-Jensen, A., Scholtens, I., Peelen, T., et al. (2016). Advances in DNA metabarcoding for food and wildlife forensic species identification. *Anal. Bioanal. Chem.* 408 (17), 4615–4630. doi:10.1007/s00216-016-9595-8
- Stanford, A. M., Harden, R., and Parks, C. R. (2000). Phylogeny and biogeography of Juglans (Juglandaceae) based on matK and ITS sequence data. *Am. J. Bot.* 87 (6), 872–882.
- Sun, X. Q., Zhu, Y. J., Guo, J. L., Peng, B., Bai, M. M., and Hang, Y. Y. (2012). DNA barcoding the Dioscorea in China, a vital group in the evolution of monocotyledon: use of matK gene for species discrimination. *PLoS One* 7 (2), e32057. doi:10.1371/journal.pone.0032057
- Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., and Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* 21 (8), 2045–2050. doi:10.1111/j.1365-294X.2012.05470.x
- Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., et al. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* 35 (3), e14. doi:10.1093/nar/gkl938
- Thomsen, P. F., Kielgast, J., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T., et al. (2012). Monitoring endangered freshwater biodiversity using environmental DNA. *Mol. Ecol.* 21 (11), 2565–2573. doi:10.1111/j.1365-294X.2011.05418.x
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45 (W1), W6–W11. doi:10.1093/nar/gkx391
- Turova, P., Stekolshchikova, E., Baygildiev, T., Shpigun, O., Rodin, I., and Stavrianidi, A. (2018). Unified strategy for HPLC-MS evaluation of bioactive compounds for quality control of herbal products. *Biomed. Chromatogr.* 32 (12), e4363. doi:10.1002/bmc.4363
- Viglietti, G., Galla, G., Porceddu, A., Barcaccia, G., Curk, F., Luro, F., et al. (2019). Karyological analysis and DNA barcoding of pompia citron: a first step toward the identification of its relatives. *Plants* 8 (4), 83. doi:10.3390/plants8040083
- Xia, W., Zhang, B., Xing, D., Li, Y., Wu, W., Xiao, Y., et al. (2019). Development of high-resolution DNA barcodes for Dioscorea species discrimination and phylogenetic analysis. *Ecol. Evol.* 9 (18), 10843–10853. doi:10.1002/ece3.5605
- Xie, S., Zhu, Q., Qu, W., Xu, Z., Liu, X., Li, X., et al. (2019). sRNAPrimerDB: comprehensive primer design and search web service for small non-coding RNAs. *Bioinformatics* 35 (9), 1566–1572. doi:10.1093/bioinformatics/bty852
- Yang, J., Ye, H., Lai, H., Li, S., He, S., Zhong, S., et al. (2012). Separation of anthraquinone compounds from the seed of *Cassia obtusifolia* L. using recycling counter-current chromatography. *J. Separ. Sci.* 35 (2), 256–262. doi:10.1002/jssc.201100535
- Yu, X., Tan, W., Zhang, H., Gao, H., Wang, W., and Tian, X. (2019). Complete chloroplast genomes of *ampelopsis humulifolia* and *ampelopsis japonica*: molecular structure, comparative analysis, and phylogenetic analysis. *Plants* 8 (10), 410. doi:10.3390/plants8100410
- Zhang, D., Gao, F., Jakovlić, I., Zou, H., Zhang, J., Li, W. X., et al. (2020). PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour.* 20 (1), 348–355. doi:10.1111/1755-0998.13096
- Zhang, X., Li, N., Yao, Y., Liang, X., Qu, X., Liu, X., et al. (2016). Identification of species in tripterygium (celastraceae) based on DNA barcoding. *Biol. Pharm. Bull.* 39 (11), 1760–1766. doi:10.1248/bpb.b15-00956

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

Copyright © 2020 Yu, Tan, Gao, Miao and Tian. 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) and the copyright owner(s) 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.