Characterization and novel Est-SSR marker development of an important Chinese medicinal plant, *Morinda officinalis* How (Rubiaceae)

Boyong Liao  
Shiou Yih Lee, Kaikai Meng, Qianyi Yin, Cuiying Huang, Qiang Fan  
State Key Laboratory of Bio-control, Guangdong Provincial Key Laboratory of Plant Resources, School of Life Sciences, Sun Yat-sen University, Guangzhou, Guangdong, P. R. China  

**ABSTRACT**  
*M. officinalis* How (Rubiaceae) is well known as one of the four most popular medicinal plants in the southern region of China. In order to assess its genetic diversity and assist the development of its molecular assisted-selection breeding in China, we constructed the transcriptome database of *M. officinalis* using fresh leaf material and developed novel expressed sequence tag–simple sequence repeat (EST-SSR) markers for this important medicinal plant species. Our study identified 8064 potential EST-SSRs from 6668 unigenes. For perfect repeat motifs, the mononucleotide repeats were most abundant (4221; 52.34%), followed by the trinucleotide repeats (1588; 19.69%), and the dinucleotide ones (1040; 12.90%). The seven most commonly occurring repeat motifs were A/T (4661), AG/CT (588), AAG/CTT (408), AC/GT (353), ATC/ATG (338), AT/AT (296) and AGG/CCT (277). Of a total of 5605 SSR primer sets that were successfully designed *de novo*, 100 were selected for validation analysis. There, 24 primer sets were polymorphic across the 96 *M. officinalis* individuals collected from four different populations in Guangdong province. Based on cross-species transferability test, 19 and 22 primer sets produced successful amplification in *M. umbellata* and *M. parvifolia*, respectively. These novel EST-SSR markers with high cross-species transferability between *M. officinalis* and its closely-related species could serve as aids to future studies in genetic diversity, conservation, assisted-selection breeding and genes associated with the medical uses of the genus.

**INTRODUCTION**

*M. officinalis* How (Rubiaceae) is a perennial vine plant, naturally found in the wild in the Indo-Malaysian region, i.e. in China, India, Japan, Philippines, Sri Lanka, Thailand, Vietnam and other countries [1,2]. Its long succulent roots were harvested and air-dried to be made into Chinese traditional medicine, *Bajitian*, currently known as one of the four major southern medicinal materials in China. Nowadays, the species is widely applied in the treatment of male impotence, female infertility, irregular menstruation, hypogastric-cold pains and rheumatoid arthritis [3,4].

As a medicinal source, the roots of *Bajitian* contain diverse chemical components, such as polysaccharides, anthraquinones, iridoids, amino acids and microelements, which have anti-aging, anti-osteoporosis, immune-enhancement properties and others [5–7]. Today, mass cultivation of *M. officinalis* has been promoted in many places of the southern regions of China, yet its wild populations have been greatly threatened due to unsustainable harvesting activities [2]. A recent assessment carried out by Zhang et al. [8] reported the natural population of *M. officinalis* to be on the edge of extinction, especially in Guangdong Province. It was listed as a ‘Category III’ protected species in the Chinese National Key Protected Wild Plants [9]. Although researches on *M. officinalis* have mostly centered on its pharmacological properties, He et al. [10] proposed that *M. officinalis* growing in different regions exhibited variable phenotypic characters in its chemical properties, which was further supported by Zhu et al. [11] and Zhou et al. [12]. More and more studies have begun to pay attention to its molecular mechanism of action. For example, Ding et al. [13]...
used random amplified polymorphism DNA (RAPD) markers to investigate the genetic diversity of 5 populations of *M. officinalis* in Guangdong Province of China, and the intergenic spacer region *psbA-trnH*, and internal transcribed spacer (ITS) were further applied in the identification of specific cultivars of *Bajitian* [14,15]. Recently the complete chloroplast genome structure of this medicinal plant was characterized [16]. In order to design suitable conservation plans for the natural populations and identify superior breeding materials among these cultivars, an appropriate fingerprinting technique is required in looking into the intra-specific variation across *M. officinalis* natural populations and cultivars, as well as identifying gene pools and producing useful markers for future plant breeding programs.

At present, the advanced development in life science technologies has allowed researchers to obtain complete genomes of non-model plants, providing convenience to identify and work on regulating genes toward secondary metabolism biosynthetic pathways, thus obtaining information of their potential active ingredients. Although pharmacologically related studies are well-documented for *M. officinalis*, to date, there is no transcriptomic work reported for this valuable medicinal plant. Therefore, in this study, we aimed to establish the first transcriptome library for *M. officinalis* using next generation sequencing (NGS). In addition, we took the initiative to develop novel expressed sequence tag simple sequence repeat (EST-SSR) primers for *M. officinalis*, as well as to test on their cross-transferability between its closely-related species, *M. parvifolia* and *M. umbellata* subsp. *ovovata*.

**Materials and methods**

**Plant materials**

For the transcriptomic study, fresh leaf samples were collected from an ex-situ *M. officinalis* tree planted in the greenhouse of Sun Yat-sen University, originating from a natural population in Deqing (DQ), Guangdong province. For the EST-SSR study, fresh leaf samples were collected from 96 individuals of four *M. officinalis* natural populations (Conghua (CH), *n* = 26; Dianbai (DB), *n* = 19; DQ, *n* = 26; Fengkai (FK), *n* = 25), whereas fresh leaf samples from six individuals were collected for *M. parvifolia* and *M. umbellata* subsp. *ovovata*, both originating from natural populations in FK, Guangdong province. All collected fresh leaf samples were kept in zip-lock bags filled with silica gel prior to transfer back to the laboratory. Voucher specimens were deposited at the Herbarium of Sun Yat-sen University (Supplemental Table S1).

**DNA extraction**

A total of 150 g fresh leaf samples were pulverized into fine powder using mortar and pestle, with the aid of liquid nitrogen. DNA extraction was carried out using the cetyltrimethylammonium bromide (CTAB) method [17]. The extracted DNA was quantified using Nano-drop™ (Thermo Fisher, USA) and stored at −20°C until further use.

**RNA extraction**

A total of 150 g fresh leaf samples were pulverized into fine powder using mortar and pestle, with the aid of liquid nitrogen. Total RNA was isolated based on a modified CTAB method. During the extraction process, residual DNA was treated with DNase I (Takara Biotechnology, Japan) for 45 min at 37°C, and the quality and concentration of the extracted total RNA was visualized under 2.0% agarose gel, quantified using Nano-drop™ (Thermo Fisher, USA).

**De novo assembly and unigene annotation**

The cDNA library was constructed using TruSeq RNA Library Preparation Kit version 2.0 (Illumina, USA) based on the manufacturer’s protocol and sequenced using HiSeq 2000 sequencing system (Illumina, USA). Amplicons were sequenced from both paired ends to obtain 125-bp short reads. Raw reads were filtered by eliminating reads containing ‘N’ bases or more than 10% bases with a Q-value less than 20 using NGS toolkit V 2.3.3 [18]. Subsequently, all the clean reads were de novo-assembled to construct contigs using Trinity v 2.4.0 [19] based on its default settings; then the contigs were connected to form transcripts. Paired-end reads were incorporated to fill up the gap between these transcripts in order to obtain unigenes that were later deposited to the NCBI GenBank as short read archive (SRA; Accession no. SRP149778).

Unigenes were annotated by matching against the Gene Ontology (GO) database (http://www.geneontology.org), based on the condition where by the e-value cut-off was 1.0e−5, and provided a minimum of 20 best hits, using Diamond software [20]. Web Gene Ontology (WEGO) plot annotation was conducted to perform GO functional classification and to illustrate the distribution of gene functions at a macro-level [21].

**EST-SSR primers development**

Short sequence repeats (SSR) with more than five repeat motifs within the assembled unigenes were
identified using MicroSatellite identification tool (MISA, http://pgrc.ipk-gatersleben.de/misa/). A total of 5605 primer sets were designed successful by MISA (Supplemental Table S2). The primers were designed by MISA based on the following parameters: (1) primer length ranging between 18 and 23 bp, with 21 bp as optimum; (2) amplicon size ranging between 100 and 300 bp; (3) an optimum annealing temperature of 52–65°C; and (4) a percentage of GC content at 40–60%, with 50% as optimum. One hundred SSRs were randomly selected from the 5605 primers to design SSR-specific markers using Primer3 (http://pgrc.ipk-gatersleben.de/misa/primer3.html) by incorporating the online Perl scripts [22]. The selected 100 primer sets were synthesized (Sangon Biotech (Shanghai) Co., Ltd., Guangzhou, China) for polymerase chain reaction (PCR) amplification in the subsequent experiments.

SSR-PCR amplification and cross-transferability test

PCR solution was prepared in a total volume of 20.0 μL, containing 10× EasyTaq® Buffer (containing 200 mmol/L Tris-HCl pH 8.3, 200 mmol/L KCl, 100 mmol/L (NH₄)₂SO₄, 20 mmol/L MgSO₄; TransGen Biotech, China), 0.25 mmol/L of each deoxyribonucleoside triphosphate (dNTP), 0.2 μmol/L of each primer (forward and reverse), 1 U of EasyTaq® DNA polymerase (TransGen Biotech, China), and 25 ng of genomic DNA template. The PCR amplification was conducted in a T100™ thermal cycler (Bio-Rad, USA), programmed at the following settings: an initial denaturing step at 94°C for 2 min; followed by 35 cycles of 94°C for 30 s, annealing temperature (Tₐ) for 30 s, and 72°C for 40 s; and a final extension at 72°C for 5 min. For the EST-SSR primer set PCR amplification test, four M. officinalis individuals were used to assess if the primers were working, and at the same time to check if the amplicons yielded within the range of the expected band size in a 2.0% agarose gel, prior to the amplicons yielded within the range of the mers were working, and at the same time to check if M. officinalis For the EST-SSR primer set PCR amplification test, four individuals were used to assess if the primers were working, and at the same time to check if the amplicons yielded within the range of the expected band size in a 2.0% agarose gel, prior to the amplicons yielded within the range of the mers were working, and at the same time to check if M. officinalis

Results and discussion

In China, the cultivation of M. officinalis was first reported in Deqing of Guangdong province and Nanjing of Fujian province as a renewable medicinal resource. Yet its wild populations were greatly threatened by large-scale wild harvesting activities [26]. Previous studies of M. officinalis in part of the southern region of China based on RAPD and ISSR markers, and ITS and chloroplast sequencing showed the apparent genetic structure in M. officinalis [26,27]. Researches on chemical compounds showed that populations collected from different regions always demonstrated obvious variations in their morphological characters and chemical compositions [28]. These researches also showed that the wild resources of M. officinalis may contain ample useful chemical compounds, which calls for urgent protection of its natural populations and scientific researches on its genetic resources. Therefore, by utilizing the high-throughput sequencing technology, we sequenced the transcriptome of M. officinalis, producing tens of thousands of genetic sequences, which could be helpful for studies on gene functions [29]. These novel SSR markers could be greatly useful in further studies on the genetic diversity of M. officinalis to lay out plans for the protection of the species.

Characterization of Est-SSRs

This is the first report on the transcriptome and development of the working EST-SSR primer sets for M. officinalis using NGS. A total of 6668 selected unigenes containing SSRs were identified from 51049 examined sequences (Table 1). The average length and N₅₀ value of the unigenes (801 bp and 1587 bp) were at least comparable to other species of the same family, such as Coffea canephora (662 bp and 868 bp) and C.arabica (663 bp and 832 bp) [30], thus indicating that the transcript quality was satisfactory in this study. Comparing to other species of Rubiaceae, the EST-SSRs frequency

### Table 1. Summary of the SSR searches for the unigenes of M. officinalis using MISA.

| Source | Number |
|--------|--------|
| Total number of sequences examined | 51049 |
| Total size of examined sequences (bp) | 40900392 |
| Total number of identified SSRs | 8064 |
| Number of SSR-containing sequences | 6668 |
| Number of sequences containing more than 1 SSR | 1125 |
| Number of SSRs present in compound formation | 494 |

and expected heterozygosity (He) were estimated using GenAlEx 6.501 [25].
for *M. officinalis* was 1 per 5.084 kb (8064 SSRs in 40.9 Mb), which was higher than *C. canephora* (1/7.73 kb), and lower than *Gardenia jasminoides* (1/3.516 kb) [31,32].

A total of 8064 SSRs were identified from 6668 selected unigenes (Table 1). Among the identified SSRs, 7570 were perfect repeats (Supplemental Table S2), and 494 were imperfect repeats (Table 2). A total of 457 SSRs were perfect compound repeats. Among these perfect SSRs, mononucleotide repeats (52.34%) were the most common, followed by trinucleotide (19.69%), dinucleotide (12.90%), tetranucleotide (1.98%), pentanucleotide (0.79%) and hexanucleotide (0.50%) repeats. Among all SSRs, the mononucleotide repeat units also had the biggest percentage number of repeat units and the trinucleotide repeat units ranked second with a percentage of 22.81% (1839), which was 1.47 times when compared to the dinucleotide motifs (15.45%) (Table 3). The repeat unit that came with the highest frequency was 10 (27.96%); the longest repeat motif was the mononucleotide A, with a total of 42 repeating bases (Table 3). Three trinucleotide motifs, AAG/CTT (408), ATC/ATG (338) and AGG/CCT (277), ranked three, five and seven in the greatest copy numbers of motif type. Yet it is known that shifts in trinucleotide SSRs, which are in multiple of three nucleotide bases, do not cause changes in EST reading frames as codons are represented with three nucleotide bases. Therefore, the consistent multiplication of trinucleotide SSRs without fail in *M. officinalis* indicated a positive selection pressure for the SSRs in genic sequences [33].

### Categories and annotations of unigenes containing SSRs

In this study, 6046 unigenes of the 6668 unigenes were annotated in the GO database and divided into

### Table 2. Frequencies of SSR repeat motif type distribution from *M. officinalis* transcriptome.

| Repeat motif type | Mono- | Di- | Tri- | Tetra- | Penta- | Hexa- | Total | Frequency (%) |
|-------------------|-------|-----|------|--------|--------|-------|-------|---------------|
| Perfect           | 4221  | 1040| 1588 | 160    | 64     | 40    | 7570  | 93.87%        |
| Imperfect         | 494   |     |      |        |        |       | 494   | 6.13%         |
| Total             | 8064  |     |      |        |        |       | 8064  | 100%          |

### Table 3. Distribution of different repeat sequences in EST-SSRs of *M. officinalis*.

| No. of repeat units | Mono- | Di- | Tri- | Tetra- | Penta- | Hexa- | Total | Frequency (%) |
|---------------------|-------|-----|------|--------|--------|-------|-------|---------------|
| 5                   | 0     | 0   | 941  | 109    | 47     | 21    | 1118  | 13.86%        |
| 6                   | 0     | 468 | 411  | 56     | 15     | 17    | 967   | 11.99%        |
| 7                   | 0     | 282 | 240  | 2      | 3      | 5     | 532   | 6.60%         |
| 8                   | 0     | 175 | 170  | 5      | 0      | 1     | 351   | 4.35%         |
| 9                   | 0     | 108 | 17   | 5      | 0      | 0     | 130   | 1.61%         |
| 10                  | 2151  | 81  | 19   | 2      | 0      | 1     | 2255  | 27.96%        |
| 11                  | 1020  | 36  | 20   | 0      | 0      | 0     | 1076  | 13.34%        |
| 12                  | 561   | 47  | 11   | 0      | 0      | 0     | 619   | 7.68%         |
| 13                  | 303   | 12  | 2    | 0      | 0      | 0     | 317   | 3.93%         |
| 14                  | 212   | 6   | 3    | 0      | 0      | 0     | 221   | 2.74%         |
| 15                  | 129   | 8   | 1    | 0      | 0      | 0     | 138   | 1.71%         |
| 16                  | 77    | 11  | 1    | 0      | 0      | 0     | 89    | 1.10%         |
| 17                  | 45    | 4   | 1    | 0      | 0      | 0     | 50    | 0.62%         |
| 18                  | 37    | 2   | 0    | 0      | 0      | 0     | 39    | 0.48%         |
| 19                  | 30    | 2   | 0    | 0      | 0      | 0     | 32    | 0.40%         |
| 20                  | 18    | 2   | 0    | 0      | 0      | 0     | 20    | 0.25%         |
| 21                  | 23    | 1   | 2    | 0      | 0      | 0     | 26    | 0.32%         |
| 22                  | 13    | 0   | 0    | 0      | 0      | 0     | 13    | 0.16%         |
| 23                  | 13    | 0   | 0    | 0      | 0      | 0     | 13    | 0.16%         |
| 24                  | 9     | 1   | 0    | 0      | 0      | 0     | 10    | 0.12%         |
| 25                  | 7     | 0   | 0    | 0      | 0      | 0     | 7     | 0.09%         |
| 26                  | 10    | 0   | 0    | 0      | 0      | 0     | 10    | 0.12%         |
| 27                  | 5     | 0   | 0    | 0      | 0      | 0     | 5     | 0.06%         |
| 28                  | 6     | 0   | 0    | 0      | 0      | 0     | 6     | 0.07%         |
| 29                  | 4     | 0   | 0    | 0      | 0      | 0     | 4     | 0.05%         |
| 30                  | 2     | 0   | 0    | 0      | 0      | 0     | 2     | 0.02%         |
| 31                  | 2     | 0   | 0    | 0      | 0      | 0     | 2     | 0.02%         |
| 32                  | 3     | 0   | 0    | 0      | 0      | 0     | 3     | 0.04%         |
| 33                  | 1     | 0   | 0    | 0      | 0      | 0     | 1     | 0.01%         |
| 34                  | 3     | 0   | 0    | 0      | 0      | 0     | 3     | 0.04%         |
| 35                  | 2     | 0   | 0    | 0      | 0      | 0     | 2     | 0.02%         |
| 37                  | 2     | 0   | 0    | 0      | 0      | 0     | 2     | 0.02%         |
| 42                  | 1     | 0   | 0    | 0      | 0      | 0     | 1     | 0.01%         |
| Total               | 4690  | 1246| 1839 | 179    | 65     | 45    | 8064  | 100.00%       |

| Frequency           | 58.16%| 15.45%| 22.81%| 2.22%  | 0.81%  | 0.56% | 100%  |               |
three major categories according to their functions, namely biological process, cellular component, and molecular function. In the biological process category, 25 different processes were identified, whereby most unigenes were involved in the cellular processes (n = 6041; 8.06%), followed by the cellular component organization or biogenesis (n = 5924; 7.90%), and metabolic process (n = 5887; 7.85%) (Figure 1a). For

Figure 1. Distribution of unigenes containing SSRs in M. officinalis transcriptome categorized according to their biological functions: biological process (a), cellular component (b) and molecular function (c), using the WEGO [21].
the cellular component category, 13 different cellular components were recorded and most unigenes were involved in the cell (n = 5515; 16.06%), followed by the cell parts (n = 5514; 16.06%) and organelles (n = 5190; 15.11%) (Figure 1b). In the molecular function category, nine different functions were recorded, whereby most unigenes were involved in catalytic activities (n = 5459; 33.45%), followed by binding (n = 4221; 25.87%) and transporter activity (n = 3237; 19.84%) (Figure 1c).

When we matched the *M. officinalis* transcriptome against the GO database, although some returned with low match identity, most of the genes were able to be correctly annotated. This showed that *M. officinalis* has low probability of presence of rare or unknown genes. As *bajitian*, the root of *M. officinalis*, is a raw medicinal material. We reported the presence of gene synthases, such as β-fructofuranosidase (EC:3.2.1.26), which is involved in inulin-type oligosaccharide, galactosum and sucrose (PATH: ko00500) pathways in metabolism. However, we seemed not able to detect several pharmacologically-related genes which are responsible for the production of chemical compounds such as polysaccharide, anthraquinones and terpenoids in *bajitian* [6,10,34]. As *bajitian* is processed from *M. officinalis* roots, whereas our specimen for the transcriptome sequencing was from its leaves, although the changes are small, it is possible that the genes responsible for these chemical compounds may be specific to the roots [35]. Alternatively, the differences in reported gene synthases could be caused by phenotypic changes, whereby a comprehensive comparison of bioactive compounds and pharmacological activities between different geographical origins could perhaps be able to link components to the related genes in this species [36].

**Polymorphism and cross-species transferability**

In this study, a total of 5605 primers were designed from the 6668 unigene sequences that contained SSRs. However, we randomly selected 100 primer sets from the list to check on their capability of successful amplification. As a result, 59 of them were recorded with successful amplification; while 24 of them showed polymorphisms in all four populations (Table 4). Often, EST-SSR primer sets failed to amplify due to a few reasons such as existing of large introns, primer sequences designed across splice sites, poor sequence quality, or assembly errors. The primer sequences used in this study were deposited in the NCBI GenBank under the accession numbers: MF496207-MF496230 (Supplemental Table S3). Based on this study, the linkage disequilibrium (LD) test showed no significantly linked pairs of primers after a Bonferroni correction, so these primers could be used

| Locus | DB (n = 19) | FK (n = 25) | CH (n = 26) | DQ (n = 26) | \( M. \) umbellata subsp. obovata |
|-------|------------|------------|------------|------------|----------------------------------|
| MO02  | 0.789 0.702 | 4 0.840 0.686 | 5 0.769 0.663 | 5 0.615 0.763 | + + |
| MO04  | 7 1.000 0.752 | **8 0.880 0.818** | 7 1.000 0.777 | **7 0.962 0.750** | + + |
| MO05  | 6 1.000 0.787 | **7 0.820 0.704** | 11 0.962 0.857 | **12 1.000 0.834** | + + |
| MO12  | 7 0.947 0.731 | **7 1.000 0.778** | 7 1.000 0.717 | **8 0.923 0.772** | + + |
| MO19  | 7 0.737 0.699 | **4 0.480 0.630** | 6 0.654 0.697 | **7 0.615 0.723** | + + |
| MO26  | 6 1.000 0.713 | **9 0.880 0.824** | 7 0.962 0.777 | **7 0.923 0.819** | + + |
| MO30  | 8 0.947 0.823 | **9 0.800 0.836** | 8 0.769 0.822 | **8 0.808 0.777** | -- -- |
| MO33  | 10 0.632 0.676 | **8 0.960 0.767** | 9 0.885 0.771 | **11 0.538 0.699** | -- -- |
| MO37  | 6 0.842 0.708 | **6 0.840 0.721** | 4 0.923 0.733 | **4 0.808 0.689** | + + |
| MO38  | 8 0.947 0.812 | **11 1.000 0.873** | 10 0.962 0.855 | **11 0.769 0.865** | + + |
| MO39  | 8 1.000 0.817 | **9 0.920 0.810** | 8 0.962 0.779 | **11 0.808 0.878** | + + |
| MO41  | 5 1.000 0.716 | **6 0.960 0.734** | 5 1.000 0.684 | **5 0.808 0.758** | + + |
| MO43  | 7 0.474 0.673 | **7 0.560 0.734** | 6 0.654 0.816 | **9 0.538 0.821** | + + |
| MO47  | 4 0.947 0.655 | **3 0.920 0.64** | 3 0.846 0.629 | **4 0.731 0.641** | + + |
| MO53  | 6 0.789 0.661 | **5 0.400 0.718** | **7 0.269 0.701** | **7 0.385 0.763** | + + |
| MO57  | 6 0.842 0.704 | **8 1.000 0.796** | **7 0.932 0.791** | **6 0.923 0.768** | + + |
| MO60  | 5 1.000 0.699 | **6 1.000 0.736** | 7 1.000 0.753 | **5 0.962 0.768** | + + |
| MO61  | 5 0.947 0.705 | **4 0.960 0.672** | 5 0.855 0.709 | **5 0.923 0.672** | + + |
| MO63  | 7 0.842 0.767 | **7 0.600 0.814** | 9 0.115 0.761 | **7 0.462 0.815** | + + |
| MO88  | 5 0.947 0.719 | **5 0.920 0.662** | 4 0.962 0.675 | **6 1.000 0.695** | + + |
| MO89  | 3 0.737 0.561 | **3 0.600 0.675** | 4 0.500 0.734 | **4 0.462 0.723** | + + |
| MO90  | 3 0.737 0.483 | **5 0.560 0.554** | 6 0.577 0.633 | **4 0.769 0.576** | + + |
| MO94  | 7 0.231 0.539 | **7 0.440 0.613** | 6 0.385 0.649 | **8 0.423 0.658** | + + |
| MO96  | 6 0.947 0.780 | **8 0.840 0.784** | 6 0.615 0.711 | **8 0.885 0.818** | + + |

Notes: DB, Dianbai; FK, Fengkai; CH, Conghua; DQ, Deqing. A, number of alleles; He, expected heterozygosity; Ho, observed heterozygosity; HWE, P value of Hardy–Weinberg test; ns, not significant; *P < 0.05, **P < 0.01, ***P < 0.001; n, number of individuals collected for each population. +, primers produced successful amplification in all individuals; --, primers could not produce amplicons in any individual.
alone to test the populations genetic parameters. The number of alleles (A) ranged from three to ten in the population DB, from three to eleven in the population FK and CH, and from Four to twelve in the population DQ. The Ho and He values ranged, respectively, from 0.211 to 1 and from 0.483 to 0.817 in population DB, from 0.400 to 1 and from 0.554 to 0.873 in population FK, from 0.115 to 1 and from 0.629 to 0.857 in population CH, and from 0.385 to 1 and from 0.576 to 0.878 in population DQ. HW tests showed that three populations of FK, CH and DQ significantly deviated from the Hardy–Weinberg equilibrium in most loci, and DB had most loci according to the Hardy–Weinberg equilibrium (Table 4).

For the cross-species transferability test, 19 primer set were able to amplify in both M. parvifolia and M. umbellata subsp. obovata (Table 4), while three primer sets (MO57, MO60 and MO63) were able to amplify in M. parvifolia but not in M. umbellata subsp. obovata. From this study, we believe that M. officinalis has a closer evolutionary relationship with M. parvifolia when compared to M. umbellata subsp. obovata. Also, based on the high success of the amplification rate in the cross-species transferability test, we foresee that the newly developed M. officinalis EST-SSR primer sets could work well on other Morinda species, as well as other members of the family Rubiaceae, thus contributing as a powerful genetic tool to provide useful genetic information for future biotechnology purposes.

Conclusions

EST-SSRs are known to be valuable gene function markers, which is useful in genetic diversity assessment across germplasm resources and marker-assisted breeding programs. Therefore, the M. officinalis transcriptome and the newly developed EST-SSR primer sets will be beneficial to future research on gene association with active pharmaceutical ingredients, genetic diversity, population structure and improving the understanding of the regulatory processes of pharmacologically-related compounds in M. officinalis, specifically towards the development of a renewable medicinal source for bajitian.

Acknowledgments

The authors thank Dr. Yelin Huang of Sun Yat-sen University for contributing ideas to this study and Mr. Weizheng Tan of Sun Yat-sen University for assisting in sample preparation.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Special Program for Science and Technology Basic Research of Ministry of Science and Technology of the People’s Republic of China (2013FY111500) and The Fourth National Survey on Chinese Material Medical Resources Program for State Administration of Traditional Chinese Medicine of the People’s Republic of China (2017-152-003). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ORCID

Boyong Liao http://orcid.org/0000-0003-1051-9810
Qiang Fan http://orcid.org/0000-0003-4254-6936
Wenbo Liao http://orcid.org/0000-0001-6839-9511
Sufang Chen http://orcid.org/0000-0003-3920-7618

References

[1] How FC. Description of a Chinese medical plant “Pa Chit T’ien”. J Univ Chin Acad Sci. 1958;7:325–328.
[2] Chen T, Luo XR, Zhu H, et al. 2003. Rubiaceae, Wu ZY, Raven PH, Hong DY, editors. Flora of China. Beijing, China: Science Press; St. Louis, MO: Missouri Botanical Garden Press; Vol. 19, pp. 220–230.
[3] Wu ZQ, Chen DL, Lin FH, et al. Effect of bajijiasu isolated from Morinda officinalis F. C. How on sexual function in male mice and its antioxidant protection of human sperm. J Ethnopharmacol. 2015;164:283–292.
[4] Zhang JH, Xin HL, Xu YM, et al. Morinda officinalis How. A comprehensive review of traditional uses, phytochemistry and pharmacology. J Ethnopharmacol. 2018;213:230–255.
[5] Yong JP, Lu CZ, Huang SJ, et al. Chemical components isolated from the roots of Morinda officinalis. Chem Nat Compd. 2015;51(3):548–549.
[6] Yang F, Su YF, Zhao ZQ, et al. Anthraquinones and iridoids from Morinda officinalis. Chem Nat Compd. 2016;52(6):989–991.
[7] Zhao XS, Wei JH, Yang MH. Simultaneous analysis of iridoid glycosides and anthraquinones in Morinda officinalis using UPLC-QqQ-MS/MS and UPLC-Q/TOF-MSE. Molecules. 2018;23(5):E1070.
[8] Zhang RJ, Li Q, Qu M, et al. Investigation on germplasm resources of Morinda officinalis How. Mod Chin Med. 2016;18(4):482–487.
[9] The Ministry of Agriculture and the State Administration of Forestry. National key protected wild plants (first batch). Beijing, China: The State Council of the People’s Republic of China; 1999.
[10] He CB, Li L, Tang FX, et al. Isolation and structure characterization of polysaccharide from Morinda officinalis How. Chem J Chin Univ. 2009;30:2391–2395.
[11] Zhu MY, Wang CJ, Gu Y, et al. Extraction, characterization of polysaccharides from *Morinda officinalis* and its antioxidant activities. Carbohydr Polym. 2009;78: 497–501.

[12] Zhou B, Chang J, Wang P, et al. Qualitative and quantitative analysis of seven oligosaccharides in *Morinda officinalis* using double-development HPTLC and scanning densitometry. Biomed Mater Eng. 2014;24(1): 953–960.

[13] Ding P, Liu J, Yang TC, et al. Genetic diversity of *Morinda officinalis* by RAPD. Chin Trad Herb Drug. 2008;39:1869–1872.

[14] Ding P, Liu J, Qiu JY, et al. Genuineness of *Morinda officinalis* How germplasm inferred from ITS sequences variation of nuclear ribosomal DNA. Yao Xue Xue Bao. 2012;47(4):535–540.

[15] Xu Z, Ran Z, Yang X, et al. Cluster analysis of *Morinda officinalis* How in different geographical populations based on *psb*-*trn*-*H* and rDNA ITS sequences. J Southern Agri. 2018;49(12):2364–2370.

[16] Zhang RJ, Li Q, Gao JL, et al. The complete chloroplast genome sequence of the medicinal plant *Morinda officinalis* (Rubiaceae), an endemic to China. Mitochondr DNA A. 2016;27(6):4324–4325.

[17] Doyle JJ, Doyle JL. A rapid DNA isolation procedure from small quantities of fresh leaf tissues. Phytochem Bull. 1987;19:11–15.

[18] Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS One. 2012;7(2):e30619.

[19] Haas BJ, Papanicolaou A, Yassour M, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8(8):1494–1512.

[20] Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12(1):59–60.

[21] Ye J, Zhang Y, Cui HH, et al. WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. Nucleic Acids Res. 2018;46(W1):W71–W75.

[22] Untergasser A, Cutcutache I, Koressaar T, et al. Primer3-new capabilities and interfaces. Nucleic Acids Res. 2012;40(15):e115.

[23] Roussset F. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Resour. 2008;8(1):103–106. 01931.x.

[24] Van Oosterhout C, Hutchinson WF, Wills DPM, et al. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes. 2004;4(3):535–538.

[25] Peakall R, Smouse PE. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. Bioinformatics. 2012;28(19): 2537–2539.

[26] Liu YJ, Huang Y, Rong JD, et al. Genetic diversity analysis of *Morinda officinalis* by ISSR markers. J Fujian Coll Forestry. 2011;31(3):203–206.

[27] Singh DR, Srivastava AK, Srivastava A, et al. Genetic diversity among three *Morinda* species using RAPD and ISSR markers. Ind J Biotechnol. 2011;10(3): 285–293. http://hdl.handle.net/123456789/12110.

[28] Ding P, Chu TL, Xu YJ. The chromatographic fingerprint of component of *Morinda officinalis* in different germplasms. West China J Phama Sci. 2006;21(1): 12–14.

[29] Singh AK, Chaurasia S, Kumar S, et al. Identification, analysis and development of salt responsive candidate gene based SSR markers in wheat. BMC Plant Biol. 2018;18(1):249.

[30] Mondego JM, Vidal RO, Carrazzolle MF, et al; Brazilian Coffee Genome Project Consortium. An EST-based analysis identifies new genes and reveals distinctive gene expression features of *Coffea arabica* and *Coffea canephora*. BMC Plant Biol. 2011;11(1):30.

[31] Poncet V, Rondeau M, Tranchant C, et al. SSR mining and analysis mapping and genomic variants screening. BMC Genomics. 2018;19(1):e291.

[32] Deng SY, Wang XR, Zhu PL, et al. Development of polymorphic microsatellite markers in the medicinal plant *Gardenia jasminoides* (Rubiaceae). Biochem Syst Ecol. 2015;58:149–155.

[33] Han ZZ, Ma XY, Wei M, et al. SSR marker development and inraspecific genetic divergence exploration of *Chrysanthemum indicum* based on transcriptome analysis. BMC Genomics. 2018;19(1):e291.

[34] Zhang HY, Miao HM, Wei LB, et al. Identification of a SiCL1 gene controlling leaf curling and capsule indehiscence in sesame via cross-population association mapping and genomic variants screening. BMC Plant Biol. 2018;18(1):e296.

[35] Jiang KM, Huang D, Zhang DW, et al. Investigation of inulins from the roots of *Morinda officinalis* for potential therapeutic application as anti-osteoporosis agent. Int J Biol Macromol. 2018;120:170–179. 08.082.

[36] Zhao M, Xu WF, Shen HY, et al. Comparison of bioactive components and pharmacological activities of *Ophiopogon japonicus* extracts from different geographical origins. J Pharm Biomed Anal. 2017;138: 134–141.