Chinese Journal of Natural Medicines

Molecular quantification of herbs (Herb–Q): a pyrosequencing–based approach and its application in *Pinellia ternata*

Yifei PEI, Ziyi LIU, Dade YU, Xiangyu ZHANG, Wei SUN, Xiaofang CHEN, Xue FENG, Xiwen LI

Citation: Yifei PEI, Ziyi LIU, Dade YU, Xiangyu ZHANG, Wei SUN, Xiaofang CHEN, Xue FENG, Xiwen LI, Molecular quantification of herbs (Herb–Q): a pyrosequencing–based approach and its application in *Pinellia ternata*, Chinese Journal of Natural Medicines, 2024, 22(7), 663–672. doi: 10.1016/S1875–5364(24)60594–7.

View online: https://doi.org/10.1016/S1875–5364(24)60594–7

Available online at www.sciencedirect.com
ScienceDirect
Chinese Journal of Natural Medicines 2024, **22**(7): 663-672
doi: 10.1016/S1875-5364(24)60594-7
Chinese Journal of Natural Medicines

•Original article•

Molecular quantification of herbs (Herb-Q): a pyrosequencing-based approach and its application in *Pinellia ternata*

PEI Yifei[1Δ], LIU Ziyi[1Δ], YU Dade[1], ZHANG Xiangyu[2], SUN Wei[1], CHEN Xiaofang[3], FENG Xue[1*], LI Xiwen[1*]

[1] *Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China;*
[2] *Bijie Institute of Traditional Chinese Medicine, Bijie 551700, China;*
[3] *The Key Laboratory for Health Industry of Bijie, Bijie Medical College, Bijie 551700, China*

Available online 20 Jul., 2024

[ABSTRACT] Variations in herb dosage due to species adulteration and dosing inaccuracies can substantially affect clinical safety and efficacy. Accurate species quantification remains challenging, as current methods often yield inconsistent results. This study introduces a novel pyrosequencing-based technique, termed herb molecular quantification (Herb-Q), designed to precisely quantify herbal products. We evaluated its effectiveness using *Pinellia ternata* and five of its adulterants. Initially, we assessed commonly used DNA barcodes with sequences from a public database, identifying two candidate regions, Maturase K (*mat*K) and internal transcribed spacer 2 (ITS2), for screening specific single nucleotide polymorphism (SNP) loci, allowing for species-specific identification. These loci were validated by amplifying and sequencing genomic material from collected samples. Our validation studies showed that Herb-Q demonstrated excellent linearity, accuracy, repeatability, and detection limits. We established quantitative standard curves with high R^2 values (> 0.99) to enable precise species quantification, which were combined with external standards to provide clear and accurate visual quantification results. The average bias in quantifying the tuber of *P. ternata* was 2.38%, confirming that Herb-Q can accurately identify and quantify herbal product constituents. Moreover, the entire quantification process took less than 4 h. This study presents a novel, rapid method for accurately quantifying species in herbal products and advances the application of DNA barcoding from species identification to quantitative detection.

[KEY WORDS] Quantification; Adulterant; DNA barcodes; Single nucleotide polymorphism
[CLC Number] R917 **[Document code]** A **[Article ID]** 2095-6975(2024)07-0663-10

Introduction

Herbal plants are integral to the medical industry, particularly in countries like China, India, and other regions where traditional medicine is widely practiced [1, 2]. Moreover, herbs constitute a primary ingredient in various dietary supplements, including pills, capsules, tablets, powders, and liquid formulations [3]. However, the average adulteration rate of herbal products (HPs is alarmingly high, averaging 27%, as evidenced by the analysis of 5957 types of HPs across 37 countries [4]. The excessive use of adulterants in HPs often leads to severe consequences [3, 5]. For example, ashwagandha, a widely used herb in Ayurvedic medicine and a common component in herbal dietary supplements, is generally considered safe; nevertheless, cases of liver injury have been reported, likely due to overdosage [6]. In some cases, adulterated HPs can cause significant harm; for instance, Clematidis Armandii Caulis and Aristolochiae Manshuriensis Caulis, which are adulterants of Akebiae Caulis, have been associated with renal failure and even death [3, 7]. Thus, accurate and reliable methods for herb species quantification are essential to mitigate the health risks associated with herbal product consumption.

Traditional methods such as microscopic examination, morphological analysis, and chemical constituent analysis

[Received on] 10-Feb.-2024
[Research funding] This work was supported by the Scientific and Technological Innovation Project of China Academy of Chinese Medical Sciences (Nos. CI2021A04106 and CI2021A03910), the National Key Research and Development Program of China (No. 2019YFC1710601), the Fundamental Research Funds for the Central Public Welfare Research Institutes of China (Nos. ZZ15-YQ-033, ZXKT21026 and ZXKT23004), the Major Special Project of Scientific and Technological Cooperation of Bijie (No. 2021-02), the Advantageous Chinese Medicinal Materials R&D Talent Base Project of Bijie, Guizhou Province (No. RCJD2020-21), and Bijie Technology Innovation Platform and Talent Team (Bikehe [2023] No.66-BJZDSYS 2024-05).
[*Corresponding author] Tel/Fax: 86-10-8408-4107, E-mails: xfeng0413@icmm.ac.cn (FENG Xue); xwli@icmm.ac.cn (LI Xiwen)
[Δ]These authors contributed equally to this work.
These authors have no conflict of interest to declare.

have limitations in accurately quantifying individual species, as they typically provide qualitative results or only quantitatively analyze certain components [8, 9]. In contrast, advanced barcoding-based technologies [8], including quantitative real-time polymerase chain reaction (PCR) [10], high-resolution melting [11], vector control quantitative analysis [12], and droplet digital PCR [13, 14], have shown efficacy in herb quantification research. However, these methods encounter difficulties in accurately quantifying closely related target species [12-15]. The primary source of adulteration in HPs is the intentional or unintentional mixing of closely related species at low taxonomic levels due to their morphological or olfactory similarities [16-17]. Consequently, there is an urgent need for a user-friendly, precise, and reliable quantification method.

Pyrosequencing is a powerful tool for accurately quantifying individual species, especially those at low taxonomic levels [18, 19]. Pyrosequencing-based species identification (SpeID) assays have been successfully employed for identifying hair tissues from species such as *Equus caballus* and *E. asinus* [20]. Additionally, base ratios at specific single nucleotide polymorphism (SNP) loci obtained through pyrosequencing of a 1∶1 mixture of *Lonicerae japonica* and its adulterant species from the same genus demonstrate this technique's potential for species quantification [21]. Therefore, pyrosequencing-based SNP analysis is particularly suitable for accurately quantifying herb species at low taxonomic levels.

In this study, we introduce a pyrosequencing-based approach to facilitate precise and straightforward quantification of herb species, termed Herb-Q. The primary objective was to develop and validate this method for quantifying *Pinellia ternata* and its low-taxonomic-level adulterants. We comprehensively tested the methodological specifications of Herb-Q and assessed its quantitative capability by measuring bias. Additionally, we evaluated the feasibility of Herb-Q as a user-friendly and efficient approach for quantifying herbs, aiming to provide technical support for market regulation and trade traceability.

Materials and Methods

Test materials

A total of 107 samples from six members of the Araceae family were collected across 15 provinces in China (Table S1). These samples included Pinellia Rhizoma (PR) and tubers of *P. pedatisecta* Schott. (PPR), *P. cordata* N. E. Brown. (PDR), *Typhonium flagelliforme* (Lodd.) Blume. (TFR), *T. blumei* Nicolson and Sivad. (TBR), and *Arisaema heterophyllum* Blume. (AHR). Among these, *P. ternata* has official approval for medicinal use in PR [22], while the other species are commonly used as adulterants of PR in the market. To ensure authenticity, all collected samples were identified using internal transcribed spacer 2 (ITS2) barcoding. The tubers were then washed with water, sliced, and dried at 56 °C for 12 h. After drying, the samples were ground into a fine powder for genomic DNA extraction. To evaluate the validity of the developed assay, mixed samples of PR and a randomly selected adulterant, TBR powder, were prepared. The mixed proportions of these two materials are described in detail in the Detection Methodology section. Specimens and DNA from all samples were deposited in the herbarium of the Institute of Chinese Materia Medica, China Academy of Chinese Medicinal Sciences (Beijing, China).

DNA extraction and Sanger sequencing

DNA was extracted using the Universal Genomic DNA Kit (Tiangen Biotech, Beijing Co., China). The DNA template was amplified with the primers ITS2-S2F/S3R and Maturase K (*mat*K)-390F/1326R using an optimized reaction system and PCR procedure (Table S2) on an Applied Biosystems 2720 Thermal Cycler (AppliedBiosystem, Madrid, Spain). The PCR products were then sequenced using Sanger sequencing at Tsingke Biotechnology Co., Ltd. (Beijing, China). High-quality ITS2 or *mat*K sequences were uploaded to NCBI (https://www.ncbi.nlm.nih.gov/) for BLAST analysis.

SNP selection and pyrosequencing

High-quality PCR products were prepared for pyrosequencing using the PyroMark®Q48 Autoprep instrument (Qiagen, Hilden, Germany). Following the manufacturer's guidelines, primers, enzymes, substrates, sepharose bead mixture, and dNTPs were sequentially added to the reagent chamber, and the pyrosequencing reaction was automatically performed according to the preset protocol [23]. Primers were designed based on specific SNP loci that showed species conservation and species-specific variations within the *mat*K and ITS2 sequences (Table S3 and Figs. S1−S6). These SNP loci were selected to ensure precise differentiation between species (Tables S4a and S4b). All primers were synthesized by MDBio, Inc. (Beijing, China). For pyrosequencing, the input sequences at each SNP locus ranged from 16 to 35 bases. Importantly, when using a forward sequencing primer, the reverse primer must be biotin-labeled at the 5′ end to facilitate the pyrosequencing process.

Developmental validation

Methodological validation was conducted to thoroughly evaluate the efficacy of Herb-Q in molecular identification [12-14]. This validation process included assessments of linearity, limit of detection (LOD), limit of quantitation (LOQ), accuracy, and repeatability [24]. TBR was selected as an adulterant to mix with PR for these evaluations.

To determine the quantification range, TBR powder was mixed with PR in varying proportions: 1%, 2%, 4%, 6%, 8%, 10%, 20%, 30%, 40%, and 50% (*Wt/Wt*), corresponding to PR proportions of 99%, 98%, 96%, 94%, 92%, 90%, 80%, 70%, 60%, and 50% (*Wt/Wt*). Each experiment was performed in triplicate. Linearity was assessed by examining the relationship between the chemiluminescent signal values at the target SNP site and the proportion of each species.

The LOD and LOQ of Herb-Q were evaluated using TBR powders at proportions of 8%, 6%, 4%, 2%, and 1%. The LOD was defined as the ability to detect the target in ≥ 95% of replicates, while the LOQ was defined as achieving a relative standard deviation (RSD) of ≤ 25% [14]. Each proportion was tested in twenty replicates.

Accuracy and repeatability were verified using mixtures with known proportions. TBR powders were tested at propor-

tions of 15%, 25%, and 35%, with corresponding PR proportions of 85%, 75%, and 65%, respectively. Each experiment included six replicates.

Establishing the quantitative molecular assay

Quantitative standard curves for each SNP site were established by mixing specific species with PR at proportions of 10%, 30%, 50%, 70%, and 90%, with three replicates for each proportion. A linear relationship was confirmed when the coefficient of determination (R^2) was 0.99.

PPR was randomly selected as an adulterant for PR, and PDR was used as an external standard material with a known weight. This standard material was chosen for its unique SNP sites suitable for the Herb-Q method and was not present in the mixture to be quantified. To simulate contamination or preliminary processing, small amounts of mineral substances or other species were randomly added to the mixtures. Sample 1 was a mixture weighing 203.2 mg and composed of 72.1 mg PR powder, 20.3 mg PPR powder, 8.5 mg starch, and 102.3 mg PDR powder. Sample 2 was a mixture weighing 204 mg and composed of 71.6 mg of PR powder, 21.5 mg of PPR powder, 9.6 mg of wheat flour, and 101.3 mg of PDR powder. Sample 3 was a mixture weighing 201.3 mg and composed of 71.1 mg PR powder, 19.6 mg PPR powder, 9.6 mg starch and wheat flour, and 101 mg PDR powder. The samples were thoroughly mixed, and DNA was extracted followed by PCR amplification. PCR products were then used for pyrosequencing of SNP sites of the external standard substances and PR. The total weights of PR, PPR, and PDR were calculated using Equation (1), followed by the calculation of the weight of PR in the test sample (*W_ternata*) using Equation (2):

$$W\_all = W\_cordata / R\_cordata \quad (1)$$

$$W\_ternata = W\_all \times R\_ternata \quad (2)$$

Where *W_cordata* represents the actual weight of PDR in the test sample; *R_cordata* is the ratio of the chemiluminescent signal of a unique genotype for PDR at a specific SNP locus; *W_all* is the calculated sum of PR, PPR, and PDR; and *R_ternata* is the ratio of the chemiluminescent signal of a unique genotype for PR at a specific SNP locus.

The Herb-Q approach used to obtain the weight of adulterants was considered applicable when the bias was less than 25% [24]. The bias was calculated as follows:

$$Bias = \frac{|W\_ternata - A\_ternata|}{A\_ternata} \times 100\% \quad (3)$$

Where *A_ternata* represents the actual weight of PR in the test sample, and *W_ternata* is the calculated weight of PR in the test sample.

Results

Sequence validation, SNP loci verification, and primer design

Among the species studied, only the tubers of *P. ternata* have official approval for medicinal applications in PR [22]. However, once powdered and with the outer skin removed, the tubers of *P. ternata* and its five adulterants are visually indistinguishable (Fig. 1A).

Preparation of sequences

The selection of SNP loci for pyrosequencing requires both appropriate specificity and suitable conservation [25]. To identify suitable sequences, we evaluated common DNA barcodes, including ITS2, *mat*K, *rbc*L, and *psb*A-*trn*H. The *mat*K sequence was deemed the most suitable due to its optimal number of SNP loci at low taxonomic levels, covering both the genus and species levels. Although the ITS2 sequence exhibited high variation at the genus level, it maintained a moderate number of SNP loci at the species level. The *rbcL* sequence showed slight variation at the family level, while the *psbA-trnH* sequence was highly variable at the species level. Consequently, *mat*K and ITS2 sequences were selected for quantifying PR and its adulterants using Herb-Q.

ITS2 sequencing results from 107 samples were subjected to BLAST analysis in the NCBI database to confirm the authenticity of the voucher samples. Detailed sequences of these samples are provided in Data S1. As of February 2023, all public 257 *mat*K sequences from all species of the genera *Pinellia*, *Arisaema*, and *Typhonium*, along with 175 ITS2 sequences from all species of the genus *Pinellia*, were downloaded from the NCBI database. Combined with 107 *mat*K and 57 ITS2 sequences obtained from the voucher samples, a total of 364 *mat*K and 232 ITS2 sequences were used for the Herb-Q analysis.

Specific SNP sites and primers for Pinellia ternata

One specific SNP locus, *mat*K-34 (Table S5), in the *mat*K sequence was identified to distinguish the *P. ternata* in the mixture. Genotype G was observed at the *mat*K-34 site for *P. ternata*, whereas genotype A was observed in other species. This locus was considered specific among the 364 sequences analyzed (Table S5). However, sequence specificity verification suggested that species such as *P. peltata* and *P. polyphylla*, belonging to the same genus *Pinellia*, might also occasionally exhibit genotype G at this site. Notably, *P. peltata* and *P. polyphylla* are uncommon in China and have not been reported as adulterants of *P. ternata*. Therefore, *mat*K-34 is a specific SNP locus for *P. ternata* among its closely related species. The allele frequencies determined by pyrosequencing were consistent with expectations (Fig. 2), confirming the utility of this locus for identifying *P. ternata* adulteration.

Specific SNP sites and primers for genus Pinellia

P. pedatisecta, belonging to the same genus as *P. ternata*, is a common adulterant with an adulteration rate as high as 85% [26]. To rapidly identify adulterant species at the genus level in test samples, specific SNP loci for the genus *Pinellia* were screened. Initially, *mat*K-31, *mat*K-352, and *mat*K-404 sites in the *mat*K sequence (Table S6) were considered potential SNP loci for identifying adulterants in *Pinellia* mixtures. However, multiple consecutive bases (≥ four sites) were the same genotype (T) at the *mat*K-31 and *mat*K-404 loci could affect quantification. Consequently, the *mat*K-352 locus was deemed more suitable for further verification. At this locus, genotype T was observed in *Pinellia*,

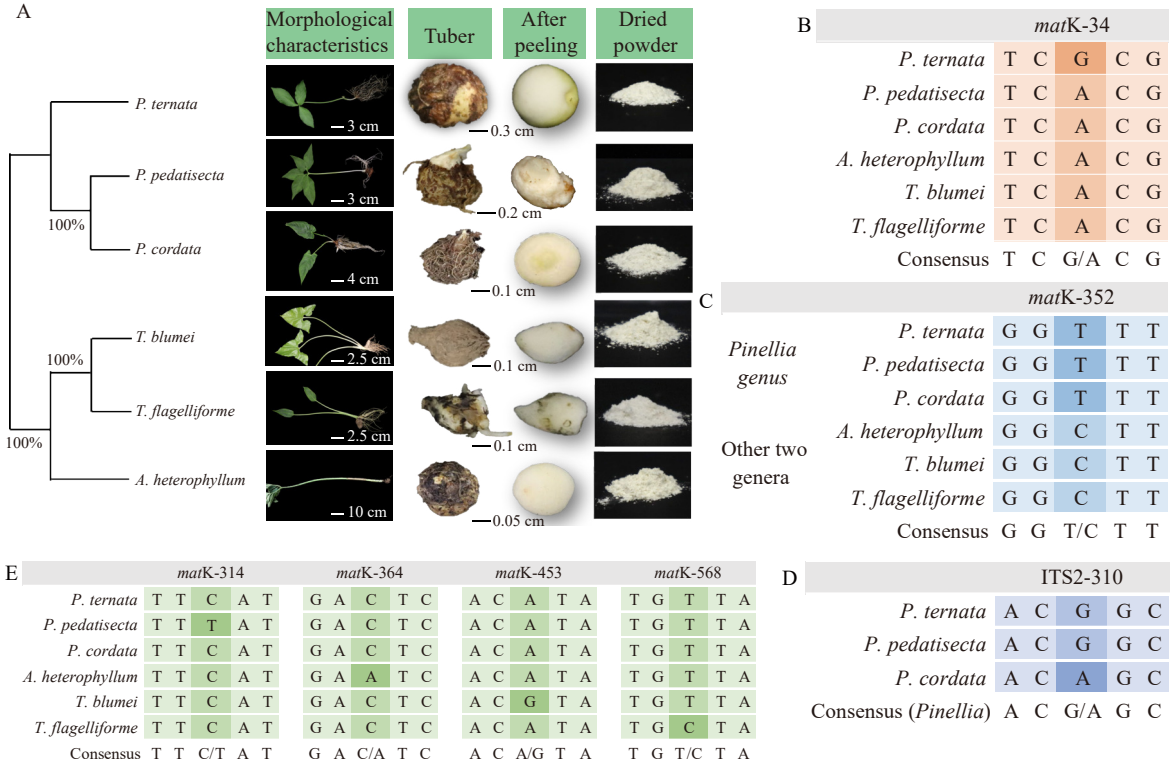


Fig. 1 Characteristics of PR and its adulterants (A) and the SNP sites verified for its quantitative adulteration detection (B−D). Genotype G observed at the *mat*K-34 site in PR and genotype A in other adulterants (B). Genotype T observed at the *mat*K-352 site in *Pinellia* and genotype C in other species (C). Genotype A observed at the ITS2-310 site in PDR and genotype G in PR and PPR (D). The weights of adulterant PDR were verified based on the pyrosequencing results of SNP *mat*K-352 combined with ITS2-310. Genotype T observed at the *mat*K-314 site in PPR and genotype C in PR and other adulterants (E). Genotype A observed at the *mat*K-364 site in AHR and genotype C in PR and other adulterants (E). Genotype G observed at the *mat*K-453 site in TBR and genotype A in PR and other adulterants (E). Genotype C observed at the *mat*K-568 site in TFR and genotype T in PR and other adulterants (E). PR, tuber of *Pinellia ternata*; PPR, tuber of *P. pedatisecta*; PDR, tuber of *P. cordata*; AHR, tuber of *Arisaema heterophyllum*; TBR, tuber of *Typhonium blumei*; TFR, tuber of *T. flagelliforme*.

while other species in the genera *Arisaema* and *Typhonium* exhibited genotype C. The allele frequencies for each species at *mat*K-352 were as expected (Fig. 2).

Specific SNP sites and primers for five individual adulterants

Thirteen sites in the *mat*K sequence were screened as candidate SNP sites for identifying individual adulterants (Table S7). From this analysis, specific SNP loci were identified for the following adulterants: PPR: *mat*K-314 (genotype T), AHR: *mat*K-364 (genotype A), and TBR: *mat*K-453 (genotype G), TFR: *mat*K-568 (genotype C). In contrast, other species exhibited different genotypes at these loci: genotype C at *mat*K-314, genotype C at *mat*K-364, genotype A at *mat*K-453, and genotype T at *mat*K-568. The allele frequencies at these SNP loci for each species were consistent with expectations, successfully verifying their specificity (Figs. S7–S9).

Specific SNP sites suitable for identifying *P. cordata* rhizoma (PDR) were not found in either the *mat*K or ITS2 sequences. Therefore, a combination of two SNP sites was utilized for PDR identification. Initially, the *mat*K-352 locus was used to determine the presence of any adulterant species from the genus *Pinellia*. Subsequently, the ITS2-310 locus was employed for further differentiation within the genus *Pinellia*. At the ITS2-310 locus (Table S8), genotype A was observed in PDR, while genotype G was observed in PR and PPR. The allele frequencies for each species at the ITS2-310 locus in *Pinellia* matched expectations (Fig. S10). This approach allowed for the successful identification of PDR in mixtures by combining the results from the ITS2-310 and *mat*K-352 loci.

A total of seven SNP sites from *mat*K and ITS2 sequences were used for analysis. Initially, the weight of authentic PR in the mixtures was determined based on pyrosequencing results from the *mat*K-34 site (Fig. 1B). Following this, the weights of adulterants from the genus *Pinellia* or other genera in the mixtures were established using the *mat*K-352 site (Fig. 1C). Finally, the weight of PDR adulterants was confirmed by combining the results from SNP loci *mat*K-352 combined with ITS2-310 (Fig. 1D). The weights of adulterants PPR, AHR, TBR, and TFR were verified through pyrosequencing results from their respective specific SNP loci: *mat*K-314, *mat*K-364, *mat*K-453, and *mat*K-568 for TFR (Fig. 1E). This comprehensive approach ensured accurate identification and quantification of the various adulterants in herbal mixtures.

Methodological testing

The pyrosequencing results for samples mixed at spe-

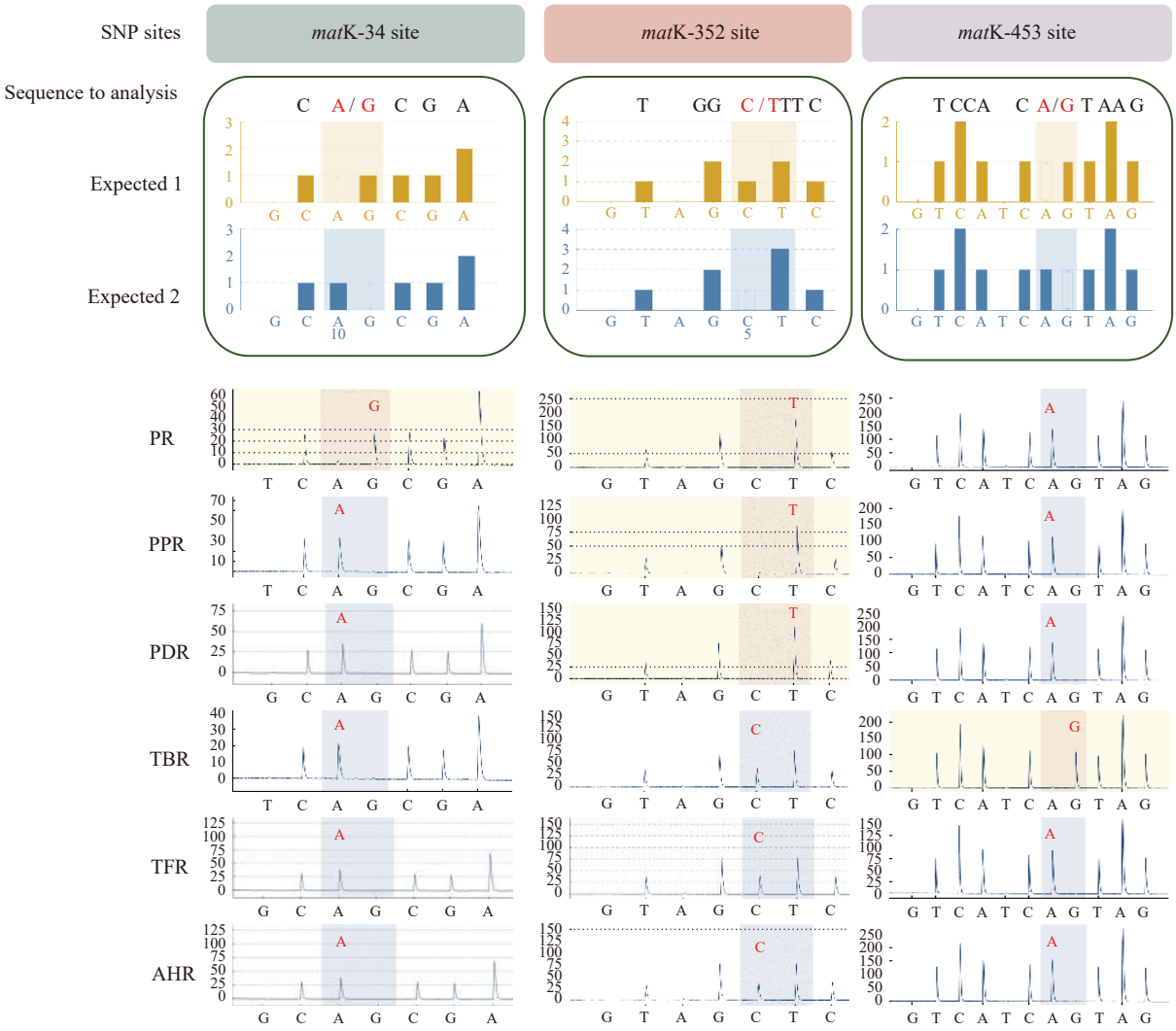


Fig. 2 Verification of the specific SNP loci in PR, genus *Pinellia*, and TBR at the *mat*K-34, *mat*K-352, and *mat*K-453 sites. PR, tuber of *Pinellia ternata*; PPR, tuber of *P. pedatisecta*; PDR, tuber of *P. cordata*; AHR, tuber of *Arisaema heterophyllum*; TBR, tuber of *Typhonium blumei*; TFR, tuber of *T. flagelliforme*.

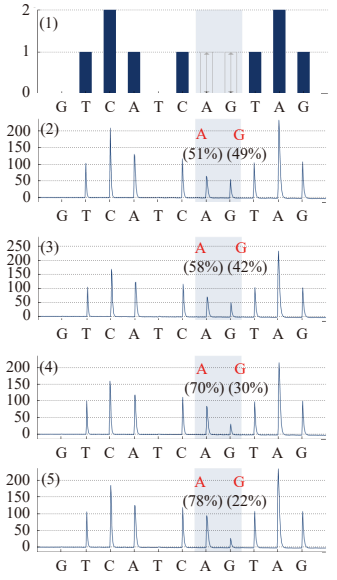
cified ratios are presented in Fig. 3A to evaluate linearity. The chemiluminescent signal values of the genotypes at the *mat*K-453 site were consistently and accurately detected using TBR powder at the specified ratios (except for the 1%). The ratio of chemiluminescence signal values for each base in the specific SNP was consistent with the ratio of the mixed materials. However, usable data could not be obtained for the 1% adulteration ratio. Detailed data from the three replicates for all proportions are presented in Table S9. The coefficient of determination (R^2 = 0.9948) indicated a strong linear relationship between the expected genotype frequency and the measured values, with a fitted line slope of 0.99 (Fig. 3B).

A stable detection rate of ≥ 95% was achieved for detecting 2% adulterants in mixtures (Fig. 3C). Consequently, the LOD for quantifying PR was determined to be 2%. The LOQ was also established at 2%, as the RSD for the 1% proportion of TBR in the PR powder exceeded 25% (Fig. 3C). Data from the 20 replicates for all five proportions are presented in Table S10. For the specified ratios, the RSD values (Fig. 3D) were < 25%, indicating that Herb-Q exhibits good reproducibility for PR quantification.

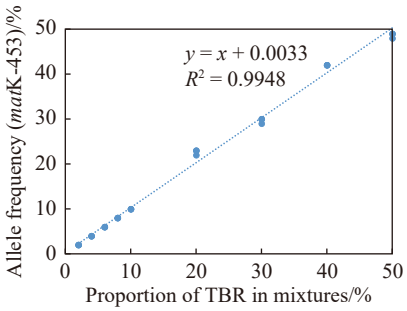
Quantitative standard curve for each SNP locus

Quantitative standard curves were established for each SNP locus by correlating the proportions (*Wt*/*Wt*) of each adulterant with the rates of chemiluminescent signal values for the specific genotypes at each SNP site. For the PR-specific SNP site *mat*K-34, pyrosequencing was conducted on mixtures containing PR powder at proportions of 90%, 70%, 50%, 30%, and 10%, with three replicates for each proportion (Table S11). The average chemiluminescent signal values corresponding to genotype G were observed at proportions of 90%, 70%, 53%, 30%, and 10% (Fig. 4A). The regression equation correlating the proportion of PR (x) to the rate of chemiluminescent signal values of genotype G (y) was determined to be $y = 0.99x + 0.009$ (Fig. 4B), with an R^2 value of 0.9961, indicating excellent linearity. For other specific SNP sites, pyrograms and the regression equations for mixtures of each adulterant with PR in different proportions

A



B



C

Proportion of TBR/G	Max value /%	Min value /%	Mean value /%	RSD /%
8	8	8	7	9
6	5	6	9	6
4	3	4	11	5
2	2	3	13	3
1	0	2	85	4

D

Actual proportion/%	Six parallels/%						Measured proportion/%	RSD/%	Deviation/%
35% TBR/65% PR	33/67	35/65	35/65	34/66	33/67	36/64	34/66	3/2	1/1
25% TBR/75% PR	24/76	25/75	25/75	25/75	24/76	26/74	25/75	2/1	0.6/0.6
15% TBR/85% PR	16/84	17/83	16/84	17/83	17/83	17/83	17/83	3/1	0.5/0.5

Fig. 3 Pyrograms of 10 different proportions at the *mat*K-453 site and the qualitative results for Herb-Q. A1 is the expected pyrogram of Herb-Q at the *mat*K-453 site. A2−A10 are the pyrograms of mixtures at the *mat*K-453 site with TBR powder at proportions of 50%, 40%, 30%, 20%, 10%, 8%, 6%, 4%, and 2% (*Wt/Wt*), respectively. B is the regression equation of TBR mixtures, with R^2 being 0.9948. C shows the LOD and LOQ for TBR adulteration in PR by Herb-Q (n = 20). D is the verification results of quantification for samples with known proportions. PR, tuber of *Pinellia ternata*; TBR, *Typhonium blumei*.

are shown in Figs. S11–S16. Each regression analysis demonstrated R^2 values exceeding 0.99, confirming the high accuracy and reliability of the quantitative measurements. The Herb-Q method effectively detected and quantified the chemiluminescent signal values of the genotypes for PR and its five adulterants.

Calculation of quantification results

The quantification results were calculated based on the actual weights of each constituent substance used to prepare samples 1−3 (Fig. 5A). The expected pyrograms displayed chemiluminescent signals for genotype A at the ITS2-310 site and genotype G at the *mat*K-34 site (Figs. 5B1 and 5C1). Representative pyrograms from one of the three replicates of the pyrosequencing results for these samples are shown in (Figs. 5B2−5B4 and 5C2−5C4). The total weight of all substances, *W_all*, was calculated using the chemiluminescent signal rates of genotype A at the ITS2-310 site (Equation 1). The weight of PR, *W_ternata*, was calculated using the chemiluminescent signal rates of genotype G at the *mat*K-34 site (Equation 2). Fig. 5 shows one pyrogram. Figs. 5B2−5B4 and 5C2−5C4 illustrate one pyrogram from the three replicates for samples 1−3. (Figs. 5D and 5E) of the pyrosequencing results for samples 1−3. The average chemiluminescent signal rates for genotype A at the ITS2-310 site, denoted as *R_cordata*, were 50%, 50%, and 51%, which closely matched the actual values of 53%, 52%, and 53%, respectively. Using Equation 1, the average total weights, *W_all*, for samples 1 to 3 were estimated to be 204.6, 204, and 199.3 mg, respectively. The estimated weights of PR, *W_ternata*, for samples 1 to 3 (Fig. 5F) were 75.0, 71.4, and 69.1 mg, respectively, using Equation 2. These estimates were compared to the actual weights of 72.1, 71.6, and 71.1 mg, respectively. The bias values for samples 1−3, calculated using Equation 3, were all below 25%, with values of 4.05%, 0.3%, and 2.81%, respectively. These results demonstrate that the Herb-Q approach can accurately quantify mixtures of PR and its adulterants, confirming its reliability and precision in differentiating and measuring the components within complex herbal mixtures.

Discussion

Adulteration of HPs poses a significant challenge, with the global adulteration rate being alarmingly high [4]. This issue severely impacts the medical industry, especially in the formulation of traditional and supplementary medicines, where the precise quantification of herbs is crucial [1, 4]. Current methods for herb species quantification have several limitations, including the need for high-specificity species sequences and difficulties in quantifying species at low taxonomic levels [8, 15]. Pyrosequencing technology, an affordable and high-throughput DNA sequencing method, has been widely employed for genotyping and species identification [19, 27-29]. This technique has shown remarkable sensitivity and specificity in quantitatively determining genotypes and assessing mutation loads of disease-causing mutations [30]. Genotyping a single mutation locus represents a novel approach for quantifying herb species [25]. In this study, the

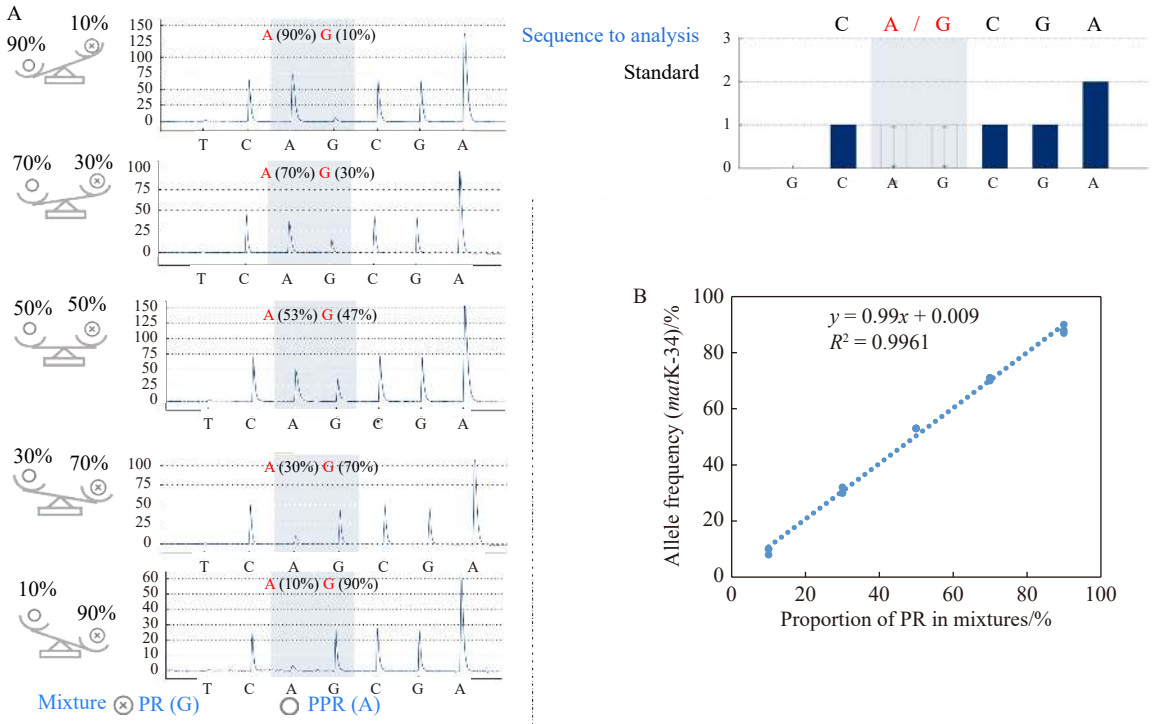


Fig. 4 Pyrograms of different proportions of 90%, 70%, 50%, 30%, and 10% of PR in mixtures at the *mat*K-34 site and the coefficient of determination for linearity. A shows the expected pyrogram of Herb-Q at the *mat*K-34 site, and the pyrograms of mixtures at the *mat*K-34 site using PR powder at proportions of 10%, 30%, 50%, 70%, and 90% (*Wt/Wt*). B shows the regression equations of PR mixtures, with the R^2 0.9961. PR: tuber of *Pinellia ternata*; PPR: tuber of *P. pedatisecta*.

SpeID method was utilized for qualitative species analysis [20], providing a foundation for validating the Herb-Q method. PR was chosen as a case study to establish our new method due to its high adulteration rate (66%) [26]. PR is a toxic species, and its common adulterants belong to the same family (genus) at a low taxonomic level. ITS2, one of the universal DNA barcodes for plants, has been widely used to identify qualitative adulteration of PR [8, 26], but quantitative adulteration detection has not yet been reported. This study introduces Herb-Q, a novel pyrosequencing-based molecular quantification assay that can effectively quantify closely related species. The feasibility of this method was preliminarily verified, demonstrating its ability to accurately quantify target species even in the presence of mineral substances or species used as adjuvants. Herb-Q shows promise as a user-friendly and accurate approach for herb species quantification, with potential applications in quality control.

Considering the distinctive features of Herb-Q, a structured workflow for herb species quantification in medicinal herb adulteration research has been proposed. The process consists of four main steps, as depicted in Fig. 6. The initial step involves the collection of diverse and accurate genetic data for the target species and related species within the same family or genus. Sequences should be obtained from public databases, and additional samples should be collected from various geographical locations to enhance nucleotide diversity for each species. The second step involves the validation of SNP sites for each target species. Specific SNP sites must exhibit high levels of conservation within the species of interest to ensure stability and reliability. These SNP sites should also demonstrate a high degree of specificity when aligned with different species, enabling accurate species identification. PCR and pyrosequencing primers are then designed for each validated SNP locus. These primers are verified through qualitative identification of each species using pyrosequencing techniques, ensuring the robustness and effectiveness of the selected SNP loci for Herb-Q analysis. The third step involves methodological testing to assess various performance parameters, including the LOD, LOQ, accuracy, linearity, and repeatability of the developed method, as recommended by Engl [24]. Quantitative curves are established through multiple mixed proportions and quantification (*W/W*) results using an external standard material. Notably, the external standard species, which serves as an adulterant absent from the mixture, is chosen based on the qualitative analysis conducted earlier in the study.

The methodological testing results indicated that the Herb-Q approach achieved an LOQ of 2% when applied to determine PR and its adulterants. The platform provides accurate visualization results of chemiluminescent signals up to several digits without decimal points [20, 25], contributing to the slight average bias of 2.38% between the true and measured weights of the test samples. This error, encompassing the entire process from DNA extraction to the final quantitative results, is considered acceptable given the complexity of the analysis. Notably, the 2.38% deviation is significantly lower than the reference bias value of 25% and the 7.3% bias reported in other studies using molecular methods [12, 24].

A

	PR/ mg	PPR/ mg	Other impurities/mg	PDR/ mg	Total/ mg
Sample 1	72.1	20.3	8.5	102.3	203.2
Sample 2	71.6	21.5	9.6	101.3	204.0
Sample 3	71.1	19.6	9.6	101.0	201.3

F

	Results of PR	
	Weight/mg	Bias/%
Sample 1	75.0	4.05
Sample 2	71.4	0.30
Sample 3	69.1	2.81

E

	Results of matK-34 (G/A)				
	Actual value	1	2	3	Average value
Sample 1	37%/63%	38%/62%	36%/64%	36%/64%	37%/63%
Sample 2	37%/63%	35%/65%	35%/65%	35%/65%	35%/65%
Sample 3	37%/63%	35%/65%	35%/65%	35%/65%	35%/65%

D

	Results of ITS2-310 (A/G)				
	Actual value	1	2	3	Average value
Sample 1	53%/47%	51%/49%	50%/50%	49%/51%	50%/50%
Sample 2	52%/48%	50%/50%	50%/50%	49%/51%	50%/50%
Sample 3	53%/47%	50%/50%	51%/49%	51%/49%	51%/49%

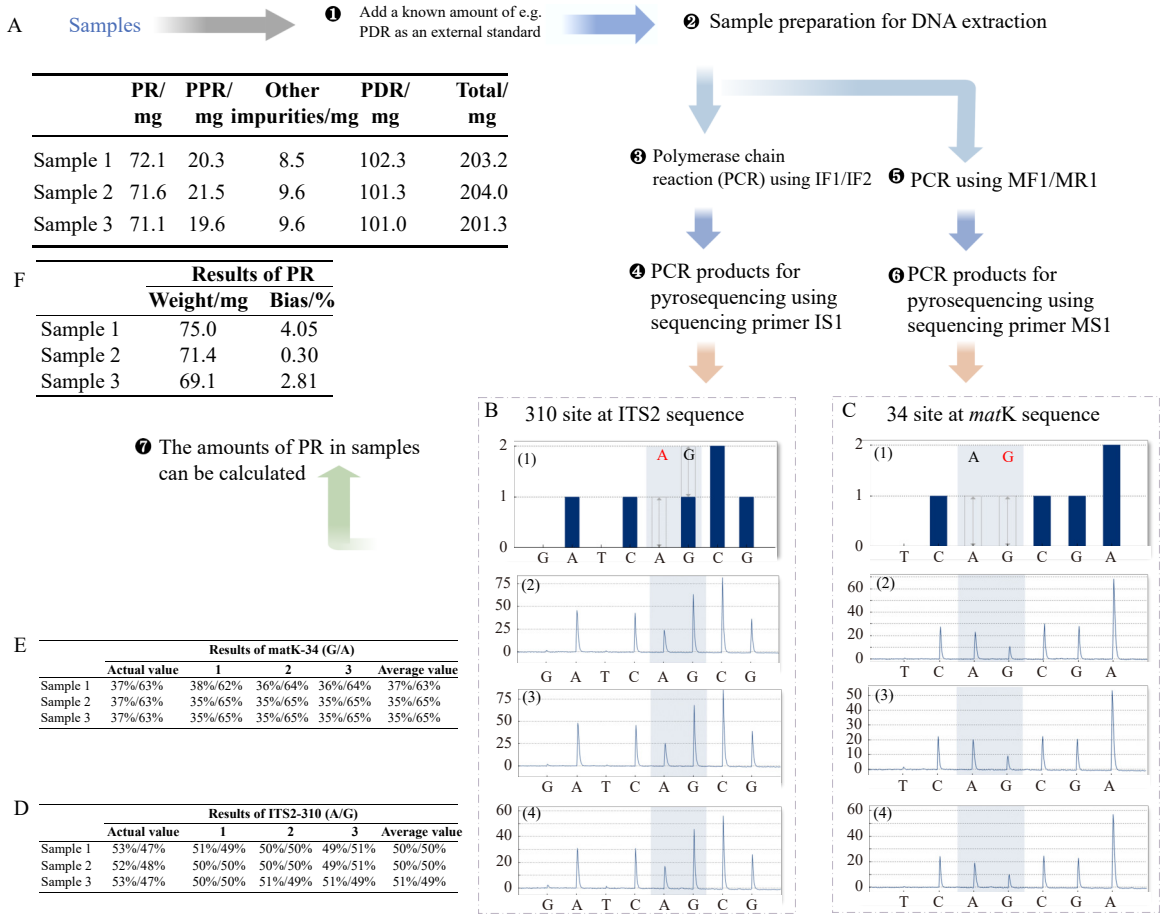


Fig. 5 Flow chart and results of the Herb-Q approach. A shows the true values of components making up samples 1 to 3 based on weight. B1 shows the expected pyrogram at the ITS2-310 site, and the ratio of the chemiluminescent signal of genotype A is represented as the proportion of *W_cordata* in test samples. C1 is the expected pyrogram at the *mat*K-34 site, and the ratio of the chemiluminescent signal of genotype G is presented as the proportion of *W_ternata* in test samples. B2 and C2 are the pyrograms of sample 1 at the ITS2-310 and *mat*K-34 sites, respectively. B3 and C3 are the pyrograms of sample 2 at the ITS2-310 and *mat*K-34 sites, respectively. B4 and C4 are the pyrograms of sample 3 at the ITS2-310 and *mat*K-34 sites, respectively. D represents the actual values and test values of the chemiluminescent signal of genotypes at the ITS2-310 site of samples 1 to 3. E shows the actual and test values of chemiluminescent signals of genotypes of samples 1 to 3 at the *mat*K-34 site. F shows the calculated results of *W_ternata* of samples 1 to 3 and their bias. PR: tuber of *Pinellia ternata*; PPR: tuber of *P. pedatisecta*; PDR: tuber of *P. cordata*.

In the process of quantifying herb species using Herb-Q, detecting SNP loci specific to the target species is the first step. This initial detection helps determine whether the mixture under analysis contains the target species and any potential adulterants. SNP loci that distinguish different genera of adulterants are preferentially selected to narrow the scope of potential adulterants. Once the genus of the adulterants has been determined, appropriate SNP sites are selected for pyrosequencing to quantify the adulterants in the mixture. Thus, selecting appropriate SNP loci in Herb-Q is crucial [25]. Choosing an incorrect SNP site for the target species can result in inaccurate output data and significant deviations in the quantification (*Wt/Wt*) results from their true values. Ideally, the selected SNP site should be free of adjacent identical bases before and after the SNPs and should have suitable regions around the SNPs for designing PCR and sequencing primers [23]. To ensure the selection of the most suitable SNP sites, as many sequences as possible from the genome in a multiple sequence alignment of the target species should be included. However, it is not always possible to obtain SNP sites for some adulterants of a target species due to the lack of whole-genome or plastid genome sequences available for alignment. This challenge can be resolved by obtaining refined genome sequence data from the target species and its adulterants [31]. Currently, quantifying adulterants in complex herbal patent medicines using Herb-Q is challenging due to incomplete DNA extraction from processed herbs or potential reagent-related impacts on nucleic acid extraction efficiency. However, this assay is still applicable for the quantitative detection of target herbs when they are added to herbal patent medicines in simple forms, such as powders, pills, or tablets. With continuous optimization of the DNA extraction process, the application scope of the Herb-Q method is expected to expand gradually.

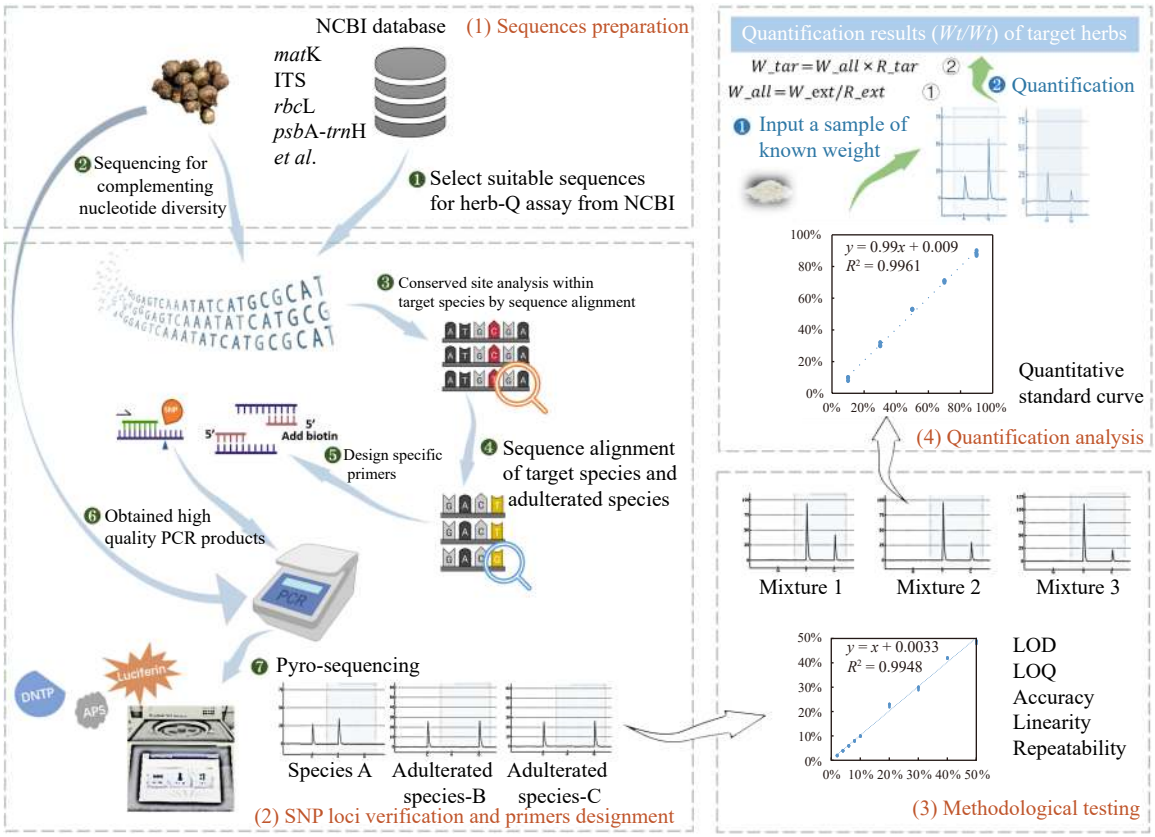


Fig. 6 Process flow of Herb-Q: sequence and supplement selection; SNP loci verification and primer designment; methodological testing, including LOD, LOQ, accuracy, linearity, and repeatability; and establishment of quantitative curves with multiple mixed proportions and quantification (*Wt/Wt*) results obtained using an external standard material. *W_ext* represents the actual weight of the external standard material in the test sample. *R_ext* is the ratio of the chemiluminescent signal of unique genotype for the external standard material at the specific SNP locus. *W_all* is the calculated weight of the test mixture sample and external standard material. *R_tar* is the ratio of the chemiluminescent signal of unique genotype for the target species at the specific SNP locus. The weight of PR in the test sample *W_tar* is the weight of the target species in the test sample. PR: tuber of *Pinellia ternata*. Figure generated with BioRender (https://biorender.com/).

Herb-Q offers several significant advantages for the quantification of herb species over existing methods. The assay boasts reduced operating time, cost, and expertise requirements, as well as enhanced visualization of output data, making it an invaluable tool for researchers and practitioners in the field [20]. The visual output of Herb-Q enhances objectivity in the quantification process, minimizing the influence of subjective parameter selection. This assay uses no more than 35 bases, has a turnaround time of less than 0.5 days from DNA extraction to sequencing, and costs less than $10 per sample. The two key properties of pyrosequencing, generating a peak only when the correct nucleotide is dispensed and incorporated and having peak heights proportional to the number of incorporated nucleotides, form the foundation for the accurate quantification of single herbs using Herb-Q [25]. Compared to other probe-based or enzyme-cut methods [12-14], Herb-Q is more user-friendly and accessible, especially for detecting species at low taxonomic levels. Herb-Q has a wide range of applications, including the quantification of plant herbs, animal-derived herbs, and fungi. It can be used in various contexts to ensure safety and regulatory compliance, including 1) herb products, 2) food products, 3) agricultural products such as spices, and 4) environmental samples such as soil and water to ensure they meet the safety regulatory standards.

Conclusions

Adulteration of herbs significantly impacts economic trade and the medical industry, making accurate quantification essential. This study successfully established the Herb-Q approach for precise quantitative analysis of species, especially at low taxonomic levels. PR, a toxic herb, and its five adulterants were selected as a case study to validate this assay. Herb-Q demonstrated a LOD and LOQ of 2%, with outstanding linearity and accuracy. The Herb-Q method proved effective in detecting adulterated species, even in the presence of mineral substances or other species used as adjuvants, with an average bias of just 2.38%. This approach provides a novel means of identifying adulterants in mixed powders, advancing DNA barcoding from species identification to quantification. Herb-Q offers crucial technical support for market regulation, ensuring clinical safety and traceability.

Supplementary Information

Supporting information of this paper can be requested by sending E-mails to the corresponding authors.

References

[1] Mukherjee PK. *Quality Control and Evaluation of Herbal Drugs* [M]. Elsevier Inc.: Amsterdam, Netherlands, 2019.

[2] Moreira J, Machado M, Dias-Teixeira M, *et al*. The neuroprotective effect of traditional Chinese medicinal plants: a critical review [J]. *Acta Pharm Sin B*, 2023, **13**(8): 3208-3237.

[3] Gurley BJ, Mcgill MR, Koturbash I. Hepatotoxicity due to herbal dietary supplements: past, present and the future [J]. *Food Chem Toxicol*, 2022, **169**: 113445.

[4] Ichim MC. The DNA-based authentication of commercial herbal products reveals their globally widespread adulteration [J]. *Front Pharmacol*, 2019, **10**: 1227.

[5] Srirama R, Santhosh KJU, Seethapathy GS, *et al*. Species adulteration in the herbal trade: causes, consequences and mitigation [J]. *Drug Safety*, 2017, **40**(8): 651-661.

[6] Inagaki K, Mori N, Honda Y, *et al*. A case of drug-induced liver injury with prolonged severe intrahepatic cholestasis induced by Ashwagandha [J]. *Kanzo*, 2017, **58**: 448-454.

[7] Martena MJ, van-der Wielen JCA, van-de Laak LFJ, *et al*. Enforcement of the ban on aristolochic acids in Chinese traditional herbal preparations on the dutch market [J]. *Anal Bioanal Chem*, 2007, **389**(1): 263-275.

[8] Chen S, Yin X, Han J, *et al*. DNA barcoding in herbal medicine: retrospective and prospective [J]. *J Pharm Anal*, 2023, **13**(5): 431-441.

[9] MacKeigan PW, Garner RE, Monchamp ME, *et al*. Comparing microscopy and DNA metabarcoding techniques for identifying cyanobacteria assemblages across hundreds of lakes [J]. *Harmful Algae*, 2022, **113**: 102187.

[10] Villa C, Costa J, Oliveira MBPP, *et al*. Novel quantitative real-time PCR approach to determine safflower (*Carthamus tinctorius*) adulteration in saffron (*Crocus sativus*) [J]. *Food Chem*, 2017, **229**: 680-687.

[11] Sun W, Li J, Xiong C, *et al*. The potential power of Bar-HRM technology in herbal medicine identification [J]. *Front Plant Sci*, 2016, **7**: 367.

[12] Zhao B, Xiong C, Li J, *et al*. Species quantification in complex herbal formulas-vector control quantitative analysis as a new method [J]. *Front Pharmacol*, 2020, **11**: 483193.

[13] Xu W, Zhu P, Xin T, *et al*. Droplet digital PCR for the identification of plant-derived adulterants in highly processed products [J]. *Phytomedicine*, 2022, **105**: 154376.

[14] Yu N, Xing R, Wang P, *et al*. A novel duplex droplet digital PCR assay for simultaneous authentication and quantification of *Panax notoginseng* and its adulterants [J]. *Food Control*, 2022, **132**: 108493.

[15] Lee H, Choi S, Cho I, *et al*. Development and application of a reverse transcription droplet digital PCR assay for detection and quantification of *Plantago asiatica* mosaic virus [J]. *Crop Prot*, 2023, **169**: 106255.

[16] Sheidai M, Tabaripour R, Talebi SM, *et al*. Adulteration in medicinally important plant species of *Ziziphora* in Iran market: DNA barcoding approach [J]. *Ind Crop Prod*, 2019, **130**: 627-633.

[17] Techen N, Parveen I, Pan Z, *et al*. DNA barcoding of medicinal plant material for identification [J]. *Curr Opin Biotech*, 2014, **25**: 103-110.

[18] Förster S, Schumann E, Eberhard WW, *et al*. Discrimination of alleles and copy numbers at the *Q* locus in hexaploid wheat using quantitative pyrosequencing [J]. *Euphytica*, 2012, **186**(1): 207-218.

[19] Karlsson AO, Holmlund G. Identification of mammal species using species-specific DNA pyrosequencing [J]. *Forensic Sci Int*, 2007, **173**(1): 16-20.

[20] Ghemrawi M, Fischinger F, Duncan G, *et al*. Developmental validation of SpeID: a pyrosequencing-based assay for species identification [J]. *Forensic Sci Int-Gen*, 2021, **55**: 102560.

[21] Xu S, Zhao X, Zhang F, *et al*. Study on authenticity identification method of honeysuckle based on pyrosequencing analysis technology [J]. *Curr Biotechno*, 2020, **10**(6): 637-645.

[22] *Pharmacopoeia of People's Republic of China* (Part I) [M]. China Medical Science Press, Beijing, 2020: 123-124.

[23] Ronaghi M. Pyrosequencing for SNP genotyping [J]. *Methods Mol Biol*, 2003, **212**: 189-195.

[24] Engl. Definition of minimum performance requirements for analytical methods of gmo testing [EB/OL]. 2015: http://gmo-crl.jrc.ec.europa.eu/doc/MPR%20Report%20Application%2020_10_2015.pdf

[25] Harrington CT, Lin EI, Olson MT, *et al*. Fundamentals of pyrosequencing [J]. *Arch Pathol Lab Med*, 2013, **137**(9): 1296-1303.

[26] Zhang T, Xu F, Ruhsam M, *et al*. A nucleotide signature for the identification of Pinelliae Rhizoma (Banxia) and its products [J]. *Mol Biol Rep*, 2022, **49**(8): 7753-7763.

[27] Nash PA, Silva-Pinheiro P, Minczuk MA. Genotyping single nucleotide polymorphisms in the mitochondrial genome by pyrosequencing [J]. *J Vis Exp*, 2023, **192**: e64361.

[28] Ghemrawi M, Tejero NF, Duncan G, *et al*. Pyrosequencing: current forensic methodology and future applications: a review [J]. *Electrophoresis*, 2023, **44**(1-2): 298-312.

[29] Melville LA, Redman E, Morrison AA, *et al*. Large scale screening for benzimidazole resistance mutations in *Nematodirus battus*, using both pyrosequence genotyping and deep amplicon sequencing, indicates the early emergence of resistance on UK sheep farms [J]. *Int J Parasitol-Drug*, 2020, **12**: 68-76.

[30] Felsberg J, Wolter M, Seul H, *et al*. Rapid and sensitive assessment of the IDH1 and IDH2 mutation status in cerebral gliomas based on DNA pyrosequencing [J]. *Acta Neuropathol*, 2010, **119**(4): 501-507.

[31] Pei YF, Leng L, Sun W, *et al*. Whole-genome sequencing in medicinal plants: current progress and prospect [J]. *Sci China Life Sci*, 2024, **67**(2): 258-273.

Cite this article as: PEI Yifei, LIU Ziyi, YU Dade, *et al*. Molecular quantification of herbs (Herb-Q): a pyrosequencing-based approach and its application in *Pinellia ternata* [J]. *Chin J Nat Med*, 2024, **22**(7): 663-672.

Professor LI Xiwen, the director of Pharmacology Research Center, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences. His major is pharmcogonosy and research subjects are herbal identification, cultivation and quality evaluation. He has published 70 SCI papers on internationally recognized academic journals including *Sci China Life Sci*, *APSB*, *Ind Crop Prod*, *JHM*, and they have been cited more than 5600 times. Prof. LI joined the Editorial Board of many journals including *CJNM*, *CHIN MED-UK*, *APSB*. He won 1 second prizes of State Scientific and Technological Progress and 3 first prizes in Provincial and Ministerial Level.