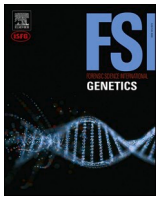
Forensic Science International: Genetics 83 (2026) 103440

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen



Forensic botany clues: Development of a novel *Osmanthus fragrans* STR multiplex system

Kewei Liang [a], Pengfei Li [b], Wei Zhang [b], Chun Shi [b], Bin Zhou [b], Ting Wu [c], Fu Ren [d,e,f,*], Fei Guo [d,e,f,*]

[a] School of Forensic Science and Technology, Criminal Investigation Police University of China, Shenyang, Liaoning 110854, PR China
[b] Suzhou Industrial Park Branch of Suzhou Public Security Bureau, Suzhou, Jiangsu 215127, PR China
[c] Department of Forensic Science and Technology, Hunan Police College, Changsha, Hunan 410138, PR China
[d] Shenyang Medical College, Shenyang, Liaoning 110034, PR China
[e] Liaoning Province Key Laboratory for Phenomics of Human Ethnic Specificity and Critical Illness (LPKL-PHESCI), Shenyang, Liaoning 110034, PR China
[f] Shenyang Key Laboratory for Phenomics, Shenyang, Liaoning 110034, PR China

ARTICLE INFO

Keywords:
Forensic science
Forensic botany
Osmanthus fragrans
Short tandem repeats
Individual identification
Developmental validation

ABSTRACT

Comprehensive research on human DNA genetic markers shows that short tandem repeats (STRs) have found extensive applications in individual identification and kinship testing. Botanical STRs are less commonly employed in forensic science due to the distinctive structure of plant cells and the incomplete state of genome sequencing. This study developed a novel 12-plex system for the individual identification of *Osmanthus fragrans* (*O. fragrans*), a species widely distributed across the world's temperate zones, to furnish essential botanical evidence for forensic investigations at crime scenes. Initially, a method for extracting plant DNA from trace samples was improved, which utilized glass beads to break cell walls and yielded significantly higher DNA quantities compared to the traditional method. Subsequently, five novel highly polymorphic STR loci (GHSTR1, GHSTR2, GHSTR4, GHSTR5, and GHSTR6) with trinucleotide repeats were identified from the *O. fragrans* reference genome (GCA_019395295.1) and characterized (expected heterozygosity > 0.8, polymorphic information content > 0.8, and power of discrimination > 0.9). These loci were integrated with seven previously reported loci (OF5, OF7, OF8, OF9, OF13, OF19, and OSM063) to establish a multiplex system. Developmental validation was conducted in accordance with ISFG and SWGDAM guidelines. The system demonstrated substantial species specificity, high sensitivity (optimal input: 50–500 pg), robust tolerance to common PCR inhibitors (e.g., ≤500 μM hematin and ≤400 ng/μL humic acid), and reliable performance in mixture studies (detecting minor contributors at 1:19–19:1 ratios). PCR conditions were further optimized to 58°C annealing, 20 min elongation, and 28 cycles. The system showed high precision (standard deviation <0.1 bp), accuracy (size deviation within ±0.5 bp), and inter-laboratory concordance. Population analysis of 273 samples revealed the high power of the system effectiveness (combined power of discrimination = $1 - 1.0045 \times 10^{-19}$ and combined probability of exclusion = 0.993111589879955) despite observed Hardy-Weinberg equilibrium deviations. Ultimately, the system established definitive links between botanical evidence from suspects/victims and primary crime scenes in two real-world cases. In conclusion, the findings of the present study demonstrate the applicability of a novel STR multiplex system for individual identification of *O. fragrans* trace samples in forensic science. Without these forensic botany clues, these cases may not have been resolved.

1. Introduction

Crime scenes involving dense vegetation are typically structurally complex and often lack closed-circuit television surveillance, making it challenging for law enforcement to obtain photographic evidence or identify eyewitnesses. Moreover, the chance of recovering crucial forensic evidence, such as DNA left by the suspects in these environments, is extremely low [1–4]. However, suspects or their vehicles may inadvertently transport botanical evidence such as leaves, flower petals, and pollen from crime scenes. Therefore, identifying species or

* Corresponding authors at: Shenyang Medical College, Shenyang, Liaoning 110034, PR China.
E-mail addresses: rf@symc.edu.cn (F. Ren), nevergfaye@gmail.com (F. Guo).

https://doi.org/10.1016/j.fsigen.2026.103440
Received 8 November 2025; Received in revised form 27 January 2026; Accepted 28 January 2026
Available online 29 January 2026

individuals from botanical samples retrieved at a crime scene and comparing them with trace plant DNA derived from a suspect's body or personal belongings can help determine whether or not the suspect was present at the crime scene [5–7]. Additionally, forensic investigators can utilize botanical evidence to establish positive matches between the location of a body's disposal and the primary crime scene [8–10].

Two key issues must be addressed for forensic botany clues to be effectively utilized. First, the outer layer of plant cells is unique, distinguished from the structure of human cells [11]. To extract nuclear DNA from a plant, it is necessary first to break down the cell wall. The common method of soaking the sample in liquid nitrogen and then grinding it by hand into a powder is both strenuous and time-consuming. Additionally, this method suffers from substantial specimen loss and potential cross-contamination, and it performs poorly when extracting trace amounts of plant DNA for forensic applications. Second, classical plant DNA typing methods commonly employ random/anonymous markers, such as random amplified polymorphic DNA (RAPD) [12], inter-simple sequence repeat (ISSR) [13], and amplified fragment length polymorphism (AFLP) [14]. While these methods are relatively simple, their reliability and repeatability of the outcome cannot always be guaranteed [15], and they are not well compatible with the genetic analyzers commonly used in forensic laboratories. Given the broad range of applications in human samples, short tandem repeats (STRs), amplified by fluorescent-labeling multiplex polymerase chain reaction (PCR) and detected by capillary electrophoresis (CE), have the potential to address the aforementioned shortcomings [16–18]. One challenge when applying STRs in forensic botany is that, in most cases, the basic characteristics (e.g., inheritance, genomic location, and polymorphism) of STRs are not always available for the species of interest. In addition, when a kit is generated for forensic purposes and intended for use in real cases, it must be developed and validated according to the guidelines of the International Society of Forensic Genetics (ISFG) [19] and the Scientific Working Group on DNA Analysis (SWGDAM) [20] recommendations for the use of non-human DNA.

This study focuses on *Osmanthus fragrans* (*O. fragrans*), a widely distributed species across southern China. It begins with the improvement of the plant DNA extraction method for trace samples, progresses to the discovery of highly polymorphic STR loci to enhance the power of discrimination, and culminates in the development of a novel multiplex system to identify *O. fragrans* individuals. Additionally, developmental validation is conducted according to the SWGDAM guidelines, including species specificity, sensitivity, stability, precision, accuracy, population, mixture, and PCR-based studies. Lastly, two cases involving the individual identification of *O. fragrans* to determine crime scenes without other investigative clues are provided. Resolving such cases might have been more difficult without these forensic botany clues.

2. Materials and methods

2.1. Sample preparation

O. fragrans is widely distributed in southern China (Figure S1) and also occurs extensively worldwide. An *O. fragrans* specimen was selected from the Suzhou Botanical Garden as a positive control (PTC) and applied to sensitivity, stability, concordance, and PCR-based studies. The PTC species was authenticated by sequencing a two-locus barcode system (*rbcL* + *trnH-psbA*), as recommended by Ferri et al. [21], and by comparison with sequences available on GenBank release 268.0 and Barcode of Life Data System (BOLD) v5.0. A total of 273 *O. fragrans* leaf samples were collected from botanical gardens at various locations, including Suzhou City in Jiangsu Province (N = 105), Hunan Province (N = 60), Zhejiang Province (N = 40), Guangxi Zhuang Autonomous Region (N = 18), Fujian Province (N = 16), Yunnan Province (N = 15), Sichuan Province (N = 9), Guangdong Province (N = 5), and Shanghai City (N = 5), which were used for accuracy, population, stutter, and balance studies. The species of population samples were confirmed by botanical specialists through morphological identification analysis. In-house allelic ladders were utilized for precision studies. Double-distilled water (ddH_2O) was used as the negative control (NTC) from DNA extraction to CE processes.

Species specificity was assessed using 11 common plant samples (*Aspidistra elatior*, *Hedera nepalensis*, *Ilex chinensis*, *Magnolia denudata*, *Morus alba*, *Nandina domestica*, *Photinia serratifolia*, *Phyllostachys viridis*, *Populus tomentosa*, *Salix babylonica*, and *Trachycarpus fortune*) from the Suzhou Botanical Garden. Also, this study included human female control DNA 9947 A (Thermo Fisher Scientific, MA, USA), non-human animals (cattle, sheep, horse, pig, chicken, duck, goose, dog, and rabbit), and purified DNA from microorganisms (*Candida albicans*, *Escherichia coli*, *Staphylococcus aureus*, and *Streptococcus salivarius*).

Sensitivity studies were conducted using DNA extracted from the PTC, subsequently diluted to 1000 pg, 500 pg, 200 pg, 100 pg, 50 pg, 20 pg, and 10 pg. All serial dilutions were tested in triplicate.

Stability studies were performed using two common forensic inhibitors: 10 mM hematin (NuHigh Biotech, Suzhou, China) and 2 mg/mL humic acid (NuHigh Biotech). The final concentrations of the inhibitors varied as follows: 0–1000 μM for hematin and 0–700 ng/μL for humic acid. All inhibitors were prepared in advance and added to 500 pg PTC prior to PCR.

Mixture studies were tested in triplicate using a series of mixtures of two different *O. fragrans* samples at various ratios (19:1, 9:1, 4:1, 1:1, 1:4, 1:9, and 1:19) while holding the total input DNA mixed constant at 1000 pg. To facilitate data interpretation, the two samples were chosen to have relatively few overlapping STR alleles.

PCR-based studies were evaluated in triplicate by amplifying 500 pg PTC at various thermal cycling parameters, including annealing temperatures of 56, 58, 60, 62, and 64 °C, final elongation times of 0, 5, 10, 15, and 20 min, and cycle numbers of 26, 27, 28, 29, and 30 cycles.

2.2. Improved plant DNA extraction method

A leaf sample (0.3 cm × 0.3 cm, approximately 0.02 g) was taken, rinsed sequentially with 75 % ethanol and sterile water, and then shredded into small pieces within a 1.5 mL centrifuge tube. The 30 μL of 0.2 mm glass beads (Zhifengyi Tech, Shenzhen, China), 10 μL of 0.5 mol/L dithiothreitol (DTT; Promega, WI, USA), 100 μL of 10 % sodium dodecyl sulfate (SDS; Promega), and 10 μL of 1 mg/mL Proteinase K (Merck, Darmstadt, Germany) were added, vortexed for 3–5 min to ensure complete rupture of the cell walls, and then digested and lysed at 65 °C for 30 min. Plant DNA was extracted and purified using the GOMag™ Plant DNA Kit (GeneOn BioTech, Changchun, China) with some **modifications**: (1) **300** μL of lysis solution gPL was added, followed by incubation at 65 °C for 30 min (mixing by inverting 2–3 times during this period), and then removed from the incubator. (2) **100** μL of extraction solution gEX was added, vortexed to mix thoroughly, and placed in a freezer at –20°C for 5 min, centrifuged at 12,000 ×*g* for 5 min, and then **500** μL of the supernatant was transferred to a new 1.5 mL centrifuge tube. (3) 30 μL of magnetic beads S0 and **500** μL of anhydrous ethanol were added, vortexed to mix well, and then left at room temperature for **10** min for adsorption. (4) After vortexing twice during adsorption, the centrifuge tube was placed on a magnetic rack for 1 min, and then all the liquid from the tube was aspirated carefully and discarded. (5) **700** μL washing buffer BW1 containing ethanol was added to wash the sample, which was then covered with a centrifuge tube lid, vortexed to mix for 1 min, returned to the magnetic rack for magnetic separation for 1 min, the liquid in the lid aspirated, and then all the remaining liquid aspirated and discarded. (6) **700** μL of 80 % ethanol was added, and the sample was washed according to the method in step 5. (7) The centrifuge tube was placed on a magnetic rack and dried at **56** °C with the lid open for **2–3** min. (8) 50–100 μL of ultrapure water was added, vortexed to mix well, and left to stand at room temperature for 5 min (mixing by vortexing twice during this period). After vortexing, the tube was placed immediately onto the magnetic rack for magnetic

separation for 1 min. The solution was then transferred to a 1.5 mL centrifuge tube and stored at –20 °C for later use.

DNA was quantified using the Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific) on the Qubit™ 4.0 Fluorometer according to the user guide [22].

2.3. Discovery of O. fragrans STR loci

Based on the 23 chromosomal genome sequences of *O. fragrans* archived in NCBI (GenBank ID: GCA_019395295.1), SSR Hunter v1.3 software [23] was used to search for STR sequences with the following parameters: definition (unit_size/min_repeats): 2/15, 3/10, 4/8, 5/7, and 6/7; interruptions (max_difference_between_2_STRs): 300. Then, two DNA fragments with STR sequences were selected from each chromosome (Table S1), with the genetic distance between each fragment being > 10 Mb, indicating no genetic linkage theoretically.

Primer Premier v5.0 software (PREMIER Biosoft, CA, USA; http://www.premierbiosoft.com/primerdesign/index.html) was used to design primers for the STR sequences identified. The design requirements were as follows: (1) primer length of 15–30 bp, with a common range of 18–27 bp, but not exceeding 38 bp; (2) PCR product length of 100–450 bp; (3) annealing temperature of 55–65°C; (4) primer sequence GC content of 40–60 %; (5) primer sequences with low similarity within the template, especially at the 3′ end; (6) attempting to avoid using base A at the 3′ end of the primer; (7) selection of primers with low ΔG values at the 3′ end (absolute value not exceeding 9), and primers with relatively higher ΔG values at the 5′ end and middle; (8) use of the Primer-BLAST function in the NCBI browser (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) to evaluate the specificity of the primers; (10) all forward primers labeled with fluorescent dye FAM.

The designed primers were evaluated in single-plex reactions for fragment-size variability (i.e., locus polymorphisms) and amplification performance across a range of samples. Five primer pairs with high locus polymorphisms and specificity were chosen for this preliminary phase.

2.4. Multiplex design and allelic ladder creation

Five newly identified *O. fragrans* STR loci (GHSTR1, GHSTR2, GHSTR4, GHSTR5, and GHSTR6), along with seven previously reported loci (OF5, OF7, OF8, OF9, OF13, OF19, and OSM063) with high polymorphisms [24,25], were divided into four groups, each containing three STR loci. Primers were redesigned using Primer Premier, ensuring that the three loci within each group yielded fragment sizes ranging from 100 bp to 450 bp. AutoDimer v1.0 software [26] was used to check for potential hairpins, mismatches, and dimers within and between primers. Forward primers in four groups were subsequently labeled with the fluorescent dyes FAM, HEX, TAMRA, and ROX. The initial concentration of each primer was set at 0.4 μM, and the final concentration was adjusted to achieve inter-locus balance.

Allelic ladders were created from a combination of individual templates representing the range of alleles observed in the populations. Amplicons of different homozygous alleles at a single locus were individually cloned into plasmids using the TOPO® TA Cloning® Kit (Thermo Fisher Scientific). Subsequently, the cloned products were diluted, mixed, reanalyzed, and balanced to generate single allelic ladders. Finally, those single allelic ladders were mixed to combine a 'cocktail' in appropriate portions [27]. All ladder alleles were sequenced to confirm the repeat number and motif. Sanger sequencing was performed on the Applied Biosystems® 3500 Genetic Analyzer (Thermo Fisher Scientific) using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific) following the manufacturer's instructions. Raw data were analyzed using the Sequencing Analysis Software v5.3.1 (Thermo Fisher Scientific).

2.5. PCR amplification

The final multiplex system was conducted in a total volume of 10 μL, containing 2 μL of 2 × NuHi® SU PCR Mix (NuHigh Biotech), 2 μL of primer set (GeneRal Biol, Chuzhou, China), 2 μL ddH_2O, and 4 μL of template DNA (500 pg). The thermal cycling program was configured on a VeritiPro™ Thermal Cycler (Thermo Fisher Scientific) as follows: enzyme activation at 96°C for 5 min; amplification for 28 cycles of denaturation at 96°C for 30 s, annealing at 58 °C for 2 min, and extension at 72°C for 30 s; final elongation at 72°C for 20 min; followed by a hold at 4°C.

2.6. Capillary electrophoresis

The Applied Biosystems® 3500 Genetic Analyzer was set with Dye Set G5 to process data from the five dyes: FAM (blue), HEX (green), TAMRA (yellow), ROX (red), and LIZ (orange), using an appropriate matrix. The electrophoresis system consisted of a mixture of 10 μL Hi-Di™ Formamide (Thermo Fisher Scientific) and GeneScan™ 500 LIZ™ Size Standard (Thermo Fisher Scientific) with 1 μL of amplified products or allelic ladders. The bin and panel for the multiplex system were programmed for genotyping, with virtual bins added to increase genotyping efficiency. Raw data were analyzed using the GeneMapper™ *ID-X* Software v1.5 (Thermo Fisher Scientific) with a peak amplitude threshold of 50 relative fluorescence units (RFU).

2.7. Statistical analysis: precision, accuracy, population genetics, stutter percent, and balance

Precision was measured by calculating the mean and standard deviation (SD) of size for all 60 allelic ladder alleles across multiple capillaries from three consecutive runs on the same instrument. Accuracy was evaluated by calculating the size differences between 72 sample alleles and their corresponding allelic ladder alleles.

The reference population dataset was analyzed using STRAF v2.2.2 [28], including the number of alleles (N_{all}), observed heterozygosity (H_{obs}), expected heterozygosity (H_{exp}) or genetic diversity (GD), polymorphic information content (PIC), match probability (MP), power of discrimination (PD), power of exclusion (PE), typical paternity index (TPI), and tests of Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD). In addition, the combined power of discrimination (CPD) and the combined probability of exclusion (CPE) were calculated. A neighbor-joining (NJ) phylogenetic tree was developed and visualized using MEGA6 [29] based on the F_{ST} matrix.

The stutter percentage was measured by dividing the height of the (N – 1) stutter peak by the height of the associated allele peak. These measurements were then used to determine stutter filter recommendations for genotyping analyses of electrophoresis profiles [30].

The fluorescence balance of peak height from STR loci was computed in accordance with the following principles: (1) the balance of heterozygous alleles (intra-locus) was defined as the peak height ratio (PHR) between the lower and the higher peak height at a specific locus; (2) the balance within one group (intra-color) was determined by first averaging the heterozygous peak heights and halving the homozygous peak heights, and subsequently calculating the ratio of the minimum peak height to the maximum among different loci labeled by the same fluorescent dye; and (3) the balance among different groups (inter-color) was evaluated in a manner analogous to the intra-color but across all loci, regardless of dye label [30].

3. Results

3.1. Comparison between the traditional and improved extraction methods

Twelve *O. fragrans* leaves were collected to evaluate different DNA

extraction methods (Figure S2). The traditional method for extracting plant DNA typically involved grinding plant samples with liquid nitrogen, followed by complex lysis procedures with cetyltrimethylammonium bromide (CTAB) [31,32]. The same sample was divided into two sub-samples (one of 0.02 g for our improved method and one of 2.00 g for the traditional method). The concentration of the resulting DNA templates was measured, and Fig. 1 shows that the improved method performed significantly better than the traditional method ($P = 0.0080$, One-way ANOVA).

3.2. Multiplex system and STR characterization

Out of the 46 loci examined across a range of samples, 39 were characterized by either a non-polymorphic or a low polymorphic nature, while two exhibited poor PCR amplification performance. Thus, five loci (GHSTR1, GHSTR2, GHSTR4, GHSTR5, and GHSTR6) identified in this study and seven previously reported loci (OF5, OF7, OF8, OF9, OF13, OF19, and OSM063) were incorporated in the multiplex system. Table 1 lists genomic location and primer-related information (sequence, distance from repeat motif, melting temperature, concentration, fragment size, and dye label) at each locus. This multiplex system included six simple, four compound, and two complex repeat motif STRs having seven dinucleotide and five trinucleotide repeat units. Fig. 2 shows the internal size standard and allelic ladders for the 12-plex system, along with the profile generated from 500 pg PTC. The barcode sequences (*rbcL* + *trnH-psbA*) from the PTC were subjected to a BLASTn search in GenBank, yielding 100 % identity with *Osmanthus fragrans* (7 hits, the highest number as shown in Figure S3). Similarly, the *rbcL* sequence was 100 % identical with *Osmanthus heterophyllus* (8 hits, the highest number) and *Osmanthus fragrans* (3 hits, the second highest number) using the BOLD Identification Engine. The comprehensive analysis supported the identification of PTC species as *Osmanthus fragrans*, suitable for the following validation studies. Sequencing results of 60 allelic ladders are attached in Figure S4. All five novel *O. fragrans* STR loci belonged to the simple trinucleotide repeat motif, with some variation in the repeat and flanking regions. The $[TG]_7/[TG]_4$ variants were observed starting from the 4th nucleotide in the downstream flanking region of GHSTR4, resulting in CE fragment sizes that are related not only to the repeat number in the repeat region but also to length variants in the flanking region. At GHSTR5, the presence of TTA/TTT/ATG variants within the repeat region disrupted the continuity of $[TTG]_n$, but the fragment size remained unchanged. Additional characteristics (inheritance and polymorphism) of 12 *O. fragrans* STRs will be presented in Section 3.7.

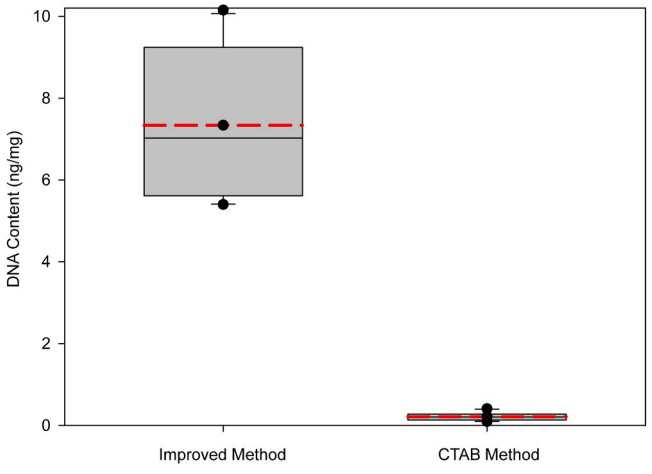


Fig. 1. The comparison between the improved and traditional extraction methods. The improved DNA extraction method developed in this study yields an average of 7.3375 ng of DNA per milligram of leaf from 12 leaf samples. On the other hand, the traditional CTAB method only manages to extract an average of 0.2119 ng of DNA per milligram of leaf from the same 12 leaf samples. The gray box represents the 25th to 75th percentile of the data; the whiskers show the data range; the red horizontal dashed line within the box indicates the mean.

3.3. Species specificity

In silico results showed that each primer pair matched the corresponding STR sequence of *O. fragrans* exactly when aligned with Primer-BLAST at NCBI, but completely mismatched the DNA sequences of other species. In wet-lab tests, the *O. fragrans* sample yielded complete 21 peaks at 12 STRs, whereas other species showed no reproducible peaks above 50 RFU (Figure S5).

3.4. Sensitivity

Results showed that the optimal mass of the DNA template for the 12-plex system ranged from 500 pg to 50 pg with average allele peak heights between 9526 RFU and 409 RFU (Fig. 3A). When the mass of the DNA template increased to 1000 pg, complete genotypes of STRs could be obtained with the average peak height of 13871 RFU and appearance of broad and split peaks caused by excessive peak heights (>30000 RFU at GHSTR1 and GHSTR5) and the occurrence of high (N – 1) stutter ratios (>50 % at the dinucleotide-repeat OF13), which were likely to lead to confusing results. However, to continue decreasing the mass of DNA template, average allele detection rates sharply decreased from 66.67 % to 31.75 % with average peak heights from 149 RFU (20 pg) to 99 RFU (10 pg), which was unsuitable for STR genotyping. Additionally, average PHRs > 70 % were obtained at ≥ 50 pg (Fig. 3B). Intra-locus imbalance (PHR < 70 %) due to stochastic effects was usually observed at ≤ 20 pg and occasionally at 1000 pg (OF8) and 50 pg (OSM063).

3.5. Stability

The capability of the 12-plex system to yield results from DNA subjected to hematin (100, 200, 300, 400, 500, 600, 700, 800, 900, and 1000 μM) and humic acid (100, 200, 300, 400, 500, 600, and 700 ng/μL) was evaluated. Results demonstrated that the 12-plex system had a strong tolerance to substantial concentrations of inhibitors. Figure S6 illustrates that average peak heights and percentage profiles decreased as inhibitor concentrations increased for both inhibitors. Typically, the largest amplicons or loci (e.g., GHSTR2, GHSTR6, and OF8) were the first to drop out. However, OF9 occasionally dropped out at ≥ 600 μM hematin, and GHSTR1 and GHSTR5 at ≥ 900 μM hematin and ≥ 500 ng/μL humic acid. Therefore, complete profiles were obtained with hematin concentrations ≤ 500 μM and humic acid ≤ 400 ng/μL.

3.6. Precision, accuracy, and concordance

Size precision facilitates the determination of reliable and accurate genotypes. Variations may occur between runs on the same instrument due to the type and concentration of the polymer mixture, run temperature, and electrophoresis conditions. Data from 24 allelic ladder samples were collected across three consecutive runs on the 3500 Genetic Analyzer. The mean fragment size and SD value were calculated for 60 allelic ladder alleles. Figure S7 plots the average fragment size against the corresponding SD. The highest SD was 0.093 bp from allele 11 at GHSTR2, and three other higher SD values (> 0.07 bp) were observed for allele 7 at GHSTR2 and alleles 36 and 38 at OF8. However, the 3 × SDs of all allelic ladder alleles fell below 0.3 bp, thereby ensuring that the sample alleles were rarely sized outside the ± 0.5-bp window. Meanwhile, the bin and panel files for the multiplex system were programmed based on the above data.

Size accuracy of the 12-plex system on the 3500 Genetic Analyzer was assessed by calculating the size differences between sample alleles and allelic ladder alleles. Figure S8 illustrates data from a total of 563

Table 1
Characterization and primer-related information on 12 *O. fragrans* STR loci presented in the multiplex PCR system.

Locus	Physical position (GCA_019395295.1)	Reference allele (forward nomenclature)	Primer sequence (5'–3')	From repeat (bp)	Average T_m (℃)	Primer concentration (μM)	Fragment size (bp) and allele range	Dye label
GHSTR1	Chr1:53642–53713	[ATT]24	F: GTTTCAAGATCCGGAATAAAG R: AACTTTATAAAACTAGAAGCCTTGG	46 54	55	2.2	167–224 (8–27)	FAM
OF7*	Chr17:10519468–10519507	[TC]7 [AC]13	F: TGCTCCAACCGAACACTA R: ATTGCAAGGGATCTTATCCTATA	74 86	54	1.0	225–277 (13–39)	FAM
OF13*	Chr12:4744280–4744348	[TG]15 AG T [GA]18	F: CCGCATTAACACCCAAACTT R: CAAAAGAAGACCAATCCCGA	219 63	58	2.2	345–397 (13–39)	FAM
GHSTR5	Chr8:327104–327139	[TTG]12	F: GGGGATTTTGTTGTTTCGTA R: GATTACAATTATGGAAGGAATTTATAGA	2 80	56	1.8	131–191 (3–23)	HEX
GHSTR4	Chr7:507682–507741	[TAT]20	F: AGACACTGACCTCCAACATACGAA R: CGTGTCCGAAAGTGACTGATT	34 134	59	1.0	208–268 (3–23)	HEX
GHSTR2	Chr1:201826–201864	[TAT]13	F: ATGCTTCTTATCGTGCTTTCATTTA R: TGGTTCAATGTGGCTTCCTCT	6 312	59	1.0	370–415 (3–18)	HEX
OF9*	Chr11:4558741–4558764	TA [TC]5 [AC]6	F: GGACCAATTAAATCAAGCAAGT R: GGCTTAGAATCTACATGCCTTCG	54 11	58	4.0	121–165 (5–27)	TAMRA
OF5*	Chr1:27623657–27623680	[TG]10 [AG]2	F: AAATCCCAACTCACAAGGCA R: TCTGCTGCACTCTTCATTGC	53 95	57	1.0	181–233 (3–29)	TAMRA
GHSTR6	Chr15:116965–117027	[ATT]21	F: TGACACTATTTTCACCATTGCG R: ATTTATGAATTACCTCGAAAGATAGA	187 93	58	7.0	343–397 (5–23)	TAMRA
OSM063	Chr15:19695557–19695566	[GA]5	F: ATGTACACTAGCTATTCTACCAACAA R: TCTCATTTTCACTCAACTCACGC	62 3	58	0.4	124–164 (5–25)	ROX
OF19	Chr4:30396574–30396623	[AT]3 [GT]12 [GA]2 [GT]4 [GA]2 [GT]2	F: CTTGCAGACCAAGAATATATTAAAA R: TGCGATAGGTTACAGGCAGAATAG	67 140	58	5.5	177–245 (13–47)	ROX
OF8	Chr3:32229454–32229525	[GT]5 [GA]5 TA [GA]3 [GT]10 TT [GT]2 TT [GT]8	F: CAGCAGTATTATTCGACGGAG R: CTGGGCCCTTTTCTGTTCTT	293 29	57	2.9	390–474 (15–57)	ROX

* Primers are designed based on the reverse complementary GenBank reference sequence.

alleles from 72 samples. All sample alleles were sized within ±0.5 bp of their corresponding allelic ladder alleles, with 64.12 % falling within ±0.25 bp. This demonstrated that the system can reliably determine genotypes and accurately identify the microvariant (partial-repeat) alleles that differ by a single nucleotide from complete-repeat alleles.

Another three laboratories in Kunshan, Suzhou, and Zhangjiagang participated in the validation (blind testing from DNA extraction to STR genotyping in triplicate) and were supplied with PTC samples. Results showed that STR genotypes from PTC samples with the 12-plex system achieved full concordance with those obtained in our laboratories. The average peak heights were (7871 ± 518) RFU, (8555 ± 1557) RFU, and (8434 ± 1223) RFU, respectively, which were similar to the values in sensitivity and PCR-based studies using 500 pg PTC.

3.7. Population genetics

3.7.1. Genetic polymorphisms and forensic parameters

A total of 273 samples from nine locations in southern China were investigated, and their genotypes are listed in Table S2. Due to differences in sample sizes across locations, these samples might not accurately represent the local target populations, particularly in provinces with limited sampling scopes and sizes (< 30 samples). Therefore, they were pooled for calculation purposes. Allele frequencies and forensic parameters for each locus are shown in Table S3 and Table 2: 281 unique alleles were detected across 12 loci; OF8 had the maximum number of variants with 40 alleles, and GHSTR2 had the minimum with 14 alleles; all 12 loci showed high observed heterozygosity ($H_{obs} > 0.5$ except for OF5), expected heterozygosity ($H_{exp} > 0.8$), and polymorphism information content (PIC > 0.8). Statistically significant departures from HWE tests were observed at all 12 loci, even after applying Bonferroni's correction for multiple testing ($\alpha = 0.0042$), and different sampling locations varied from 0 (Sichuan) to 12 loci (Jiangsu and Hunan), as shown in Table S4. An HWE population is essentially a large, randomly mating population in which the frequencies of various genes and genotypes will remain unchanged from generation to generation in the absence of migration, selection, and mutation. The main reasons for the deviations from HWE at many loci in this study may be the result of the asexual reproduction in *O. fragrans*, human intervention, and sampling bias. This is commonly observed in studies of *Cannabis sativa* [33–35]. Due to deviations from HWE, a correction factor (referred to as θ or F_{ST}) should be determined (Section 3.7.2) and applied (Section 3.12.2) when interpreting data for forensic purposes, as recommended by the National Research Council (NRC) II [36]. LD tests revealed that 21 out of 66 pairwise comparisons were statistically significant after Bonferroni's correction ($\alpha = 0.0008$), and the same was true for the variety of locations (0–33 pairwise comparisons in Table S5). However, the loci were specifically chosen from separate chromosomes, with a genetic distance greater than 10 Mb between each fragment. Theoretically, there is no genetic linkage present between them. Thus, the product rule remains applicable for CPD and CPE calculations. The PD ranged from 0.9549 (GHSTR2) to 0.9836 (OF19), with $1 - 1.0045 \times 10^{-19}$ for CPD. The PE ranged from 0.1588 (OF5) to 0.5896 (OF9), with 0.993111589879955 for CPE.

3.7.2. Inbreeding and genetic differentiation

The F-statistic coefficient is an essential indicator of the degree of genetic differentiation and inbreeding in a population. It consists of three parameters: the intra-population inbreeding coefficient F_{IS}, the total-population inbreeding coefficient F_{IT}, and the inter-population differentiation coefficient F_{ST} (or θ). Table 2 shows that all 12 loci had positive values of F_{IS} and F_{IT}, where OF5 had the highest F_{IS} and F_{IT} values, indicating that this locus was more inbred than the other loci. The average F_{IS} and F_{IT} are 0.2708 and 0.3205, respectively, indicating a high degree of inbreeding within subpopulations. It may be another reason for the deviations from HWE. When $0.05 \leq F_{ST} < 0.15$, there was a moderate degree of genetic differentiation among subpopulations. The average F_{ST} across loci was 0.0656, indicating that 6.56 % of the total genetic variation was between subpopulations and 93.44 % within subpopulations. This suggests that there is a limited gene flow between subpopulations, which are not completely isolated but also do not mate

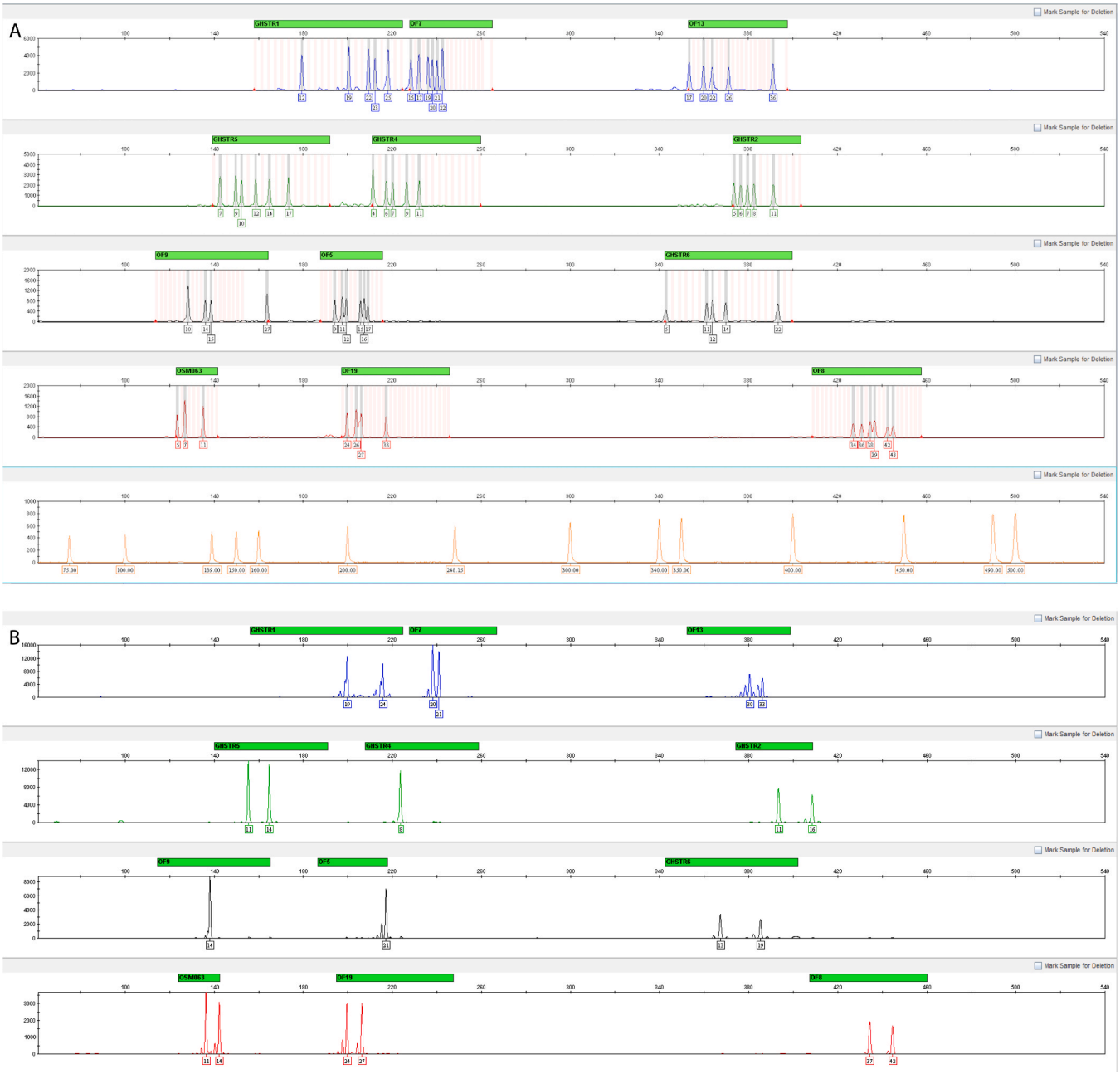


Fig. 2. Multiplex design. (A) An electropherogram of allelic ladders and an internal size standard are designed for the 12-plex system. The four dye panels for allelic ladders correspond to FAM (blue), HEX (green), TAMRA (yellow), and ROX (red) dye-labeled peaks (from top to bottom). The genotype is shown with the allele number displayed underneath each peak. The fifth panel is reserved for the internal size standard labeled the GeneScan™ 500 LIZ™ (orange). One microliter each of allelic ladders and internal size standard is simultaneously analyzed on an Applied Biosystems® 3500 Genetic Analyzer. (B) An electropherogram from 500 pg positive control (PTC) is generated using the 12-plex system.

completely at random. Geographic distance or certain obstacles may hinder individuals' migration and communication to some extent. Figure S9 indicates the clusters in a multidimensional scaling (MDS) plot and a neighbor-joining (NJ) tree based on the F_{ST} matrix by 12 loci between nine sampling locations.

3.8. Mixture

Table S6 presents the genotypes and non-overlapping alleles of the two samples used in mixture studies. As the mixture ratios increased, the percentage of unique alleles identified from the minor contributor decreased (Table S7). In triplicate, 100 % unique alleles from minor contributors (ranging from 100 to 500 pg) were identified between 9:1 and 1:9 mixture ratios; 94.74 % alleles from the minor contributor (50 pg PTC2) at a 19:1 ratio, with the largest amplicon (allele 43) dropouts observed at OF8; and 100 % alleles from the minor contributor (50 pg PTC1) at a 1:19 ratio. The lower detection limit for minor contributors generally corresponded to the sensitivity in Section 3.4.

3.9. PCR-based studies

3.9.1. Annealing temperature

Annealing temperature was tested with 500 pg PTC at 56 °C, 58 °C (standard temperature), 60 °C, 62 °C, and 64 °C. Complete profiles (18 heterozygous alleles and 3 homozygous alleles) were obtained in triplicate at 56°C and 58 °C. Results indicated that fluorescence signal

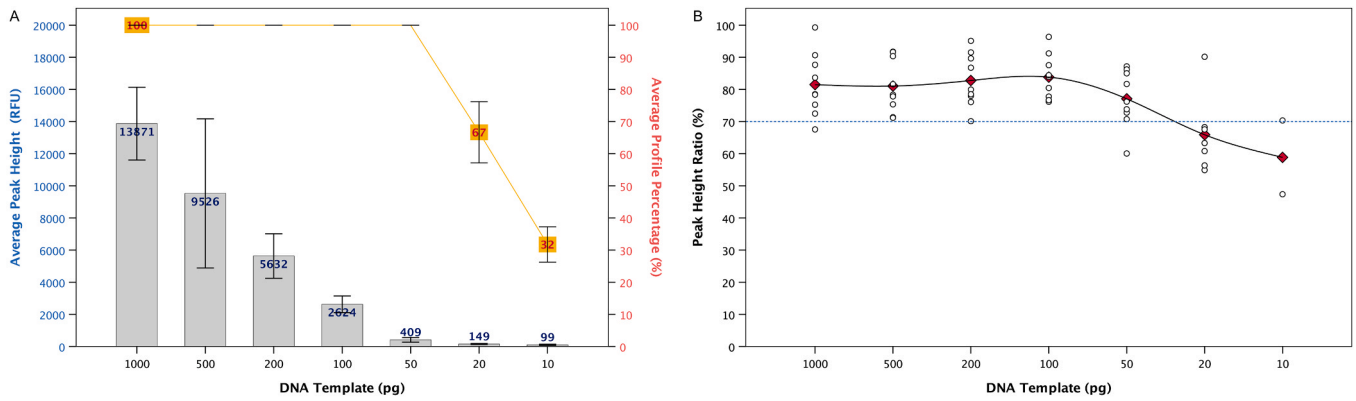


Fig. 3. Sensitivity studies. Results for serial dilutions (1000 pg, 500 pg, 200 pg, 100 pg, 50 pg, 20 pg, and 10 pg) of the positive control (PTC) were amplified with the 12-plex system. (A) The average peak height (left-hand *Y*-axis) is plotted against DNA template mass (bar graph). The average profile percentage (right-hand *Y*-axis) is also against DNA template mass (line graph). Error bars represent the plus and minus standard deviations in triplicate. (B) The peak height ratio (PHR) between lower/higher heterozygous peaks is plotted against DNA template mass. The red diamond presents the average PHR of templates per mass. The blue line serves as a reference at 70 %.

Table 2
Genetic and forensic parameters on 12 *Osmanthus fragrans* STR loci from nine locations in southern China (N = 273).

Locus	N_{all}	H_{obs}	H_{exp}	PIC	MP	PD	PE	TPI	F_{IS}	F_{IT}	F_{ST}
GHSTR1	20	0.6117	0.9164	0.9085	0.0302	0.9698	0.3052	1.2877	0.3197	0.3415	0.0320
OF7	22	0.6960	0.9320	0.9259	0.0205	0.9795	0.4221	1.6446	0.3463	0.3711	0.0379
OF13	27	0.6520	0.9459	0.9413	0.0169	0.9831	0.3580	1.4368	0.2141	0.2497	0.0453
GHSTR5	21	0.5934	0.9325	0.9265	0.0207	0.9793	0.2830	1.2297	0.3261	0.3555	0.0436
GHSTR4	20	0.6190	0.9189	0.9116	0.0271	0.9729	0.3144	1.3125	0.2722	0.3286	0.0775
GHSTR2	14	0.6703	0.8775	0.8642	0.0451	0.9549	0.3839	1.5167	0.2167	0.2671	0.0643
OF9	18	0.7949	0.8862	0.8742	0.0395	0.9605	0.5896	2.4375	0.0367	0.1841	0.1530
OF5	26	0.4652	0.9466	0.9422	0.0247	0.9753	0.1588	0.9349	0.4062	0.4355	0.0494
GHSTR6	19	0.5421	0.9237	0.9168	0.0282	0.9718	0.2271	1.0920	0.3229	0.3478	0.0367
OSM063	21	0.5897	0.9113	0.9032	0.0421	0.9579	0.2788	1.2188	0.3423	0.4309	0.1347
OF19	33	0.6593	0.9421	0.9373	0.0164	0.9836	0.3682	1.4677	0.2345	0.2631	0.0373
OF8	40	0.5861	0.9538	0.9500	0.0200	0.9800	0.2745	1.2080	0.2120	0.2711	0.0749

N_{all}: the number of alleles; H_{obs}: observed heterozygosity; H_{exp}: expected heterozygosity; PIC: polymorphism information content; MP: match probability; PD: power of discrimination; PE: power of exclusion; TPI: typical paternity index. F_{IS}: the intra-population inbreeding coefficient; F_{IT}: the total population inbreeding coefficient; F_{ST}: the inter-population differentiation coefficient.

intensity was initially increased and subsequently decreased as the annealing temperature increased (Figure S10). At 56 °C, the average peak height was 55 % (4982/9042) of the value observed at 58 °C. At 60 °C, the average peak height was 48 % (4366/9042) of that at 58 °C, accompanied by OF9 dropouts in triplicate. Average peak heights and average allele detection rates revealed a significant decline at 62°C (3689 RFU and 86 % profiles with GHSTR5 and OF9 dropouts) and 64 °C (1907 RFU and 48 % profiles with six loci dropouts). Intra-locus balance (PHR > 70 %) was only noted at all loci at 58 °C.

3.9.2. Final elongation time

Final elongation at a certain temperature ensures complete terminal nucleotide addition, producing uniform, well-shaped allele peaks. Amplified products would exhibit (n – 1) peaks (minus A) or shoulders in the absence of this step. Final elongation times of 0, 5, 10, 15, and 20 min (standard elongation time) were examined at the standard temperature (72 °C). Average peak heights (9209 RFU ± 680 RFU) and average PHRs (78 % ± 1 %) were relatively unaffected by different periods tested. Results showed that most loci presented normally shaped peaks under all conditions, but GHSTR4 and OF19 had small shoulder peaks under the 0-min elongation due to incomplete terminal nucleotide addition (Figure S11).

3.9.3. Cycle number

Cycle number provides a flexible approach to obtain the optimal performance in a PCR system. Amplification of 500 pg PTC was assessed at 26, 27, 28 (standard cycle number), 29, and 30 cycles in triplicate. Figure S12 illustrates that increased cycle number may enhance sensitivity but decrease specificity. Average percentage profiles reached 100 % at 26–30 cycles with average peak heights of (2490 ± 276) RFU, (4173 ± 302) RFU, (9484 ± 456) RFU, (13595 ± 677) RFU, and (15230 ± 677) RFU. Nevertheless, spectral bleed-through was observed in the profiles at 29–30 cycles, attributed to over-enhanced signal intensities at GHSTH1, GHSTR4, GHSTR5, and/or OF7, as well as nonspecific amplification products at OF5, OF9, and/or OF19. Furthermore, average PHRs > 80 % were observed at 26–28 cycles and > 70 % at 29–30 cycles, while inter-color balance < 10 % was observed at 26–27 cycles.

3.10. Stutter

Stutter peaks are a natural by-product of PCR amplification in STR profiles and result from strand slippage [37]. Typically, a minor peak with one repeat unit shorter than the corresponding allele is observed. The mean (N – 1) stutter ratio and its SD for each locus were calculated from 72 population samples (Table S8). It is generally observed that stutter is more prominent in shorter repeat units, so that seven previously reported dinucleotide-repeat loci generated higher stutter ratios (19.71 % averaged) than five newly found trinucleotide-repeat loci (8.40 % averaged). The average value of trinucleotide-repeat loci in *O. fragrans* is comparable to that found in humans (e.g., 9.06 % at D22S1045 [30]). The stutter filter is determined by the mean stutter

ratio plus three standard deviations; however, these values do not serve as absolute thresholds. Occasionally, some single-source samples may demonstrate stutter peaks exceeding these filters (e.g., the maximum recorded at 16.51 % at GHSTR4 in Table S8).

3.11. Balance

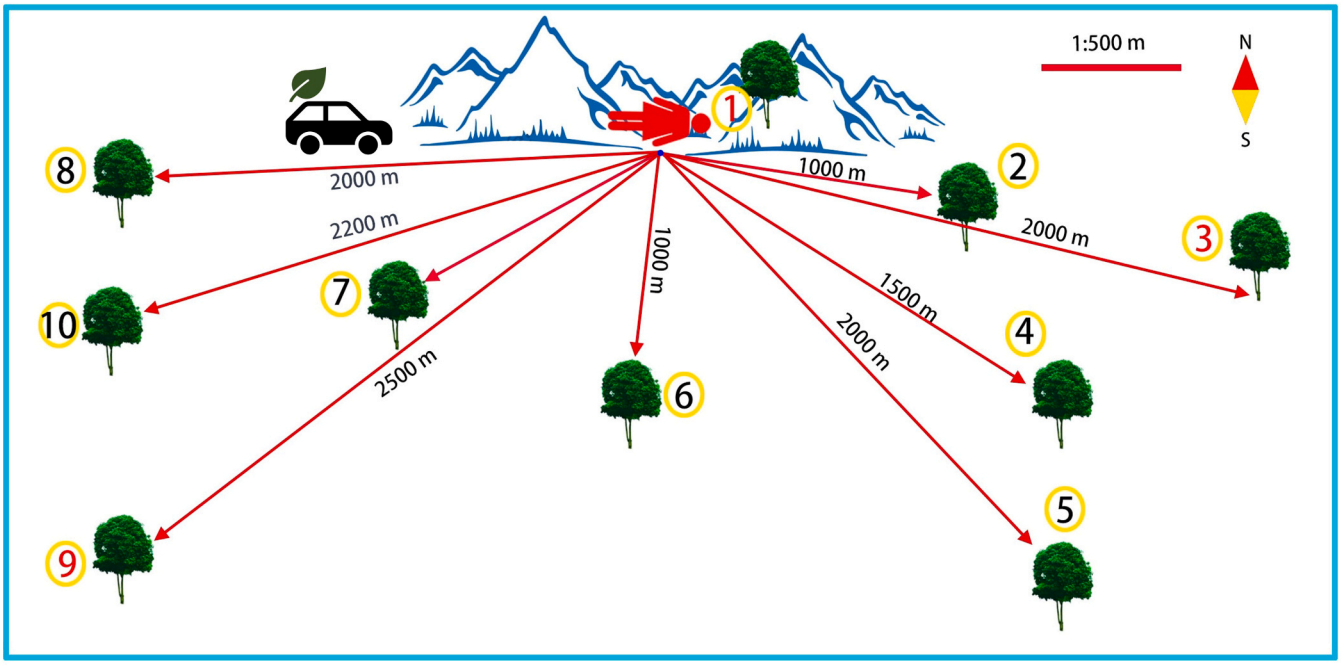
The PCR performance criterion includes intra-locus, intra-color, and inter-color balance. Intra-locus balance of greater than 70 % is desired to ensure accurate heterozygote genotyping across a range of template amounts and to facilitate mixture interpretation [38]. Additionally, intra-color balance of greater than 50 % is adopted to ensure profile completeness from inhibited or degraded samples [38]. Inter-color balance is generally in the 20 %-40 % range [38], with the recommendation > 30 % [39]. Table S9 enumerates the balance calculations derived from 72 population samples. Mean values potentially satisfied the established criteria, with intra-locus balance ranging from 70.18 % (GHSTR5) to 84.07 % (GHSTR2), and intra-FAM (blue), -HEX (green), and -ROX (red) balances ranging from 51.21 % to 61.86 %. Conversely, the intra-TAMRA (yellow) balance was only 44.44 %, attributable to lower peak heights observed at GHSTR6 in the dye panel, despite it having the highest primer concentration. Similarly, the inter-color balance (10.04 %) fell below the criteria, owing to the extreme disparity in peak heights between the red and blue dye panels.

3.12. Case studies

3.12.1. Case history

Case 1: A woman's body lay by the roadside one night in November 2023. Because the scene was isolated and there were no camera surveillance devices at the site, it was impossible to identify any suspects. After screening the footage from more distant electronic cameras, 35 vehicles were determined to have potentially passed the location of the body. An investigation of these vehicles found a half leaf inside one of them. However, the vehicle owner denied having been at the scene. The investigators collected this leaf, as well as those from the *O. fragrans* tree near the body.

Case 2: In August 2023, a male corpse was found in an overgrown vegetation spot next to a lake. There were no camera surveillance devices within a 3-kilometer radius of the body. The body and hands of the



Sample	OF5		OF7		OF8		OF9		OF13		OF19		OSM063		GHSTR1		GHSTR2		GHSTR4		GHSTR5		GHSTR6	
Leaf in the car backseat	21	21	15	15	39	50	13	14	32	39	24	31	5	9	22	22	4	8	5	10	9	14	18	18
Tree 1 (Next to the body)	21	21	15	15	39	50	13	14	32	39	24	31	5	9	22	22	4	8	5	10	9	14	18	18
Tree 2	18	18	19	20	36	44	14	15	22	29	32	34	7	11										
Tree 3	21	21	15	15	39	50	13	14	32	39	24	31	5	9	18	18	7	7	6	18	18	18	11	18
Tree 4	12	12	28	28	29	30	26	26	29	29	24	24	7	7										
Tree 5	12	12	15	22	39	45	14	15	25	33	23	23	7	7										
Tree 6	20	22	21	25	37	39	14	16	30	32	44	44	9	14										
Tree 7	12	12	22	30	31	42	12	14	28	39	26	31	7	14										
Tree 8	21	21	20	21	36	41	14	14	31	34	24	28	11	14										
Tree 9	21	21	15	15	39	50	13	14	32	39	24	31	5	9	14	15	8	8	8	24	13	21	14	18
Tree 10	8	13	21	33	39	42	13	15	26	31	26	26	7	8										

Fig. 4. Schematic diagram of the crime scene and genotypes in Case 1. Half a leaf inside the vehicle and leaves from the *O. fragrans* tree (Tree 1) near the body are collected. Leaves of nine more *O. fragrans* trees (Trees 2–10), located in different areas around the body, are also gathered. Genotypes of the evidence sample are marked in grey. Alleles of trees are concordant with those of the evidence sample marked in green and discordant in red. Alleles from three trees (1, 3, and 9) are matched using seven previously reported loci (OF5, OF7, OF8, OF9, OF13, OF19, and OSM063). However, they can be distinguished using the 12-plex system. Electropherograms are presented in Fig. S13.

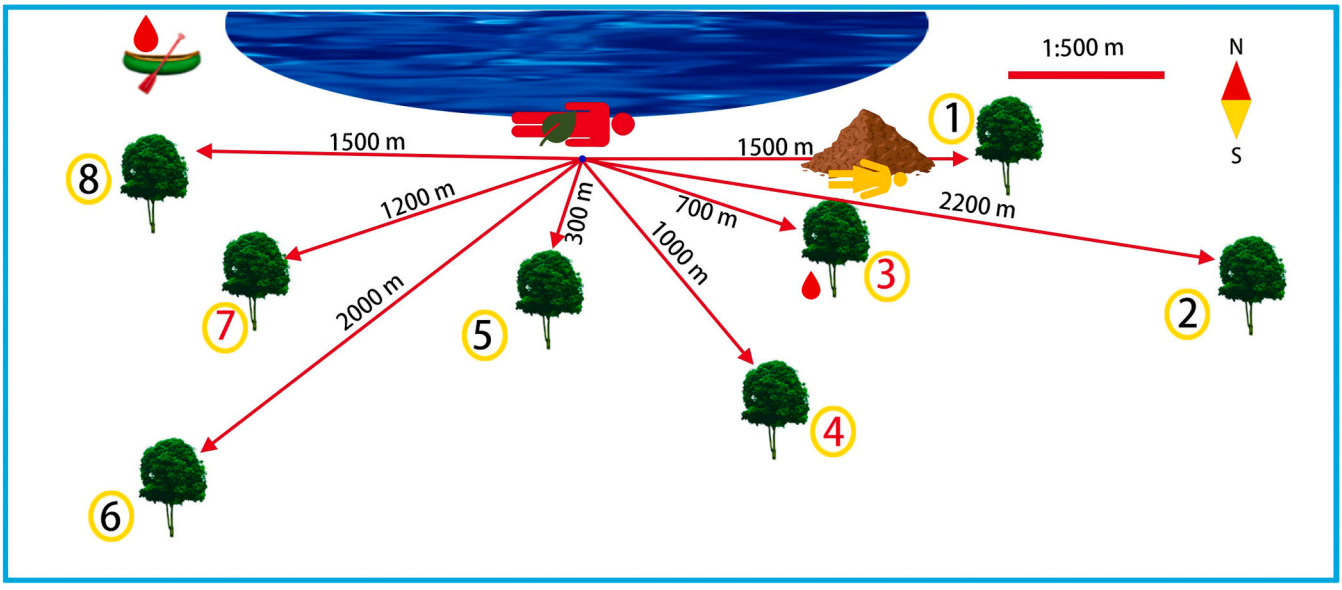
deceased exhibited knife wounds, but there was surprisingly little blood around the affected areas. A leaf was found on the body, but no trees were around. An investigation of a secluded fishing spot revealed bloodstains and a large *O. fragrans* tree. However, bloodstains found at the scene did not match the male deceased with human autosomal STR testing, which belonged to an unknown female (Amelogenin: XX). The leaves from this tree were collected by the investigators along with a leaf found on the body.

These plant samples from the above two cases were sent to our laboratory to establish the identity, i.e., whether the leaf found inside the vehicle (Case 1) or on the body (Case 2) came from one of the trees at the primary crime scene. Since botanical morphology inspection cannot serve this purpose, we have to conduct the following studies. The improved DNA extraction method was applied to the samples from Cases 1 and 2. Initially, seven previously reported loci (OF5, OF7, OF8, OF9, OF13, OF19, and OSM063) were used for individual identification. The identity was established between the half leaf in the car backseat and the *O. fragrans* tree next to the body in Case 1, as well as between the leaf on the deceased and the *O. fragrans* tree in the fishing spot in Case 2. As a precaution, we examined nine more *O. fragrans* trees from different locations around the body in Case 1 (Fig. 4). Surprisingly, alleles from three trees (including the tree next to the body) were matched. Each originated from different hills, making determining the actual crime scene impossible. A similar situation was observed in Case 2 (Fig. 5). Our analysis was related to fewer loci being detected, which prevented us from fully distinguishing the trees between the crime scene and the neighboring locations. In order to solve these cases, we must discover new *O. fragrans* STR loci (Section 2.3).

3.12.2. Case solution

Plant samples with identical genotypes at seven loci from different locations in Cases 1 and 2 were clearly distinguished using the in-house 12-plex system. Electropherograms of matching alleles in the two cases are presented in Figures S13 and S14, respectively. Fig. 4 shows that the half leaf in the car backseat indeed came from the *O. fragrans* tree next to the body (Tree 1) in Case 1 instead of Trees 3 and 9. The likelihood ratio (LR) was calculated as 3.5349×10^{27} ($\theta = 0$) based on allele frequencies in Section 3.7, and it was adjusted as 2.4275×10^{21} ($\theta = 0.03$, recommended by NRC II for small, isolated human populations where inbreeding is more likely) or 3.5357×10^{17} ($\theta = 0.07$, calculated in this study) according to NRC II equations 4.10a (for homozygous loci) and 4.10b (for heterozygous loci) [36]. Similarly, the leaf on the deceased actually belonged to the *O. fragrans* tree in the fishing spot (Tree 3) in Case 2, rather than to Trees 4 and 7 (as shown in Fig. 5). The LR value was 9.5687×10^{28} ($\theta = 0$), and its correction value was 3.2968×10^{22} ($\theta = 0.03$) or 1.8336×10^{18} ($\theta = 0.07$).

Case 1: Initially, the vehicle owner claimed he had never been to the crime scene. Once the botanical evidence established positive matches, he finally confessed that the female victim had suddenly fallen ill during a sexual encounter with him. Fearing exposure to this incident of



Sample	OF5		OF7		OF8		OF9		OF13		OF19		OSM063		GHSTR1		GHSTR2		GHSTR4		GHSTR5		GHSTR6	
Leaf on the deceased	17	17	19	22	43	43	14	14	25	32	24	33	7	14	18	18	7	16	6	18	18	18	11	18
Tree 1	8	8	15	19	28	43	14	14	25	33	25	25	7	23										
Tree 2	8	8	16	23	39	40	12	15	26	37	23	33	7	14										
Tree 3 (In the fishing spot)	17	17	19	22	43	43	14	14	25	32	24	33	7	14	18	18	7	16	6	18	18	18	11	18
Tree 4	17	17	19	22	43	43	14	14	25	32	24	33	7	14	19	24	11	16	8	8	11	14	13	19
Tree 5	21	21	20	21	37	42	14	14	31	34	23	26	11	14										
Tree 6	12	18	19	19	40	40	14	14	21	31	32	39	11	11										
Tree 7	17	17	19	22	43	43	14	14	25	32	24	33	7	14	19	23	7	11	8	10	14	14	23	24
Tree 8	18	19	28	29	40	40	14	14	22	27	26	36	11	16										

Fig. 5. Schematic diagram of the crime scene and genotypes in Case 2. A leaf on the deceased and leaves from a large *O. fragrans* tree (Tree 3) in the fishing spot are collected. Leaves of seven more *O. fragrans* trees (Trees 1–2 and 4–8), located in different areas around the body, are also gathered. Genotypes of the evidence sample are marked in grey. Alleles of trees are concordant with those of the evidence sample marked in green and discordant in red. Alleles from three trees (3, 4, and 7) are matched using seven previously reported loci (OF5, OF7, OF8, OF9, OF13, OF19, and OSM063). Nonetheless, they can be differentiated utilizing the 12-plex system. Electropherograms are presented in Fig. S14.

prostitution, the suspect left the deceased woman by the roadside and departed alone.

Case 2: Further searches were launched at the potential primary crime scene identified through botanical evidence. A female body was found buried in one of the turned-over dirt. Bloodstains in the fishing spot matched this female's DNA. Following the digital forensic investigation, the convict was apprehended and confessed to killing this woman. Additionally, the convict admitted that the deceased man was a fishing enthusiast who might have witnessed the murder and burial of the woman. The man was spotted by the convict and threatened with a knife to get into a boat, where the convict killed the man and then dumped the body into the lake. The male body was likely washed up on a far lakeside later. Finally, the latent bloodstains collected from the boat by investigators matched the deceased man's DNA.

4. Discussion

4.1. Advantages of the improved plant DNA extraction method

In forensic science, most plant specimens found at crime scenes are present in trace amounts, e.g., a half-leaf in Case 1. Therefore, extracting high-quality plant DNA in such cases is crucial to prevent contamination with exogenous DNA. The traditional CTAB method requires a large volume of specimen (at least 2.00 g), which must be ground with liquid nitrogen, resulting in some plant sample loss. Moreover, the traditional method is complex and can easily lead to contamination when specimens are ground with a mortar and pestle, which is not ideal for extracting plant DNA in forensic botany. Nowadays, there are very useful extraction methods (manual and automated) that can accurately extract DNA from as little as a few milligrams of plant material [40,41]. Experimental research indicates that our improved extraction method developed here requires only 1 % of the sample volume used in the traditional method yet yields 35 times the amount of extracted DNA per milligram under the same sample conditions (Section 3.1). The process of shredding a small amount of plant tissue and vigorously shaking it with 0.2 mm glass beads can effectively break cell walls and prevent sample loss. By modifying the DNA extraction and purification protocols, the maximum amount of usable DNA is obtained. The experiments performed in the present study used disposable reagents and consumables to avoid cross-contamination between samples. The improved DNA extraction method was highly suitable for soft plant tissues such as leaves, pollen, buds, and plant scrapes. It was not only rapid and straightforward but also highly resistant to foreign DNA contamination, making it highly effective for extracting forensic plant DNA for research purposes.

4.2. Advantages of the five novel STR loci

Historically, forensic identification of plant species was primarily achieved using the morphological method. The introduction of DNA technology has provided a powerful tool for understanding the genetic relationships between plant populations and individuals [5,21,42]. This study screened and characterized five novel STR loci—GHSTR1, GHSTR2, GHSTR4, GHSTR5, and GHSTR6—located on different chromosomes (Sections 3.2 and 3.7). All possess trinucleotide repeat units, rather than dinucleotides commonly employed in plant identification. They tend to exhibit decreased stutter propensity with longer repeat units [43], thereby facilitating the development of a multiplex system suitable for *O. fragrans* DNA analysis. Moreover, these five STRs exhibit high polymorphism ($H_{exp} > 0.8$ and PIC > 0.8) and high discriminatory capability (PD > 0.9). Therefore, they meet the criteria for individual identification within forensic contexts, highlighting the potential application of forensic botany in pointing to suspects and crime scenes. However, the identification of individuals must exclude samples from plants propagated as clones, e.g., from cuttings, grafts, and tissue culture, as these will share STR genotypes with the donor plant [42]. Due to variations in cultivation environments among asexually propagated *O. fragrans* plants, DNA methylation patterns will be examined in the future.

4.3. Advantages and limitations of the 12-plex system

By combining five newly identified *O. fragrans* STRs in this study with seven previously reported STRs (OF5, OF7, OF8, OF9, OF13, OF19, and OSM063), the 12-plex system further improves the power of the system effectiveness, e.g., individual identifications in two cases reported (Section 3.12). Developmental validation studies are conducted according to the ISFG and SWGDAM recommendations [19,20]. First, we collect 60 unique alleles observed in samples and sequence them to confirm the repeat number and potential repeat/flanking region variation (Section 3.2). Also, 164 virtual bins are added to facilitate data interpretation. High size precision (SD < 0.1 bp) and accuracy (variation within ± 0.5 bp) are confirmed for repeatability, ensuring reliable microvariant detection, and inter-laboratory concordance across genetic analyzers further validates reproducibility (Section 3.6). Second, the system demonstrates high species specificity, producing complete profiles only in *O. fragrans* with no reproducible peaks in other species (Section 3.3). PCR-based studies optimize standard thermal cycling parameters: amplification of 500 pg PTC using 58°C for annealing, 20 min for final elongation, and 28 cycles to obtain complete profiles and optimal performance with average peak heights of 8000–10000 RFU and average PHR > 0.7 (Section 3.9). Due to the characteristics of the selected PCR master mix (which includes an increased concentration of bovine serum albumin and Taq DNA polymerase, as per private communications with NuHigh Biotech), the system generates complete profiles at ≥ 50 pg input DNA (Section 3.4), deconvolutes ≥ 95 % unique alleles from minor contributors from mixtures between 1:19 and 19:1 ratio (Section 3.8), and has strong inhibitor tolerance (e.g., $\leq$500 μM hematin and $\leq$400 ng/μL humic acid in Section 3.5). Overall, forensic laboratories may use the proposed method for individual identifications of *O. fragrans* plants to decipher botanical clues for determining crime scenes. Although we constantly improve the protocols, this system still has limitations, e.g., a higher stutter ratio (~30 %) at the dinucleotide-repeat OF13 (Section 3.10) and a lower inter-color balance (~10 %) due to the red/blue dye violation (Section 3.11). We are attempting to minimize confounding stutter artifacts at dinucleotide-repeat loci by using the reduced stutter polymerase (RSP, provided by NuHigh Biotech), as shown in Figure S15. Additionally, allelic ladders are not as comprehensive at the current time as commercial human STR kits, i.e., many of the alleles are not represented in the allelic ladders. This is a major task, particularly in relation to dinucleotide-repeat STR loci. We may need to expand the sample size to identify additional unique homozygous alleles for creation. In the future, we will continue to improve these limitations.

5. Conclusion

The present study selects *Osmanthus fragrans* (*O. fragrans*), a species with a wide distribution in southern China, as the study subject. It starts by improving the plant DNA extraction method suitable for trace samples, followed by discovering highly polymorphic STR loci to enhance the power of discrimination and, ultimately, developing and validating the multiplex system to identify *O. fragrans* individuals. Moreover, two cases are reported involving the individual identification of *O. fragrans* to determine the crime scene without other investigative clues. These findings provide key clues for successfully solving the two cases. Also, they will inspire future research and development of different plant DNA applications in forensic science.

CRediT authorship contribution statement

Kewei Liang: Writing – original draft, Conceptualization. **Pengfei**

Li: Methodology. **Wei Zhang:** Validation. **Chun Shi:** Validation. **Bin Zhou:** Investigation. **Ting Wu:** Data curation. **Fu Ren:** Writing – review & editing, Supervision. **Fei Guo:** Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank Jianjun Peng at General Biological (Anhui) for providing technical support.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fsigen.2026.103440.

References

[1] I. Kasprzyk, Forensic botany: who?, how?, where?, when? Sci. Justice 63 (2023) 258–275, https://doi.org/10.1016/j.scijus.2023.01.002.
[2] Z.L. Zhang, W.B. Liang, H.B. Sun, X. Yang, L.Y. Ma, Z.X. Zheng, Forensic application of plant evidence, Fa Yi Xue Za Zhi 37 (2021) 87–90, https://doi.org/10.12116/j.issn.1004-5619.2019.490708.
[3] M. Caccianiga, C. Compostella, G. Caccia, C. Cattaneo, Contribution of plant anatomy to forensic investigation: Tree bark morphology, Forensic Sci. Int. 318 (2021) 110598, https://doi.org/10.1016/j.forsciint.2020.110598.
[4] H. Brabazon, J.M. DeBruyn, S.C. Lenaghan, F. Li, A.Z. Mundorff, D.W. Steadman, C. N. Stewart Jr, Plants to remotely detect human decomposition? Trends Plant Sci. 25 (2020) 947–949, https://doi.org/10.1016/j.tplants.2020.07.013.
[5] L.M.I. Webster, T.L. Prigge, G.J. Frankham, A guide for the validation of DNA based species identification in forensic casework, Forensic Sci. Int. Anim. Environ. 5 (2024) 100080, https://doi.org/10.1016/j.fsiae.2023.100080.
[6] M. Wesselink, A. Dragutinović, J.W. Noordhoek, L. Bergwerff, I. Kuiper, DNA typing of birch: Development of a forensic STR system for *Betula pendula* and *Betula pubescens*, Forensic Sci. Int. Genet. 35 (2018) 70–81, https://doi.org/10.1016/j.fsigen.2018.04.001.
[7] W.J. Koopman, I. Kuiper, D.J. Klein-Geltink, G.J. Sabatino, M.J. Smulders, Botanical DNA evidence in criminal cases: Knotgrass (*Polygonum aviculare* L.) as a model species, Forensic Sci. Int. Genet. 6 (2012) 366–374, https://doi.org/10.1016/j.fsigen.2011.07.013.
[8] M. Oliveira, L. Azevedo, D. Ballard, W. Branicki, A. Amorim, Using plants in forensics: State-of-the-art and prospects, Plant Sci. 336 (2023) 111860, https://doi.org/10.1016/j.plantsci.2023.111860.
[9] S. Ishak, E. Dormontt, J.M. Young, Microbiomes in forensic botany: a review, Forensic Sci. Med. Pathol. 17 (2021) 297–307, https://doi.org/10.1007/s12024-021-00362-4.
[10] H.Miller Coyle, C.L. Lee, W.Y. Lin, H.C. Lee, T.M. Palmbach, Forensic botany: using plant evidence to aid in forensic death investigation, Croat. Med. J. 46 (2005) 606–612.
[11] D.J. Cosgrove, Growth of the plant cell wall, Nat. Rev. Mol. Cell. Biol. 6 (2005) 850–861, https://doi.org/10.1038/nrm1746.
[12] J.G. Williams, A.R. Kubelik, K.J. Livak, J.A. Rafalski, S.V. Tingey, DNA polymorphisms amplified by arbitrary primers are useful as genetic markers, Nucleic Acids Res. 18 (1990) 6531–6535, https://doi.org/10.1093/nar/18.22.6531.
[13] E. Zietkiewicz, A. Rafalski, D. Labuda, Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification, Genomics 20 (1994) 176–183, https://doi.org/10.1006/geno.1994.1151.
[14] P. Vos, R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, AFLP: a new technique for DNA fingerprinting, Nucleic Acids Res. 23 (1995) 4407–4414, https://doi.org/10.1093/nar/23.21.4407.
[15] M.J. Bagley, S.L. Anderson, B. May, Choice of methodology for assessing genetic impacts of environmental stressors: polymorphism and reproducibility of RAPD and AFLP fingerprints, Ecotoxicology 10 (2001) 239–244, https://doi.org/10.1023/a:1016625612603.
[16] M. Di Nunzio, V. Agostini, F. Alessandrini, C. Barrot-Feixat, A. Berti, C. Bini, M. Bottinelli, E. Carnevali, B. Corradini, M. Fabbri, P. Fattorini, P. Garofano, S. Gino, A. Mameli, A. Marino, C. Previderè, C. Robino, C. Romano, P. Tozzo, A. Verzeletti, L. Buscemi, D. Gangitano, C. Di Nunzio, A Ge.F.I. - ISFG European collaborative study on DNA identification of Cannabis sativa samples using a 13-locus multiplex STR method, Forensic Sci. Int. 329 (2021) 111053, https://doi.org/10.1016/j.forsciint.2021.111053.
[17] R. Houston, M. Birck, S. Hughes-Stamm, D. Gangitano, Developmental and internal validation of a novel 13 loci STR multiplex method for Cannabis sativa DNA profiling, Leg. Med. (Tokyo). 26 (2017) 33–40, https://doi.org/10.1016/j.legalmed.2017.03.001.
[18] S. Köhnemann, J. Nedele, D. Schwotzer, J. Morzfeld, H. Pfeiffer, The validation of a 15 STR multiplex PCR for Cannabis species, Int. J. Leg. Med. 126 (2012) 601–606, https://doi.org/10.1007/s00414-012-0706-6.
[19] A. Linacre, L. Gusmão, W. Hecht, A.P. Hellmann, W.R. Mayr, W. Parson, M. Prinz, P.M. Schneider, N. Morling, ISFG: Recommendations regarding the use of non-human (animal) DNA in forensic genetic investigations, Forensic Sci. Int. Genet. 5 (2011) 501–505, https://doi.org/10.1016/j.fsigen.2010.10.017.
[20] cientific Working Group on DNA Analysis Methods (SWGDAM). Validation guidelines for DNA analysis methods, SWGDAM, SWGDAM, Washington, 2016. Available at: https://www.swgdam.org/_files/ugd/4344b0_813b241e8944497e99b9c45b163b76bd.pdf.
[21] G. Ferri, B. Corradini, F. Ferrari, A.L. Santunione, F. Palazzoli, M. Alu', Forensic botany II, DNA barcode for land plants: which markers after the international agreement? Forensic Sci. Int. Genet. 15 (2015) 131–136, https://doi.org/10.1016/j.fsigen.2014.10.005.
[22] Thermo Fisher Scientific, Qubit® dsDNA HS Assay Kits, P/N MAN0002326, Rev. C.0, Thermo Fisher Scientific, Waltham, 2022. Available at: (⟨https://assets.thermofisher.cn/TFS-Assets/LSG/manuals/Qubit_dsDNA_HS_Assay_UG.pdf⟩).
[23] Q. Li, J.M. Wan, SSRHunter: development of a local searching software for SSR sites, Yi Chuan 27 (2005) 808–810.
[24] Y. Ma, Development of Osmanthus SSR Primers and Establishment of SCoT Molecular Marking System, Henan University, 2015.
[25] Z.R. Zhang, D.M. Fan, S.Q. Guo, D.Z. Li, Z.Y. Zhang, Development of 29 microsatellite markers for Osmanthus fragrans (Oleaceae), a traditional fragrant flowering tree of China, Am. J. Bot. 98 (2011) e356–e359, https://doi.org/10.3732/ajb.1100241.
[26] P.M. Vallone, J.M. Butler, AutoDimer: a screening tool for primer-dimer and hairpin structures, Biotechniques 37 (2004) 226–231, https://doi.org/10.2144/04372ST03.
[27] R.A. Griffiths, M.D. Barber, P.E. Johnson, S.M. Gillbard, M.D. Haywood, C. D. Smith, J. Arnold, T. Burke, A.J. Urquhart, P. Gill, New reference allelic ladders to improve allelic designation in a multiplex STR system, Int. J. Leg. Med. 111 (1998) 267–272, https://doi.org/10.1007/s004140050167.
[28] A. Gouy, M. Zieger, STRAF 2: New features and improvements of the STR population data analysis software, Forensic Sci. Int. Genet 76 (2024) 103207, https://doi.org/10.1016/j.fsigen.2024.103207.
[29] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: molecular evolutionary genetics analysis version 6.0, Mol. Biol. Evol. 30 (2013) 2725–2729, https://doi.org/10.1093/molbev/mst197.
[30] F. Guo, H. Shen, H. Tian, P. Jin, X. Jiang, Development of a 24-locus multiplex system to incorporate the core loci in the Combined DNA Index System (CODIS) and the European Standard Set (ESS), Forensic Sci. Int. Genet. 8 (2014) 44–54, https://doi.org/10.1016/j.fsigen.2013.07.007.
[31] J.J. Doyle, J.L. Doyle, A rapid DNA isolation procedure for small quantities of fresh leaf tissue, Phytochem. Bull. 19 (1987) 11–15.
[32] S.J. Carey, L.E. Becklund, P.P. Fabre, J.J. Schenk, Optimizing the lysis step in CTAB DNA extractions of silica-dried and herbarium leaf tissues, Appl. Plant Sci. 11 (2023) e11522, https://doi.org/10.1002/aps3.11522.
[33] R. Xia, R. Tao, Y. Qu, X. Zhang, H. Yu, C. Yuan, S. Zhang, C. Li, Development and validation of a novel and fast detection method for Cannabis sativa: a 19-Plex short tandem repeat typing system, Front. Plant Sci. 13 (2022) 837945, https://doi.org/10.3389/fpls.2022.837945.
[34] R. Houston, M. Birck, S. Hughes-Stamm, D. Gangitano, Evaluation of a 13-loci STR multiplex system for Cannabis sativa genetic identification, Int. J. Leg. Med. 130 (2016) 635–647, https://doi.org/10.1007/s00414-015-1296-x.
[35] L. de Oliveira Pereira Ribeiro, E. Avila, R.F. Mariot, M.S. Fett, F.A. de Oliveira Camargo, C.S. Alho, Evaluation of two 13-loci STR multiplex system regarding identification and origin discrimination of Brazilian Cannabis sativa samples, Int. J. Leg. Med. 134 (2020) 1603–1612, https://doi.org/10.1007/s00414-020-02338-5.
[36] National Research Council Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence, National Academy Press, Washington, 1996.
[37] C. Schlötterer, D. Tautz, Slippage synthesis of simple sequence DNA, Nucleic Acids Res 20 (1992) 211–215, https://doi.org/10.1093/nar/20.2.211.
[38] P.J. Collins, L.K. Hennessy, C.S. Leibelt, R.K. Roby, D.J. Reeder, P.A. Foxall, Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpFlSTR Identifiler PCR Amplification Kit, J. Forensic Sci. 49 (6) (2004) 1265–1277.
[39] National Standardization Technical Committees for Forensic Science, Criterion of forensic science human fluorescence STR multiplex amplification reagent: GB/T 37226–2018, China Standard Press, Beijing, 2018.
[40] Qiagen, DNeasy® Plant Handbook, Rev. 03/2024, Qiagen, Germany, 2024. Available at: (https://www.qiagen.com/cn/resources/download.aspx?id=f6455f80-dc4f-4ff2-b2de-ae7a3e6c91e0&lang=en).
[41] Thermo Fisher Scientific, GMO Extraction Kit User Guide, P/N MAN0013474, Rev. D.0, Thermo Fisher Scientific, Waltham, 2024. Available at: (https://assets.thermofisher.cn/TFS-Assets/LSG/manuals/MAN0013474_GMO_ExtractionKit_UG.pdf).
[42] D.W. Hall, J.H. Byrd. Forensic Botany: A Practice Guide, first ed., Wiley-Blackwell, UK, 2012.
[43] J.M. Butler, Advanced Topics in Forensic DNA Typing: Interpretation, Academic Press, Boston, 2015.