2010 © The Japan Society for Analytical Chemistry

Notes

Potential Utility of DNA Sequence Analysis of Long-term-stored Plant Leaf Fragments for Forensic Discrimination and Identification

Hitomi S. Kikkawa,*† Ritsuko Sugita,* Rikyu Matsuki, and Shinichi Suzuki***

**Third Department of Forensic Science, National Research Institute of Police Science, 6-3-1 Kashiwanoha, Kashiwa, Chiba 277–0882, Japan*
***Biological Environment Sector, Central Research Institute of Electric Power Industry, 1646 Abiko, Abiko, Chiba 270–1194, Japan*

This study examined the potential utility of DNA sequence analysis to discriminate and identify plant material in forensic investigations. DNA was extracted from plant leaf fragments of 11 species stored for 5 to 22 years after collection. The *trnH-psbA* intergenic spacer and 316 bp of the *rbcL* gene were successfully amplified and sequenced for all fragments except for the *trnH-psbA* spacer of one sample. All of the plant samples were discriminated in pairwise comparisons of the sequences. Using a combination of local and global genetic databases is likely to provide greater reliability in search results to identify forensic samples from sequence data.

(Received April 13, 2010; Accepted June 7, 2010; Published August 10, 2010)

Introduction

Botanical evidence found at a crime scene can provide valuable information in criminal investigations. If tiny plant fragments found on a suspect can be proven to originate from a crime scene, they can be important evidence to associate the suspect with the site. In addition, taxonomic determination of an unidentified plant sample taken from a victim serves to identify the area where a crime was committed. At present, forensic plant samples are discriminated based on morphological observations.[1] However, botanical evidence at crime scenes can be fragmentary, preventing accurate species identification from morphological characteristics, and some plant organs can show strong morphological similarities among unrelated species. Therefore, techniques based on other sources of evidence, such as DNA sequences, have been investigated for their potential for identification of forensic plant samples.[2-4]

To be useful in providing botanical evidence in forensic investigations, DNA analytic methods must be applicable to long-term-stored and small fragments of plant material. In previous forensic studies of plant material, DNA was extracted from fresh samples or large amounts of tissues (0.5 – 1.0 g).[5,6] However, DNA extracted from long-term-stored forensic samples is reported to show lower polymerase chain reaction (PCR) amplification of DNA than that obtained from fresh material,[3] and botanical evidence at crime scenes often comprises small fragments of plant material. Although DNA isolated from archaeological plant samples has been analyzed successfully,[7,8] it is unclear whether DNA amplification is possible from small fragments of plants collected dozens of years previously. There is limited knowledge about the genomic composition of most plant species, and therefore whole-genome screening methods, such as RAPD[9,10] and AFLP[11,12] markers, have been utilized for DNA profiling. These methods are adequate for intact genomes, but reproducibility from degraded specimens is assumed to be problematic, preventing their use in routine forensic investigations. In recent years, direct sequencing of specific loci has been used for plant species discrimination and identification.[13-17] The length of these DNA regions is advantageous for forensic applications because many are shorter than 1 kb for most species and are more likely to be amplified successfully by PCR even in degraded samples.

In the present study, 5 mm × 5 mm (10 – 20 mg) leaf fragments from a variety of plant species stored for periods ranging from 5 to 22 years after collection were used for DNA sequence analysis. Nucleotide sequence analysis was performed, targeting the *rbcL* (ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit) gene[13,14] and the *trnH* (tRNA-His gene)-*psbA* (D1 reaction center protein of photosystem II gene) intergenic spacer,[15,16] which are commonly used as markers for phylogenetic and taxonomic studies. Both regions are relatively short and are recommended as suitable regions for species identification[17] as they evolve comparatively rapidly and therefore offer useful levels of interspecific variation in nucleotide sequence. Primers for these regions offering high universality across land plants have been developed.[14,15] Overall, these regions were considered suitable for exploration of their forensic applications. We thus examined whether plant leaf samples can be discriminated and identified using partial *rbcL* gene and *trnH-psbA* spacer sequence data.

Experimental

Materials

Healthy leaf samples of 11 plant species, which grew in areas of human habitation in Japan, were collected and their species identities were confirmed from morphological features.

† To whom correspondence should be addressed.
E-mail: kikkawa@nrips.go.jp

Table 1 Plant samples analyzed in the study

Sample No.	Family	Genus	Species	English vernacular name	Origin	Time after collection/year
1	Poaceae	*Imperata*	*cylindrica*	Blady grass	Okinawa	22
2	Poaceae	*Digitaria*	*ciliaris*	Southern crab-grass	Yamaguchi	21
3	Poaceae	*Miscanthus*	*sinensis*	Eulalia grass	Akita	21
4	Poaceae	*Pleioblastus*	*simonii*	Simon bitter bamboo	Akita	21
5	Poaceae	*Chimonobambusa*	*marmorea*	Kan-tiku	Kyoto	12
6	Poaceae	*Cortaderia*	*selloana*	Pampas grass	Kyoto	12
7	Poaceae	*Phyllostachys*	*aurea*	Fish pole bamboo	Kyoto	12
8	Poaceae	*Semiarundinaria*	*fastuosa*	Temple bamboo	Kyoto	11
9	Rosaceae	*Rosa*	Unidentified	Rose	Unidentified	11
10	Ericaceae	*Vaccinium*	*vitis-idaea*	Cowberry	Akita	6
11	Pinaceae	*Pinus*	*parviflora*	Japanese white pine	Akita	5

The samples used in this study (Table 1) were dried immediately after collection and stored at room temperature in a dry room for periods ranging from 5 to 22 years before experimental analysis. Ten samples (numbers 1 to 8, 10, 11) were stored out of direct sunlight, and sample 9 was constantly exposed directly to sunlight.

Total genomic DNA was extracted from leaf samples measuring approximately 5 mm × 5 mm (10 – 20 mg) using the DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany) in accordance with the manufacturer's instructions and stored at –20°C until analysis. The concentration of extracted DNA for each sample was determined with a U-0080D photodiode array spectrophotometer (Hitachi High-Technologies, Tokyo, Japan).

PCR amplification

A partial *rbcL* gene sequence from base positions +105 to +420 (with the translational start site in *Nicotiana tabacum* [DDBJ accession number J01450] designated as +1) was amplified by PCR with the primer set *rbcL*-F3 (5′-TAT CTT GGC AGC ATT CCG AGT AAC TCC-3′) and *rbcL*-R3 (5′-GAT TCG CAG ATC CTC CAG ACG TAG AGC-3′).[14] Amplification was carried out using a MyCycler thermal cycler (Bio-Rad Laboratories, Hercules, CA) in a total volume of 10 μL. Each reaction contained 200 μM dNTPs, 1 mM $MgSO_4$, 0.2 μM of each primer, 0.2 U KOD plus DNA polymerase (Toyobo, Osaka, Japan), 1× KOD plus reaction buffer, and 1 μL DNA extract. The PCR conditions were as follows: 98°C for 1 min, 45 cycles of 98°C for 10 s, 56°C for 30 s, and 72°C for 45 s, with a final extension at 72°C for 5 min.

The *trnH-psbA* intergenic spacer was amplified by PCR with the primer set *psbA*3f (5′-GTT ATG CAT GAA CGT AAT GCT C-3′) and *trnHf* (5′-CGC GCA TGG TGG ATT CAC AAT CC-3′)[15] as described previously. Amplification was carried out using a MyCycler in a total volume of 10 μL. Each reaction contained 200 μM dNTPs, 0.2 μM of each primer, 0.01% BSA, 0.5 U Ex Taq DNA polymerase (TaKaRa Bio, Otsu, Japan), 1× Ex Taq buffer, and 3 μL DNA extract. The PCR conditions were as follows: 98°C for 1 min, 5 cycles of 98°C for 10 s, 50°C for 30 s, and 72°C for 60 s, 35 cycles of 90°C for 10 s, 55°C for 30 s, and 72°C for 60 s, with a final extension at 72°C for 5 min.

The amplified products of both regions were detected on a 2% agarose gel and purified with the MinElute PCR Purification Kit (QIAGEN).

DNA sequencing

Cycle sequencing reactions were carried out using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) with a MyCycler in a total volume of 20 μL. Direct sequencing of the PCR products was performed using both forward and reverse primers. The cycle sequencing products were analyzed with POP-7 Performance Optimized Polymer (Applied Biosystems) and detected using an ABI PRISM 3130 Genetic Analyzer (Applied Biosystems). The partial *rbcL* sequence from base positions +132 to +393 and the full *trnH-psbA* spacer region were determined.

Data analysis

Searches were performed for similar sequences lodged in (1) the Japanese local plant DNA database of the Central Research Institute of Electric Power Industry (CRIEPI),[14] and (2) the GenBank, European Molecular Biology Laboratory (EMBL), and DNA Data Bank of Japan (DDBJ) databases using the Basic Local Alignment Search Tool (BLAST) accessible through the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/BLAST/). The CRIEPI database was established with approximately 950 *rbcL* sequences and 850 *trnH-psbA* sequences mainly from samples collected in northeastern Japan.

Results and Discussion

DNA extraction and PCR amplification

Plant samples in this study were various species and were stored in various conditions by assuming forensic unidentified plant samples. The grass family (Poaceae) are common plant species encountered in botanical evidence because of their wide range of habitats,[5] and often show similar morphological characteristics among species. Therefore, in the present study, many Poaceae samples were examined. Three commercially available kits of extracting DNA were tested; the DNeasy Plant Mini Kit, which performed PCR amplification most successfully, was selected. The concentrations of DNA recovered in the extracts ranged from 4.8 ng μL^{-1} (*Phyllostachys aurea*) to 15.6 ng μL^{-1} (*Pleioblastus simonii*).

PCR amplification of both DNA regions was successful for all 11 samples except for the *trnH-psbA* spacer of the *Rosa* sp. (Fig. 1), even though the method is used for extracts from fresh material. These results indicated the potential for DNA extraction and PCR amplification from trace amounts of long-term-stored plant samples. The *rbcL* PCR products for all samples were 316 base pairs (bp) in length (Fig. 1a), while the

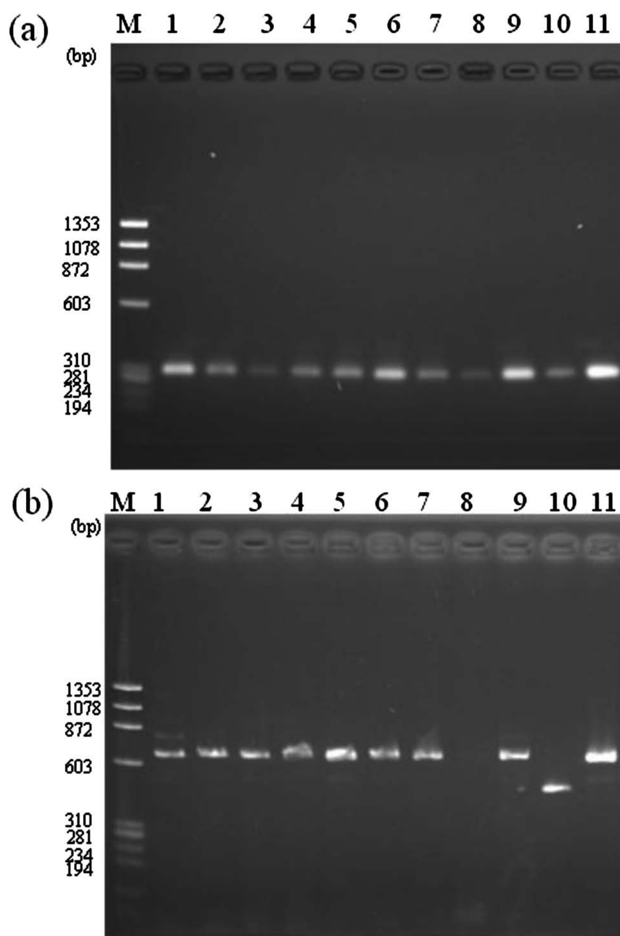

Fig. 1 PCR amplification of *rbcL* (a) and *trnH-psbA* (b). Lane: 1, *Imperata cylindrica*; 2, *Digitaria ciliaris*; 3, *Miscanthus sinensis*; 4, *Pleioblastus simonii*; 5, *Cortaderia selloana*; 6, *Chimonobambusa marmorea*; 7, *Phyllostachys aurea*; 8, *Rosa*; 9, *Semiarundinaria fastuosa*; 10, *Vaccinium vitis-idaea*; 11, *Pinus parviflora*. M, ϕX174-HaeII marker.

length of the *trnH-psbA* spacer differed widely (476 – 678 bp) among the samples (Fig. 1b). The length of the *trnH-psbA* spacer tended to be similar in related species. No correlation between the concentrations of DNA recovered and the yields of PCR products was observed. It is possible that DNA quality and affinity for primers may affect the PCR amplification. Low-quality degraded DNA extracted after an extended period of storage is difficult to amplify, especially for long regions.[3,18] In this study, the *trnH-psbA* spacer was longer than the portion of the *rbcL* gene amplified. Furthermore, the *trnH-psbA* spacer of fresh *Rosa* sp. could be amplified and its sequence is registered in CRIEPI database, raising the possibility that the *trnH-psbA* spacer of the *Rosa* sp. could not be amplified because of damage to this region during the long storage period and from direct exposure to ultraviolet rays in sunlight. Further studies are needed to understand the mechanism of this PCR failure. However, a partial *rbcL* gene sequence was obtained for the *Rosa* sp. Thus when one DNA region cannot be amplified, information about the species can be obtained from analysis of other regions, showing the utility of sequence analysis of multiple regions for identification of a specimen.

Table 2 Discrimination of pairs among the plant samples by comparison of the *rbcL* and the *trnH-psbA* sequences

Locus	Distinguishable pairs of families, %	Distinguishable pairs of genera, %	Distinguishable pairs of species, %
rbcL	100	96.4	96.4
trnH-psbA	100	100	100
rbcL + *trnH-psbA*	100	100	100

Discrimination of long-term-stored plant samples by comparison of DNA sequences

All of the PCR products were successfully sequenced using both primer sets, and alignment of the sequences was performed. Pairwise discrimination of the plant samples was performed for both the *rbcL* and *trnH-psbA* sequences. The number of pairwise comparisons was 55 for *rbcL* (= ${}_{11}C_2$) and 45 for *trnH-psbA* (= ${}_{10}C_2$). If the sequences from two different species/genus/family samples differed, the samples were deemed to be distinguishable. The results are summarized in Table 2. Comparisons of the partial *rbcL* gene and the *trnH-psbA* spacer showed that the sequences were characteristic for each species with the exception of two pairs for *rbcL* (*Pleioblastus simonii* and *Semiarundinaria fastuosa*, and *Phyllostachys aurea* and *Chimonobambusa marmorea*), but these species were discriminated by the *trnH-psbA* sequences. The results indicated that discrimination of all pairs among the plant samples was possible by comparison of the sequences for the two DNA regions.

Plant identification using searchable DNA sequence databases

To examine whether the plant samples could be taxonomically identified from the sequence data, we performed searches for similar sequences in (1) the CRIEPI database, and (2) the GenBank, EMBL, and DDBJ databases using the BLAST tool. The search results for each DNA region are shown in Table 3. The searches showed that the correct name, or the name of a closely related species, could be retrieved for all samples for both regions, indicating that the DNA amplified for each sample was authentic.

In the CRIEPI database, five samples showed 100% similarity with registered species names for both the *rbcL* and *trnH-psbA* sequences. *Cortaderia selloana* that could only be matched at the family level was not native to Japan. Most introduced species are not registered in the CRIEPI database,[14] therefore, searches for such species may only agree at the family level. The introduced species *Cortaderia selloana* was correctly identified to higher taxonomic levels (species level) in the NCBI BLAST search. Five *rbcL* and six *trnH-psbA* sequences were matched at the species level in NCBI BLAST searches. The sequence similarity of these samples ranged from 99.2% (*Miscanthus sinensis rbcL*; 2 bp difference) to 100%. However, the similarity for the *rbcL* of *Miscanthus sinensis* showed 100% homology in the CRIEPI database. Some species, such as *Miscanthus sinensis rbcL*, exhibited intraspecific sequence variation for the same region. Japan is an archipelago of islands and possesses many native and unique species. In addition, naturalized and introduced plants are also widely present in Japan, as in other countries. Thus, using a local database to identify local species and a global database for introduced species is likely to provide greater reliability in search results for unidentified forensic botanical samples. For 8 samples the correct species names was identified using a combination of the

Table 3 Results of similarity searches of *rbcL* and *trnH-psbA* sequences from the CRIEPI database and BLAST searches of NCBI databases (number of samples)

Correct identification level	*rbcL*			*trnH-psbA*		
	CRIEPI database search	NCBI BLAST search	The best similarity result[a]	CRIEPI database search	NCBI BLAST search	The best similarity result[a]
Species	5	5	7	5	6	8
Genus	3	3	2	2	3	1
Subfamily	2	3	2	3	1	1
Family	1	0	0	0	0	0

a. The best similar identification level obtained by a combination of the CRIEPI database and NCBI BLAST searches.

CRIEPI database and NCBI BLAST searches. These findings demonstrated the potential utility of DNA sequence analysis for plant sample identification in forensic cases.

In conclusion, this study demonstrated that PCR amplification and cycle sequencing of the partial *rbcL* gene sequence and the *trnH-psbA* spacer can be applied to most small fragments of plant leaves stored for periods ranging from 5 to 22 years after collection. The species successfully identified are found in areas of human habitation and therefore the technique could be useful to provide forensic evidence at crime scenes. Discrimination of all pairs among the plant samples was possible by comparison of the sequences for the two DNA regions analyzed. Utilization of both local and global genetic databases is recommended to obtain useful information about unidentified forensic samples.

Acknowledgements

The authors thank the members of the Central Research Institute of Electric Power Industry for providing the long-term plant samples, the database of Japanese plant sequences, and for their helpful advice. The authors also thank Dr. Jiro Yasuda for use of the facilities.

References

1. J. H. Bock and D. O. Norris, *J. Forensic Sci*., **1997**, *42*, 364.
2. H. Korpelainen and V. Virtanen, *J. Forensic Sci*., **2003**, *48*, 804.
3. K. J. Craft, J. D. Owens, and M. V. Ashley, *Forensic Sci. Int*., **2007**, *165*, 64.
4. E. J. Lee, S. C. Kim, I. K. Hwang, H. J. Yang, Y. S. Kim, M. S. Han, M. S. Yang, and Y. H. Lee, *J. Forensic Sci*., **2009**, *54*, 721.
5. J. Ward, R. Peakall, S. R. Gilmore, and J. Robertson, *Forensic Sci. Int*., **2005**, *152*, 121.
6. G. Ferri, M. Alu, B. Corradini, and G. Beduschi, *Int. J. Legal Med*., **2009**, *123*, 395.
7. D. L. Erickson, B. D. Smith, A. C. Clarke, D. H. Sandweiss, and N. Tuross, *Proc. Natl. Acad. Sci. U. S. A*., **2005**, *102*, 18315.
8. D. Paffetti, C. Vettori, D. Caramelli, C. Vernesi, M. Lari, A. Paganelli, L. Paule, and R. Giannini, *BMC Evol. Biol*., **2007**, *7(Suppl. 2)*, S6.
9. R. E. Veilleux, L. Y. Shen, and M. M. Paz, *Genome*, **1995**, *38*, 1153.
10. L. Congiu, M. Chicca, R. Cella, R. Rossi, and G. Bernacchia, *Mol. Ecol*., **2000**, *9*, 229.
11. P. Vos, R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau, *Nucleic Acids Res*., **1995**, *23*, 4407.
12. J. Jakse, K. Kindlhofer, and B. Javornik, *Genome*, **2001**, *44*, 773.
13. L. Gielly and P. Taberlet, *Mol. Biol. Evol*., **1994**, *11*, 769.
14. R. Matsuki, K. Shimano, S. Abe, H. Yatake, T. Takeuchi, A. Shiraki, T. Ishii, and M. Nashimoto, "*Study on the Ecosystem Sustaining a Pair of Golden Eagles —Identification of Food Plants by DNA Analysis from Animal Feces—*", **2003**, Central Research Institute of Electric Power Industry, Tokyo, 2.
15. W. J. Kress, K. J. Wurdack, E. A. Zimmer, L. A. Weigt, and D. H. Janzen, *Proc. Natl. Acad. Sci. U. S. A*., **2005**, *102*, 8369.
16. C. Sass, D. P. Little, D. W. Stevenson, and C. D. Specht, *PLoS One*, **2007**, *2*, e1154.
17. CBOL Plant Working Group, *Proc. Natl. Acad. Sci. U. S. A*., **2009**, *106*, 12794.
18. M. F. Deguilloux, M. H. Pemonge, and R. J. Petit, *Proc. Biol. Sci*., **2002**, *269*, 1039.