JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, November 2009, Vol. 54, No. 6
doi: 10.1111/j.1556-4029.2009.01196.x
Available online at: interscience.wiley.com

Jodie Ward,[1,†] *Ph.D.; Simon R. Gilmore,*[1,2] *Ph.D.; James Robertson,*[3] *Ph.D.; and Rod Peakall,*[1] *Ph.D.*

A Grass Molecular Identification System for Forensic Botany: A Critical Evaluation of the Strengths and Limitations*

ABSTRACT: Plant material is frequently encountered in criminal investigations but often overlooked as potential evidence. We designed a DNA-based molecular identification system for 100 Australian grasses that consisted of a series of polymerase chain reaction assays that enabled the progressive identification of grasses to different taxonomic levels. The identification system was based on DNA sequence variation at four chloroplast and two mitochondrial loci. Seventeen informative indels and 68 single-nucleotide polymorphisms were utilized as molecular markers for subfamily to species-level identification. To identify an unknown sample to subfamily level required a minimum of four markers or nine markers for species identification. The accuracy of the system was confirmed by blind tests. We have demonstrated "proof of concept" of a molecular identification system for trace botanical samples. Our evaluation suggests that the adoption of a system that combines this approach with DNA sequencing could assist the morphological identification of grasses found as forensic evidence.

KEYWORDS: forensic science, forensic botany, molecular identification system, grasses, indels, single nucleotide polymorphisms, polymerase chain reaction, species identification, chloroplast DNA, mitochondrial DNA

Botanical Evidence

Plant material is frequently encountered in criminal investigations but often overlooked as potential evidence due to our present inability to rapidly, accurately, and cost-effectively identify trace botanical specimens. Grasses were used as the model species in this study because they are among the plant species most likely to be encountered as forensic trace evidence. Grasses have considerable potential as contact DNA evidence and could provide links between crime scenes and individuals because of their ubiquitous nature in both urban and rural environments, their frequent utilization by humans, and the grass spikelet's morphological adaptations for seed dispersal (1).

Species Identification of Botanical Evidence

A number of methods are currently utilized to identify the biological origin of an unknown sample to higher taxonomic levels or for discriminating between closely related and morphologically similar organisms. These methods have been applied in forensic investigations (2–4), for biodiversity evaluation (5), and for the species or cultivar identification of plants (6–11). An important consideration in the development of a molecular system for species identification is that despite the availability of a range of DNA profiling techniques, not all techniques will meet forensic standards or be transferable among forensic laboratories because they suffer from problems of specificity, reproducibility, and profile complexity, and can therefore be discarded as evidence in the courtroom. In addition with many forensic samples, the quality, quantity, and purity of DNA is poor, introducing problems not routinely encountered in regular DNA profiling studies.

Forensic DNA profiling presently favors fragment analysis over direct sequencing. From a forensic perspective, an identification method needs to be robust and standardized whilst being able to be variable enough to discriminate among individuals and degraded samples. The most promising tools for forensic botany, which are currently in use in forensics involving human identification, are short tandem repeats (STRs) and single nucleotide polymorphisms (SNPs). Another molecular identification system that has risen in popularity since commencing this research is DNA barcoding.

DNA Barcoding

Hebert et al. (12) proposed that microgenomic identification systems, which employ DNA sequences from a uniform locality on the genome as taxon "barcodes," could provide a new approach to species identification. Thus far, work on animals has employed a standard region of the mitochondrial (mtDNA) genome (a highly variable 648-bp region near the 5′ end of the cytochrome oxidase subunit I [COI] gene) to provide species-specific DNA barcodes (12). DNA barcoding has been used to survey several taxa in their natural habitats. These include birds (13), amphibians (14), spiders (15), fishes (16), insects (17–20), fungi (21), and primates (22).

DNA barcoding in plants presents challenges not encountered in the early work on DNA barcoding in animals. Most significant has been the difficulty of finding a single variable barcoding gene in plants. The COI gene is not appropriate for distinguishing species of plants because genes in the mtDNA of plants evolve at extremely conservative rates (23) and exhibit limited variation. It has

[1]School of Botany and Zoology, The Australian National University, Canberra, ACT 0200, Australia.
[2]Centre for Forensic Science, Canberra Institute of Technology, GPO Box 826, Canberra, ACT 2601, Australia.
[3]National Manager, Forensic and Data Centres, Australian Federal Police, GPO Box 401, Canberra, ACT 2601, Australia.
*Funding provided by the Australian Federal Police (Forensic and Data Centres) and The Australian National University.
†Present Address: Centre for Forensic Science, Canberra Institute of Technology, GPO Box 826, Canberra, ACT 2601, Australia.
Received 24 July 2008; and in revised form 28 Oct. 2008; accepted 14 Dec. 2008.

proven to be useful only in identifying some species of algae (24). Several other loci have been proposed for the species identification of flowering plants and some authors have proposed a multilocus tiered approach for DNA barcoding of land plants (25–29).

Presently, there is no easy solution for the DNA barcoding of plants. De Ley et al. (30) proposed a combination of morphological and molecular strategies could overcome some of the criticisms associated with barcoding as a species identification tool. Here, we evaluate an alternative DNA-based identification system that could work in conjunction with the morphological identification of botanical specimens.

A Grass Molecular Identification System

Prior to the publication of many of the recent DNA-barcoding studies, we previously demonstrated the in-principle feasibility of an alternative approach—a molecular identification system for grasses (31). This identification system consisted of a series of polymerase chain reaction (PCR) assays that, like a traditional morphological taxonomic key, enabled the progressive identification of grass samples to a given taxonomic rank. We confirmed that by applying five PCR assays spanning variable insertions and deletions (indels) from two mitochondrial loci, identification was possible for our control set of 20 samples, on which the design of the PCR assays was based, and for 25 unknown grass samples (with some restrictions) (32).

The current research reported here builds on our previous study by expanding the identification system to include 100 representative grasses. Because it will never be feasible to sample all known grass species, our selection of diagnostic taxon-specific markers was made in the context of a sound phylogenetic framework. We predicted that this would maximize the likelihood of finding molecular markers at a given taxonomic level that would be diagnostic of both represented and unrepresented grass samples, thus allowing unknown samples to be identified to some taxonomic level. To assess the accuracy of this system, a blind test was conducted. We conclude with a critical evaluation of the strengths and limitations of this molecular identification system.

Building on our previous study, the specific objectives were:

- To expand the hierarchical sample design to include 100 native and exotic Australian grasses representing the nine major grass subfamilies.
- To expand genome coverage in search of potentially diagnostic indel and SNPs variation by sequencing chloroplast (*rbc*L, *trn*L-*trn*F, *rpl36*-*rps*8, *trn*T2-*rps*4) and mitochondrial (*nad* 7, *atp*A) loci.
- To construct a DNA-based molecular identification system that is built on a phylogenetically sound framework and consists of a series of taxon-specific PCR assays that enable the progressive identification of grass samples.
- To critically evaluate the forensic scope and limitations of this grass molecular identification system.

Methods

Study System

The Poaceae are one of the world's largest flowering plant families, comprising *c*. 10,000 species, inhabiting most ecological habitats (33,34). The most comprehensive subfamilial classification of the Poaceae to date was proposed by the Grass Phylogeny Working Group (GPWG) in 2001 (35). Phylogenetic analyses of these data confirmed: (i) there were 12 monophyletic subfamilies of the Poaceae; (ii) three subfamilies Anomochlooideae, Pharoideae, and Puelioideae were early diverging lineages; (iii) the remaining nine subfamilies form a clade—that is further divided into two sister clades: the PACCAD clade (Panicoideae, Arundinoideae, Chloridoideae, Centothecoideae, Aristidoideae, Danthonioideae) and the BEP clade (Bambusoideae, Ehrhartoideae, Pooideae).

Sampling

A total of 100 grass taxa representing the nine subfamilies within the BEP and PACCAD clades formed the basis of our molecular identification system (refer to Appendix 1 for species details).[1] Sample selection initially followed the phylogenetic framework of the GPWG in order to ensure a wide selection of representative samples, while at the same time providing adequate representation of the three largest subfamilies (Pooideae, Panicoideae, and Chloridoideae [36]). Grass seeds were germinated in sterile soil in glasshouse conditions to provide fresh growing leaf tissue for DNA. In addition to the 100 grass taxa, a selection of herbarium, seed, and roadside collected grass material was obtained to provide examples of forensic-type material. Voucher specimens for all fertile grass samples have been lodged with the Australian National Herbarium (CSIRO, Canberra). Refer to Appendix 1 for source details and voucher accession numbers.

DNA Extraction, PCR and DNA Sequencing

Detailed methods are provided in Ward et al. (32) and Ward (37). Here we provide only a brief overview. DNA extractions were performed using 100 mg of liquid nitrogen ground fresh leaf tissue with the exception of the forensic-type material for which either 10 seeds or 100 mg of dried leaf tissue was used. For all extractions, QIAGEN DNeasy® Plant Mini Kits (QIAGEN, Valencia, CA) were used following the manufacturer's instructions. DNA yield and quality was assessed against standards by agarose-gel electrophoresis.

In order to locate potential diagnostic characters, four chloroplast loci (*rbc*L, *trn*L-*trn*F, *rpl36*-*rps*8, *trn*T2-*rps*4) and two mitochondrial (*nad* 7, *atp*A) loci were the target of DNA sequencing. These six loci were targeted as they represented both coding and noncoding regions from two different plant genomes, they were prevalent in the phylogenetic literature, they produced single PCR products of <1500 bp in size, and they showed evidence of variation across taxa at one or more taxonomic levels.

The primer sequences and sources are shown in Table 1. Approximately 40 ng of template DNA was used in 40 μL PCR reactions consisting of 1× PCR reaction buffer (QIAGEN: 1.5 mM $MgCl_2$, Tris–HCl, KCl, and $(NH_4)_2SO_4$), 200 μM of each dNTP, 0.2 μM each of forward and reverse primer, and 1 U of Taq DNA polymerase (5 U/μL) (QIAGEN). Amplification was performed in a Corbett Research Thermal Cycler with an initial cycle at 94°C for 3 min, followed by 30 cycles of 94°C for 30 sec, 55°C for 30 sec, and 72°C for 2 min, with a final extension of 72°C for 10 min.

All PCR reactions produced single bands that were ethanol precipitated before further purification of the sequencing products using AutoSeq96 Plates (Amersham BioSciences, Piscataway, NJ) as per manufacturer's instructions. The purified sequencing products were then sequenced in both forward and reverse directions using the original primers and BigDye Terminator Cycle Sequencing Kits (Applied Biosystems, Foster City, CA) on an ABI 3100 automated sequencing instrument as per manufacturer's instructions. Sequence

[1]All appendices referred to in this paper can be freely sourced at http://www.anu.edu.au/BoZo/grass_id.

TABLE 1—*The primer name, sequence, and annealing temperature for the six primer pairs used for PCR amplification and sequencing of the six loci.*

Locus	Primer Name	Primer Sequence 5′–3′	Tm (°C)	References
cpDNA loci				
*rbc*L—subunit of ribulose 1, 5-biphosphate carboxylase coding locus	rbcL F	TCACCACAAACAGAAACTAAAGC	58	This study
	rbcL R	TGCTTTAGCTAATACACGGAAAT	58	This study
*rpl*36-*rps*8—noncoding intergenic spacer region	rpl36	ATTCTACGTGCACTCTTCCG	56	38
	rps8	CGAGGTATAATGACAGATCGAG	56	38
*trn*T2-*rps*4—noncoding intergenic spacer region	trnT2	CTGTAGGTGTAACCTTTCGC	56	38
	rps4	TCSTATTCCTGCAGTACAGG	56	38
*trn*L-*trn*F—noncoding intergenic spacer region	C	CGAAATCGGTAGACGCTACG	55	39
	F	ATTTGAACTGGTGACACGAG	55	39
mtDNA loci				
*atp*A—alpha subunit of F-1-ATPase coding locus	atpA-F1	AAGTGGATGAGATCGGTCGAG	55	40
	atpA-B1	GGCATTCGATCACAGA	55	40
nad 7—intron 1 in subunit 7 of NADH dehydrogenase locus	nad 7/1	ACCTCAACATCCTGCTGCTC	47	41
	nad 7/2r	CGATCAGAATAAGGTAAAGC	47	41

All the universal and grass-specific primer pairs were obtained from the literature as listed, except for the rbcL primer pair which was designed for this study.

chromatograms were checked and edited using Sequencher, Version 3.0 (Gene Codes Corporation, Ann Arbor, MI).

Location of Variation

Based on the sequence data from the six loci, a comprehensive phylogenetic analysis across the 100 grass samples was performed (see Ward [37] for the full details, which are beyond the scope of this paper). Importantly, this analysis revealed a phylogeny in general agreement with the GPWG phylogeny (35) indicating that this combination of six loci offered robust phylogenetic signal, with the ability to resolve the taxa and their taxonomic groupings accurately. Therefore, these six loci potentially offered a range of diagnostic markers at different taxonomic levels for the molecular identification system. However, one important finding was that a combination of loci was required to recover an accurate phylogeny. This also indicated that a single locus would not offer sufficient sequence variation to enable the identification of all grass species.

Our next task was to identify diagnostic indel and SNP variants at various taxonomic levels within the context of the phylogeny. We reasoned that by selecting diagnostic markers within a sound phylogenetic framework, this would assist in the identification of unknown samples (i.e., samples not used in the design of the identification system). To further assist our discovery of diagnostic markers, sequence alignments were divided into subfamily, tribal, or genus-specific alignments for each locus. A total of 116 informative indels were located at five of the six loci: *rpl*36-*rps*8, *trn*T2-*rps*4, *trn*L-*trn*F, *atp*A, and *nad* 7, while across all six loci (*c.* 6200 bp), 1954 variable SNP positions were located.

Primer Design and Optimization of Taxon-Specific Molecular Markers

Our molecular identification system was designed such that the PCR assays targeted informative DNA sequences surrounding variable indels or SNPs. In total, 85 diagnostic characters were identified for the design of taxon-specific molecular markers, which included 17 informative indels and 68 SNPs. These diagnostic markers were selected because they enabled a hierarchical design to our system. Using the software Primer3 (42), primer design considerations included the need for conserved sequence flanking the variable region and a fragment size of <500 bp in size to enhance the likelihood that amplification would occur with degraded DNA. Primer names, sequences, PCR conditions, and product sizes are summarized in Appendix 2.

For large indel molecular markers (>14 bps), a single forward and reverse primer was anchored in the conserved flanking regions on either side of the deletion such that differences in fragment size due to the indel were readily resolved by agarose. For smaller indel molecular markers (3–14 bps), the novel PCR assay included three primers in a single PCR reaction, as described in Ward et al. (32). Two primers (F and R) were designed in the conserved flanking regions on either side of the indel, while an internal forward insert primer (F1) was also designed to match the last three base pairs of the insertion. Thus, this assay produced two bands for an insertion (control and insert fragment), while a single band was produced for the deletion. By further manipulating the F1 primer size, it was possible to ensure that the two products could be readily resolved by agarose electrophoresis (product size differences ranged from 23 to 251 bp).

In order to identify the SNPs by an agarose-based assay, we employed an allele-specific PCR (AS-PCR) approach. Again we employed a three primer PCR assay similar to that described above to ensure a positive control product. However, two alternative PCR reactions were required for each SNP with the two allele-specific SNP primers designed to match the SNP on the 3′ end of the primer. Primer positions were again designed to ensure agarose-based resolution of the two PCR fragments (product size differences ranged from 33 to 322 bp) when the SNP was present. An additional mismatch was introduced at either the second or third base from the 3′ end of the allele-specific primer to increase the specificity of these primers following the recommendations of other studies (43–45).

Test of the Molecular Identification System

To test the performance of the molecular identification system, a total of 91 samples consisting of 24 grass species represented in the molecular identification system design, 22 grass species not included in the design, 16 forensic-type (seed, dried leaf, herbarium material) grass samples, and 24 grass samples collected from an Australian Capital Territory (ACT) roadside were assayed. In addition, five duplicate samples were included to check the reproducibility of the results. Refer to Appendix 4 for species information and voucher accession numbers. To avoid any bias in the scoring of the results, they were scored "blind" such that the species

identity was not known to the first author (JW) during the laboratory and scoring phases. A set of control samples were also amplified along with the unknown samples for each PCR assay to act as positive controls and to provide size standards.

Results

Sequence Variation

The mitochondrial coding and noncoding loci (*atp*A and *nad* 7 intron respectively) exhibited the least sequence variation (9.5% of 1210 bp and 18.48% of 1001 bp respectively), the coding chloroplast locus (*rbc*L) exhibited moderate sequence variation (20% of 800 bp), while the noncoding chloroplast spacer regions (*rpl*36-*rps*8, *trn*T2-*rps*4, *trn*L-*trn*F) exhibited the most total sequence variability (38.84% of 1066 bp, 49.2% of 998 bp, and 53.35% of 1106 bp respectively). One hundred and sixteen indels were identified; the mtDNA loci *atp*A and *nad* 7 exhibited one and 21 indels, respectively and the *rpl*36-*rps*8, *trn*T2-*rps*4, and *trn*L-*trn*F loci exhibited 18, 35, and 41 indels, respectively.

Molecular Identification System Construction

The first stage in the molecular marker design process involved designing higher-level taxonomic markers from informative indels and SNPs that would divide the samples into the two major grass clades (BEP and PACCAD clades). Next, markers were designed to enable identification for the nine subfamily divisions. Due to the large number of species in the sample design, lower-level taxonomic markers (tribe, genus, and species) were only designed for the three largest subfamilies (86 samples in total); the Pooideae, Panicoideae, and Chloridoideae. Within these three subfamilies, markers were designed to identify all 86 samples to the genus level and the majority to species level. Fourteen species within genera did not possess variation at the species level for the six loci sequenced and thus were not able to be distinguished below the genus level.

The molecular identification system was constructed so that the markers divided the samples into two groups at each step, depending on the presence or absence of an indel or SNP. Figure 1 demonstrates how the molecular identification system was constructed for the Poeae to discriminate samples at the genus and species levels. A full set of molecular identification system pathways for the 100 grass taxa at each taxonomic level are presented in Appendix 3. Figure 2 provides examples of how the taxon-specific molecular markers were visualized and scored on agarose gels.

Marker Statistics

The molecular identification system was constructed from 91 marker sets; 17 were designed from informative indels from the *trn*T2-*rps*4, *trn*L-*trn*F, *atp*A, and *nad* 7 loci, and 68 were designed from SNPs from all six loci. Of the 91 marker sets, two were used for clade identification, nine for subfamily identification, 11 for tribe identification, 22 for genus identification, and 47 for species identification. Of the three subfamilies that had lower-level taxonomic markers designed, the number of representative samples in each subfamily reflected the number of PCR assays required to identify each sample. The Panicoideae was comprised of 43 marker sets, the Pooideae subfamily 27 marker sets, and the Chloridoideae subfamily 21 marker sets. To identify an unknown sample to subfamily level required as few as four markers, and a minimum of nine and maximum of 12 markers were required for species identification.

PCR Assay Outcomes

The large indel PCR assay consistently produced bands that were readily resolved and easily scored on agarose gels. For example, samples from the BEP clade were easily distinguished from the PACCAD clade by an 80-bp deletion (Fig. 2). The smaller indel and SNP PCR assays, with their internal positive control band for PCR failure, also generated fragment size differences among taxa that were readily resolved by agarose gel electrophoresis (Fig. 2). The generic AS-PCR method for SNP detection proved to be difficult to optimize for some marker sets and it was not possible to achieve allele-specific assays in all cases. Therefore, it was necessary to individually optimize PCR conditions for some assays to ensure accurate SNP typing.

Outcomes of the Molecular Identification System Tests

All 24 grass species represented in the molecular identification system were correctly identified to all possible taxonomic levels.

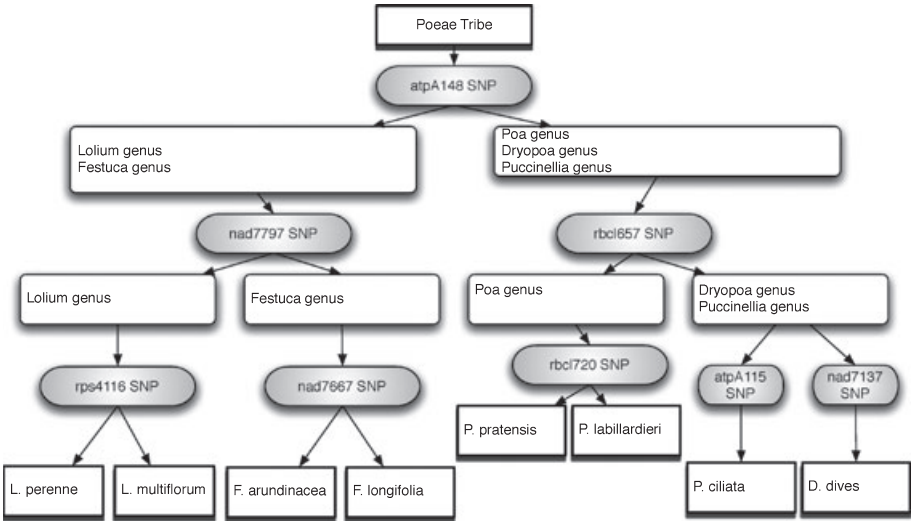

FIG. 1—*The molecular identification system pathway for the identification of grass samples from the Poeae into genera and subsequent species groups.*

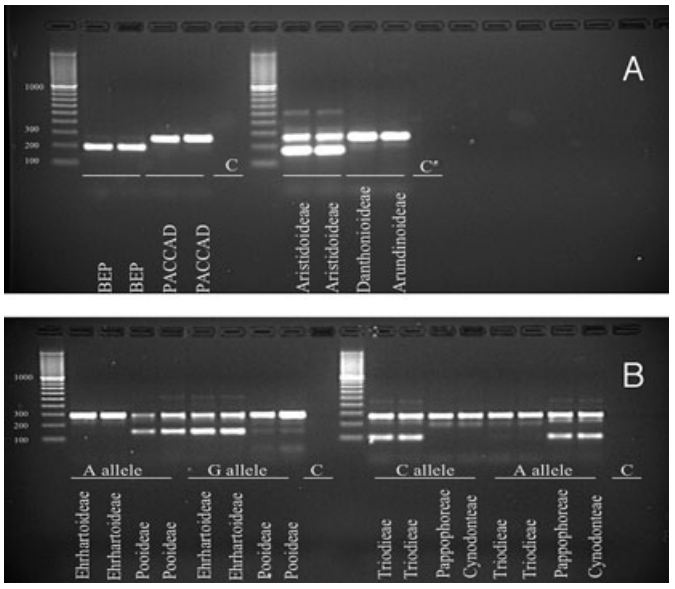

FIG. 2—*(a) The first assay demonstrates the utility of the large indel BEP/PACCAD clade marker. The second assay demonstrates the utility of a smaller indel marker that distinguishes between Aristidoideae and samples from other subfamilies. (b) The first assay demonstrates the utility of an SNP marker (A/G) that distinguishes between Pooideae and Ehrhartoideae samples and the second assay demonstrates the utility of an SNP marker (C/A) that distinguishes between Triodieae and samples from other tribes. C indicates negative control lanes, and a 100-bp ladder marker was used as a size standard and loaded in the first lane of each of the assays.*

All 22 nonrepresented species were identified correctly to the tribe level and 13 of these were correctly identified to genus. For the forensic-type samples, all samples represented in the molecular identification system design were correctly identified. This included all the samples extracted from dried plant material, grass seeds, and three out of four herbarium samples. All five duplicate samples were correctly identified at all possible taxonomic levels. All 24 roadside samples were identified correctly to the subfamily or tribe level, depending on marker availability. Sixteen samples were successfully identified to genus level and eight to species level.

The samples that were not correctly identified to genus or species level fell into one of two categories: (i) samples that did have genus/species-specific molecular markers in the molecular identification system; however, an incorrect identification indicated that the assumed taxon-specific markers were in fact not universally diagnostic across the genus/species when tested with a broader range of samples but rather only for the system representatives; (ii) samples that had no genus/species-specific markers in the molecular identification system and consequently were identified as a closely related genus represented in the system.

Discussion

The Molecular Identification System

Our molecular identification system was designed such that the PCR assays targeted specific single locus and phylogenetically informative DNA sequences surrounding variable sites and without the need for DNA sequencing of the test samples. Consideration of the reliability of the PCR assay methodology employed in the molecular identification system was particularly important given the forensic context of this study. It was essential that the scoring process was accurate, repeatable, and unambiguous and ultimately transferable across laboratories. In addition, the molecular markers were designed to be <500 bp in size to enable amplification of trace and degraded evidence material.

Two kinds of molecular markers underpinned this system: indels and SNPs. Informative indels were readily converted to PCR-based molecular markers and with our three primer assays, always provided a control band as a check for PCR failure. The PCR assays (for both large and smaller indels) required little or no marker-specific optimization and always produced unambiguous results with fragment size differences that were easily resolved with agarose electrophoresis. There are a number of SNP-typing methods available (see reviews [46–51]), which vary in their suitability for application in forensics. For the purpose of this study, we chose to apply the low-cost agarose gel-based AS-PCR method (43). However, for routine analyses, we would adopt a SNP-genotyping method that offers multiplex capability, does not require individual assay optimization, and can be run on an automated genotyping instrument.

All samples represented in the design of the molecular identification system were correctly identified at all taxonomic levels for which molecular markers were available. This confirmed that the molecular identification system was an accurate identification method and reiterated the findings from our pilot study on 20 grasses, where 100% of all samples represented in the design of the system were accurately identified (32).

The phylogenetically representative design of the molecular identification system also assisted with the identification of unknown samples not included in the original design. With the exception of several anomalous samples, test samples were correctly identified at the clade, subfamily, and tribe levels given the availability of molecular markers for those taxa. Therefore, the molecular identification system was capable of identifying unrepresented samples at higher taxonomic levels.

Misidentification, defined here as the incorrect identification of a sample at a specific taxonomic level, is an important consideration for the design of any molecular identification system. In this study, when an unknown sample lacked a representative at the genus or species level in the original design, either misidentification occurred or PCR failure or PCR fragments of inappropriate size indicated the level at which it was not possible to provide an accurate identification. Misidentification usually resulted in the identification of a sample as a closely related genus or species. In general, with care in scoring and the inclusion of positive controls for all assays, misidentification should be minimized or detected by PCR failure or unexpected fragment size.

In common with all identification systems, including DNA barcoding, the only solution to eliminate misidentification is the requirement for all target species to be represented in the system design. For the Poaceae, this would involve the unrealistic inclusion of *c*. 10,000 species. Geographic- or purpose-specific identification systems rather than whole family identification systems would minimize risks of misidentification due to a lack of representative species in the original system design. For example, a system could be readily developed for common native and exotic grasses in the ACT. This would greatly reduce the number of samples that need to be represented, enhancing the accuracy of the system as an identification method for unknown evidence samples.

The Forensic Utility of the Molecular Identification System

Our tests incorporated forensic-type samples to determine if identification of degraded, trace, or suboptimal botanical material could be achieved. Our findings demonstrated that the molecular markers employed (<500 bp in size) were suitable for amplifying

forensic-type samples from different sources of plant material. The identification of samples collected from an area of roadside in the ACT demonstrated how the system might be utilized in forensic casework. For example, if a combination of different species of grass seeds or fragments were found on a suspect's clothing, each separate grass fragment could be identified to determine the grass species present at the alleged crime scene. All roadside samples in this study were identified to at least tribe level and eight species identifications were possible. Therefore, this system could assist with the direction of an investigation by eliminating possible crime scenes based on the species composition of grasses at a specific geographic location.

Future Directions

Potential improvements to this molecular identification system include:

- Development of multiplex PCR assays in combination with high-throughput capillary electrophoresis methods to maximize the number of informative markers per reaction.
- Adoption of high-throughput methods for the detection and genotyping of SNP variation to enhance rapid screening capabilities.
- Design of geographic- or purpose-specific identification systems to target relevant species of interest.

We recommend that future development of molecular identification systems for botanical evidence should consider a two-tiered approach with the first tier employing a hierarchical molecular key to assist higher taxonomic-level identification (such as the approach taken in this study) and the second tier employing a DNA barcoding sequencing approach for identification at the lower taxonomic levels. Importantly, our extensive sequencing analysis has indicated that no single DNA sequence region will offer universal utility for identification across the diverse grass family.

A similar multigene tiered approach has been suggested by Newmaster et al. (28) who proposed a first tier of a core-coding gene such as *rbc*L and a second tier of multiple highly variable coding or noncoding genes, one of which would be implemented depending on the outcome of the first tier. A tiered approach would still require the compilation of a comprehensive sequence database of target grasses at selected loci. Thus the outcome of a higher-level taxonomic identification system could be used to assist in the selection of the most appropriate DNA region(s) to target for sequencing, given knowledge of the tribe or genus of the samples in hand, while reducing the double-sequencing requirement of the system proposed by Newmaster et al. (28). Tsai et al. (52) have demonstrated the feasibility of compiling a specific DNA sequence database of chloroplast *trn*L intron and *trn*L-*trn*F intergenic spacer sequences to assist in the identification of popular plants in Taiwan. As a starting point for developing such a database, the variable loci suggested in the plant DNA-barcoding literature could be adopted such as the chloroplast *trn*H-*psb*A or *trn*D-*trn*T intergenic spacer (26,29). Beyond these chloroplast loci, access to large segments of chloroplast noncoding DNA sequences is now available with new sets of universal primers such as those of Ebert and Peakall (53). Ebert and Peakall (54) also provided a comprehensive review of the technical resources available for accessing genetic variation in plant chloroplast genomes.

Conclusion

We have developed and tested a molecular identification system for grasses based on indel and SNP variation at multiple taxonomic levels using a multilocus approach. Our study has confirmed "proof of concept" of a new forensic technique that has the potential to provide assistance in the analysis of grass evidence. When used in conjunction with a forensic botanist, who might be able to make a preliminary morphological identification of the evidence, a DNA-based molecular identification system could confirm or extend identification of trace botanical evidence.

Acknowledgments

We thank Dan Ebert for his technical assistance in the lab, Julian Ash for assisting with the collection and identification of grass samples, and David Mallinson from the Australian National Herbarium for identifying grass samples. We also acknowledge the seed, grain, turf, and bamboo suppliers/nurseries (listed in Appendices 1 and 4); the Australian Tropical Crops and Forages Collection; the Australian Winter Cereals Collection; the Australian National Botanic Gardens; NSW Agriculture; the Australian Medicago Genetic Resource Centre; CSIRO Plant Industry; and the Australian National Herbarium for providing grass seed samples, grass material, or grass DNA.

References

1. Robertson J. Botanical and soil evidence at the crime scene. In: Horswell J, editor. The practice of crime scene investigation. Boca Raton, FL: CRC Press, 2004;317–45.
2. Baechtel FS, Comey CT. Applications of molecular biology to analyses of forensic evidence. Biologicals 1996;24:201–5.
3. Benecke M. DNA typing in forensic medicine and in criminal investigations: a current survey. Naturwissenschaften 1997;84:181–8.
4. Jobling MA, Gill P. Encoded evidence: DNA in forensic analysis. Nat Rev Genet 2004;5:739–51.
5. Karp A, Edwards KJ, Bruford M, Funk S, Vosman B, Morgante M, et al. Molecular technologies for biodiversity evaluation: opportunities and challenges. Nat Biotechnol 1997;15:625–8.
6. Morell MK, Peakall R, Appels R, Preston LR, Lloyd HL. DNA profiling techniques for plant variety identification. Aust J Exp Agric 1995;35:807–19.
7. Peakall R. PCR-based genetic markers and their applications to turfgrass breeding. Int Turfgrass Soc Res J 1997;8:243–59.
8. Caetano-Anolles G. DNA analysis of turfgrass genetic diversity. Crop Sci 1998;38(6):1415–24.
9. Korzun V. Use of molecular markers in cereal breeding. Cell Mol Biol Lett 2002;7:811–20.
10. Carita EJ. Classical and future DNA typing technologies for plants. In: Miller Coyle H, editor. Forensic botany: principles and applications to criminal casework. Boca Raton, FL: CRC Press, 2005;253–74.
11. Gostimsky SA, Kokaeva ZG, Konovalov FA. Studying plant genome variation using molecular markers. Russ J Genet 2005;41(4):378–88.
12. Hebert PDN, Cywinska A, Ball SL, DeWaard JR. Biological identifications through DNA barcodes. Proc R Soc Lond B Biol Sci 2003;270:313–22.
13. Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. Identification of birds through DNA barcodes. PLoS Biol 2004;2(10):1657–63.
14. Vences M, Thomas M, Bonett RM, Vieites DR. Deciphering amphibian diversity through DNA barcoding: chances and challenges. Philos Trans R Soc Lond B Biol Sci 2005;360:1859–68.
15. Barrett R, Hebert PDN. Identifying spiders through DNA barcodes. Can J Zool 2005;83:481–91.
16. Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN. DNA barcoding Australia's fish species. Philos Trans R Soc Lond B Biol Sci 2005;360:1847–57.
17. Smith MA, Fisher BL, Hebert PDN. DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. Philos Trans R Soc Lond B Biol Sci 2005;360:1825–34.
18. Nelson LA, Wallman JF, Dowton M. Using COI barcodes to identify forensically and medically important blowflies. Med Vet Entomol 2007;21:44–52.
19. Janzen DH, Hajibabaei M, Burns JM, Hallwachs W, Remigio E, Hebert PDN. Wedding biodiversity inventory of a large and complex

Lepidoptera fauna with DNA barcoding. Philos Trans R Soc Lond B Biol Sci 2005;360:1835–46.
20. Monaghan MT, Balke M, Gregory TR, Vogler AP. DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers. Philos Trans R Soc Lond B Biol Sci 2005;360:1925–34.
21. Summerbell RC, Levesque CA, Seifert KA, Bovers M, Fell JW, Diaz MR, et al. Microcoding: the second step in DNA barcoding. Philos Trans R Soc Lond B Biol Sci 2005;360:1897–903.
22. Lorenz JG, Jackson WE, Beck JC, Hanner R. The problems and promise of DNA barcodes for species diagnosis of primate biomaterials. Philos Trans R Soc Lond B Biol Sci 2005;360:1869–77.
23. Wolfe KH, Li W-H, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 1987;84:9054–8.
24. Saunders GW. Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. Philos Trans R Soc Lond B Biol Sci 2005;360:1879–88.
25. Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP, Haider N, et al. Land plants and DNA barcodes: short-term and long-term goals. Philos Trans R Soc Lond B Biol Sci 2005;360:1889–95.
26. Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci USA 2005;102(23):8369–74.
27. Cowan RS, Chase MW, Kress WJ, Savolainen V. 300,000 species to identify: problems, progress, and prospects in DNA barcoding of land plants. Taxon 2006;55(3):611–6.
28. Newmaster SG, Fazekas AJ, Ragupathy S. DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach. Can J Bot 2006;84(3):335–41.
29. Presting GG. Identification of conserved regions in the plastid genome: implications for DNA barcoding and biological function. Can J Bot 2006;84(9):1434–43.
30. De Ley P, De Ley IT, Morris K, Abebe E, Mundo-Ocampo M, Yoder M, et al. An integrated approach to fast and informative morphological vouchering of nematodes for applications in molecular barcoding. Philos Trans R Soc Lond B Biol Sci 2005;360:1945–58.
31. Ward J. A molecular forensic identification system for trace botanical evidence [Honours thesis]. Canberra, ACT: The Australian National University, 2002.
32. Ward J, Peakall R, Gilmore SR, Robertson J. A molecular identification system for grasses: a novel technology for forensic botany. Forensic Sci Int 2005;152:121–31.
33. Watson L, Dallwitz MJ. The grass genera of the world. Wallingford, U.K.: C.A.B International, 1992.
34. Renvoize SA, Clayton WD. Classification and evolution of the grasses. In: Chapman GP, editor. Grass evolution and domestication. Cambridge, U.K: Cambridge University Press, 1992;3–37.
35. Grass Phylogeny Working Group. Phylogeny and subfamilial classification of the grasses (poaceae). Ann Mo Bot Gard 2001;88:373–457.
36. Kellogg EA. Classification of the grass family. Flora Aust 2002;43:19–36.
37. Ward J. A grass molecular identification system for forensic botany [PhD thesis]. Canberra, ACT: The Australian National University, 2006.
38. Saltonstall K. A set of primers for amplification of noncoding regions of chloroplast DNA in grasses. Mol Ecol Notes 2001;1:76–8.
39. Taberlet P, Gielly L, Pautou G, Bouvet J. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 1991;17:1105–9.
40. Eyre-Walker A, Gaut BS. Correlated rates of synonymous site evolution across plant genomes. Mol Biol Evol 1997;14(3):455–60.
41. Dumolin-Lapegue S, Pemonge MH, Petit RJ. An enlarged set of consensus primers for the study of organelle DNA in plants. Mol Ecol 1997;6:393–7.
42. Rozen S, Skaletsky HJ. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, editors. Bioinformatics methods and protocols: methods in molecular biology. Totowa, NJ: Humana Press, 2000;365–86.
43. Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, et al. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res 1989;17:2503–16.
44. Okimoto R, Dodgson JB. Improved PCR amplification of multiple specific alleles (PAMSA) using internally mismatched primers. BioTechniques 1996;21(1):20–6.
45. Simsek M, Adnan H. Effect of single mismatches at 3'-end of primers on polymerase chain reaction. Med Sci 2000;2:11–4.
46. Gupta PK, Roy JK, Prasad M. Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Curr Sci 2001;80(4):524–35.
47. Gut IG. Automation in genotyping of single nucleotide polymorphisms. Hum Mutat 2001;17:475–92.
48. Kwok P-Y. Methods for genotyping single nucleotide polymorphisms. Annu Rev Genomics Hum Genet 2001;2:235–58.
49. Syvanen AC. Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat Rev Genet 2001;2:930–42.
50. Budowle B, Planz JV, Campbell RS, Eisenberg AJ. Single nucleotide polymorphisms and microarray technology in forensic genetics: development and application to mitochondrial DNA. Forensic Sci Rev 2004;16:21–36.
51. Sobrino B, Brion M, Carracedo A. SNPs in forensic genetics: a review on SNP typing methods. Forensic Sci Int 2005;154:181–94.
52. Tsai L-C, Yu Y-C, Hsieh H-M, Wang J-C, Linacre A, Lee JC-I. Species identification using sequences of the trnL intron and the trnL-trnF IGS of chloroplast genome among popular plants in Taiwan. Forensic Sci Int 2006;164:193–200.
53. Ebert D, Peakall R. A new set of universal de novo sequencing primers for extensive coverage of non-coding chloroplast DNA: new opportunities for phylogenetic studies and cpSSR discovery. Mol Ecol Resour 2009;9:777–83.
54. Ebert D, Peakall R. Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. Mol Ecol Resour 2009;9:673–90.

Additional information and reprint requests:
Jodie Ward, Ph.D.
Centre for Forensic Science
Canberra Institute of Technology
GPO Box 826
Canberra, ACT 2601
Australia
E-mail: jodiem.ward@cit.act.edu.au