FORENSIC MOLECULAR BOTANY: IDENTIFICATION OF PLANTS FROM TRACE EVIDENCE

Robert Bever, Ph.D. and Matthew Cimino B. Sc.
The Bode Technology Group, Inc., Springfield, VA

INTRODUCTION

Molecular systematic botany is expanding its role in forensic science. DNA based techniques have been used in criminal cases to place a suspect at the crime scene, as well as to identify strains of marijuana (1,2). These studies have used DNA obtained from individual plant specimens of a known species and have not had to overcome problems associated with DNA mixtures of unkown plant composition. Much botanical evidence occurs in the form of mixtures and consequently, is currently not used in investigations. The ability to separate and identify individual elements in botanical mixtures may provide additional information from evidence that was previously ignored or not considered valuable to case work.

The database genbank at the National Institutes of Health contains numerous plant DNA sequences that can be used to identify botanical trace evidence and include both nuclear and chloroplast genes. The most extensively characterized locus across all plant species is the large subunit of ribulose 1,5-bisphosphate carboxylase (*rbcL*), encoded in the chloroplast genome. Due to its cellular abundance and the extensive collection of DNA sequences contained in the NIH database, the *rbcL* locus was selected as the model system to develop methods for molecular analysis of trace evidence. One locus to complement *rbcL* for the investigation of trace evidence includes the ITS (internal transcribed spacer) region. The ITS region of the nuclear genome flanks the 5.8S ribosomal DNA coding region, is multi-copy, and has more sequence variability as compared to the *rbcL* gene.

Using the basic local alignment search tool (BLAST; 3), an evidentiary sequence can be compared to thousands of accessioned sequences from previously identified plants. BLAST is a web-based program that yields a list of organisms contained in the database with a sequence similar to the sequence of interest; the retrieved data starts with sequences having the greatest similarity. The matches can either be 100% or some lower percentage. Depending on the sequence of interest and the plant group that it belongs to, a high percentage match (near 100%) suggests a possible identification to the species level. However, if the match percentage is less than 100%, phylogenetic analysis is required to better characterize the sequence.

Phylogenetic methods are based on models of DNA sequence evolution and are used to determine relationships among different organisms. These models are used in analysis programs such as PAUP (4). Because the NCBI database contains a sample of representative plant species, phylogenetic methods are useful for determining or confirming the genus or family to which a sequence belongs. By constructing a nucleotide alignment that includes the evidentiary sequence and closely related sequences (of known identity) the evidentiary sequence can be classified within a genus and / or family. The most information is obtained when a sequence is identified to the species level, but useful data is also gleaned from genus or family level identification.

Sequences recovered from trace evidence may provide insight to the geographic origin of the evidence. Depending on the sample composition, a range of one to many different sequences can be recovered. Based on the ecology and habitat of the plants identified with these sequences, the geographic location can be inferred. The level of precision to which a sample's origin can be identified is often most influenced by the rarity and combination of the plants recovered. This combination of plants provides a 'description' for a sample that can match or exclude reference samples from a crime scene or other evidentiary materials.

To begin the validation process, we report two investigations that were conducted on botanical mixtures. The first mixture was made by combining purified DNA from eight previously identified plants and the second mixture was dust obtained from a piece of mock evidentiary clothing of unknown origin. We

provide a review of the methods and results of these investigations, as well as a discussion of its significance.

METHODS

Eight plant mixture

Plant tissue was collected from an area outside of the laboratory or obtained from dried herbarium specimens. The plants were identified using Gleason and Cronquist (5) and consisted of *Acer*, *Cynodon*, *Lonicera*, *Phytolacca*, *Polygonum*, *Pyrus*, *Robinia*, and *Solanum*. DNA was isolated and purified from each plant sample using the DNeasy extraction kit (Qiagen, Inc., Valencia, CA). The DNA was quantified using agarose gel visualization with commercially available DNA standards (Promega Corporation, Madison, WI). Equal concentrations of DNA were added to form a mixture. The DNA mixture was used as the genomic target with universal primers designed to amplify the gene *rbcL*, a plant specific molecular marker. The resulting PCR product was prepared for cloning using polyethylene glycol (PEG) precipitation (6). The PCR product was cloned using the Original TA Cloning Kit (Invitrogen Corp., Carlsbad, CA). Clones were screened by restriction digest using Hae III (Promega Corporation, Madison, WI). Clones of interest were sequenced with M13 primers following the standard dye-terminator protocol (Applied Biosystems, Foster City, CA). BLAST was used to identify the botanical affinity of the sequences. The PCR, cloning, and sequence identification process was repeated three times.

Dust mixture

An article of clothing (a sock) was vacuumed and particles associated with the clothing were collected on a filter. A dust sample weighing less than 0.1grams was obtained. DNA was purified from the sample with the above protocol and used in PCR. Two different reactions were used to amplify the molecular markers *rbcL* and ITS. The resulting PCR products were cloned and sequenced using the above protocols. The botanical affinity of the sequences was determined using BLAST and phylogenetic analyses with other sequences available in genbank. Analyses were conducted with PAUP software (5).

RESULTS

Eight plant mixture

In each of the three trials, all eight plants were recovered from the mixture as identified by DNA sequence analysis. The number of sequences obtained for each plant varied from among trials. Approximately 100 colonies were needed to obtain at least one sequence from every plant in the mixture. A number of clones were sequenced that contained partial sequences and could not be used in analysis. No transformants consisting of rearranged DNA were detected. Table 1 provides a summary of the three trials.

Dust mixture

ITS and *rbcL* primers successfully recovered sequences from the evidentiary dust. However, the primers for each locus did not amplify sequences for the same plants. The chloroplast gene, *rbcL*, recovered sequences from three green algae and one angiosperm, while the nuclear gene, ITS, recovered two green algae, nine angiosperms, six fungi, and numerous sequences not similar to other ITS sequences in the NCBI database.

All the *rbcL* and ITS sequences could not be identified to species. Of the nine angiosperm ITS sequences, two were identified to species, five to genus, and one was limited to family level identification. Four angiosperm families were represented and include Asteraceae, Brassicaceae, Fabaceae, and Poaceae. The *rbcL* angiosperm sequence could only be identified to the family Poaceae. Green algae sequences recovered by *rbcL* and ITS could be placed at the ordinal (Volvocales) or class (Trebouxiophyceae) level. ITS also recovered a fungal sequence that is closely related to *Fusarium* and *Gibberella*, two genera known to be crop pathogens. Figure 2 is a phylogram showing the phylogenetic

placement of an evidentiary sequence in the genus *Artemisia*. Table 2 is a list of taxa identified from the dust sample.

DISCUSSION

The work presented here suggests that botanical mixtures can be separated and that individual components can be identified in a consistent and repeatable manner. The *rbcL* locus was amplified across all the elements in the eight plant mixture and enabled each of the plants to be identified as part of the sample. Interestingly, the plants were not recovered in equal proportions across each trial. Reasons for the imbalance in recovery are unclear at this time, but may be due poor annealing of the universal primers, preferential amplification of template, or some other aspect of PCR dynamics that affect product formation. We are currently performing this experiment with ITS and hope to gain better insight to this problem with additional data.

The DNA sequences generated from the mock trace evidence identified several plants (and fungi) that have a broad global distribution. However, inferences can still be made as to the environment and ecology common to the organisms recovered. The plants identified are found in temperate climates and are not likely to be common in dry, alpine, or marine environments. The recovery of *Triticum* and a sequence closely related to fungal crop pathogens suggests an area influenced by agriculture. Identification of *Barbarea* (winter cress), *Artemisia* (ragweed), *Medicago* (medick), and *Conyza* (fleabane) suggest an area that has been recently disturbed (5), as these plants are early colonizers (weeds) of areas that have rich, damp soil, with direct sunlight. The incidence of green algae in the dust sample is consistent with an environment that is wet or damp and rich in nutrients. Additionally, several *rbcL* sequences were identified as closely related to *Prasiola* (Chlorophyceae) and may share a similar habitat to this genus; *Prasiola* can be found in moving water, adhering to rocks (7). The occurrence of the *Prasiola*-like algae suggests the source of the dust may be near rocky outcroppings of a stream or river.

Botanical mixtures are difficult to resolve with taxonomic clarity due to the lack of macroscopic characters in particulate matter. DNA based methods in concert with genbank show a promising start for accurately identifying plants in trace evidence mixtures. As systematic botany advances, more information will be added to the database. This will serve to increase the general power of this approach and will further our ability to resolve mixtures.

The botanical components of a mixture can be used for many aspects of criminal investigations. Plants can help determine a sample's geographic origin in either exact or general terms. By determining nucleotide sequences, botanical elements can also identify potential matches or exclusions among pieces of evidence. Furthermore, molecular techniques support and compliment traditional forensic botany in the identification of evidence that may have been previously ignored. In this manner, the methods reported here represent significant advances in investigative technology.

ACKNOWLEDGEMENTS:

Technical work was performed by Lauren Brinkac, Emily Hopkins, Robyn Wingrove, and Nicole Gross of The Bode Technology Group, Inc. The herbarium specimens were provided by the Norton–Brown Herbarium, University of Maryland- College Park, College Park, Maryland 20742. The authors would like to thank Dr. Ed Golenberg of Wayne State University and Dr. Charles Delwiche of the University of Maryland for their scientific discussions and technical recommendations.

REFERENCES:

1. Yoon C. Botanical witness for the prosecution. Science 1993; 260:894-895.

2. Coyle HM, Ladd C, Lee HC, Palmbach T. The green revolution: Botantical contributions to forensics and drug enforcement. Croation Medical Journal 2001; 42:340-345.

3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990; 215:403-10.

4. Swofford DL. PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4. Massachusetts: Sinauer Associates, Sunderland, 2001.

5. Gleason HA, Cronquist A. Manual of vascular plants of Northeastern United States and adjacent Canada. New York: The New York Botantical Garden, 1991.

6. Morgan DR, Soltis DE. Phylogenetic relationships among members of Saxifragaceae *sensu lato* based on *rbcL* sequence data. Ann Missouri Bot Gard 1993; 80:631-60.

7. Smith G. The fresh-water algae of the United States. London: McGraw-Hill, 1933.

RESULTS OF MIXTURE TRIALS

Species recovered VS Colonies Screened

Plant	Trial 1	Trial 2	Trial 3	Total	Percent
Cynodon	35	20	9	65	22%
Phytolacca	43	29	33	105	36%
Acer	1	2	2	5	2%
Pyrus	6	7	5	18	6%
Polygonum	1	5	6	12	4%
Robinia	8	4	3	15	5%
Solanum	7	7	10	24	8%
Lonicera	3	2	9	14	5%
Partial Inserts	6	6	25	37	12%
TOTAL	110	82	101	293	100%

Strategy for Sample Identification