



DNA barcoding of insect and plant fragments: an alternative approach to assist with soil geoattribution

Kelly A. Meiklejohn¹, Megan L. Jackson², Jack Hietpas¹, Libby A. Stern³ and James M. Robertson³

¹Counterterrorism and Forensic Science Research Unit, Visiting Scientist Program, FBI Laboratory Division; ²Virginia Commonwealth University, Forensic Science Masters Program; and ³Counterterrorism and Forensic Science Research Unit, FBI Laboratory Division

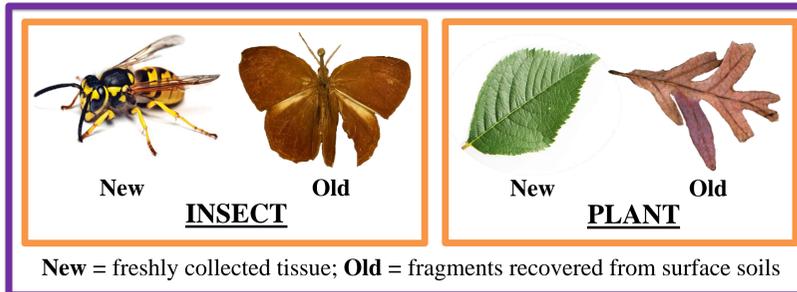


Introduction

- Soil and geological evidence can provide clues in forensic investigations
 - primary focus on constraining the circumstances of the crime
- Biological fragments are commonly recovered in soil evidence and could assist with geoattribution, but are rarely analyzed:
 - morphological ID is difficult and requires specialized expertise
 - DNA might provide a reliable and accurate method for ID
- DNA barcoding is an accepted molecular ID approach, targeting:
 - Insects:** 1 mitochondrial gene
 - COI* – species level ID
 - Plants:** 2 chloroplast genes
 - rbcL* – family/genus level ID
 - matK* – species level ID

Materials and Methods

- Developed and tested the protocol using two types of samples:



- Optimized PCR conditions for the barcode primer pairs listed in Table 1:

- facilitates amplification even for highly degraded samples (mini barcode primers)
- used nested PCR to increase *matK* amplification success, as there are no reliable mini primers

Table 1. Information on the targeted barcode regions and primer pairs used for amplification. Arrows represent the suggested order for amplifications (i.e. larger fragment followed by the shorter fragment).

Barcode region	Length (bp)	Primers	Target
<i>COI</i>	~650	LCO1490-L/HCO2198-L ¹	Universal
<i>COI</i> mini	~130	uniminibarF1/uniminibarR1 ²	Universal
<i>matK</i>	~850	matK-KIM-1R/matK-KIM-3F ³	Angiosperms
nested <i>matK</i>	~830	matK4La ⁴ /matKMALPR1 ⁵	Angiosperms
<i>rbcL</i>	~590	rbcLa-F ⁶ /rbcLa-R ⁷	Universal
<i>rbcL</i> mini	~230	rbcL1/rbcLB ⁸	Universal

References:
¹ Nelson LA, Wallman JF, Dowton M. (2007). *Med Vet Entomol.* 21: 44-52.
² Meunier I, Singer GAC, Landry JF *et al.* (2008). *BMC Genomics.* 9: 214.
³ Ki-Joong Kim, personal communication.
⁴ Wojciechowski MF, Lavin M, Sanderson MJ. (2004). *Am J Bot.* 91: 1846-1862.
⁵ Royal Botanic Garden Edinburgh, Edinburgh U.K.
⁶ Levin RA, Wagner WL, Hoch PC *et al.* (2003). *Am J Bot.* 90: 107-115.
⁷ Kress WJ, Erickson DL, Jones FA *et al.* (2009). *PNAS.* 106: 18621-18626.
⁸ Little, DP. (2013). *Mol Ecol Resour.* 14: 437-446.
⁹ Kress WJ, Erickson DL. (2012). *Methods Mol Biol.* 858.

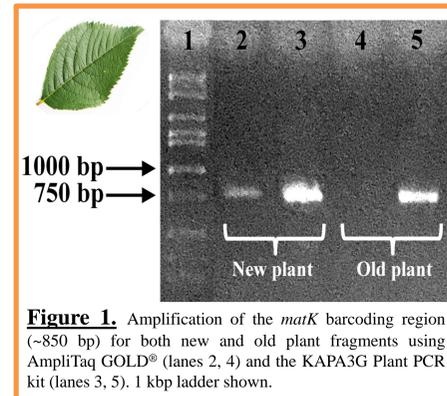
Acknowledgements.
 Visiting Scientist Program at the FBI Laboratory Division, administered by the Oak Ridge Institute of Science and Education, through an interagency agreement between the US Department of Energy and the FBI.

Disclaimer. Names of commercial manufacturers are provided for identification purposes only and inclusion does not imply endorsement by the FBI. The conclusions contained in this document are those of the authors and should not be implied as necessarily representing the official policies, either expressed or implied, of the U.S. Government.

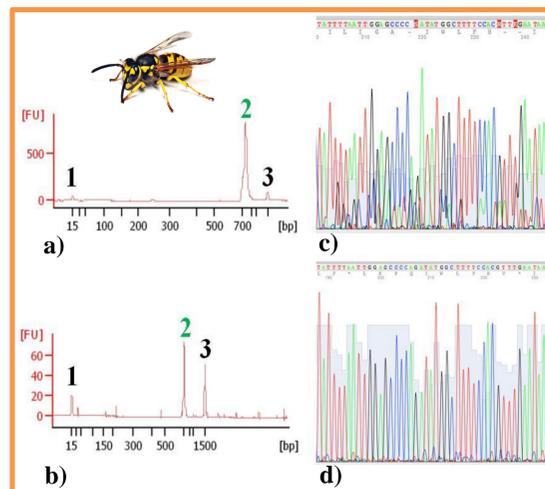
Results and Discussion

Issues encountered during protocol development

- All PCR reaction and cycling conditions produced amplicons **only** for the **new plant** (AmpliTaq GOLD®)
 - identified **old plant** amplification impacted by **inhibitors**. To overcome inhibition we trialed:
 - Post extraction clean up* (e.g. bead purification, precipitation)
 - PCR additives* (e.g. BSA, betaine, PVP)
 - Specialized Plant Enzyme* (KAPA3G Plant PCR kit)



- KAPA3G Plant PCR kit provided the strongest reproducible PCR amplicons (Figure 1).



- COI* entire sanger sequence data was **messy** for **old** insect amplicons (amplified using AmpliTaq GOLD®) (Figure 2a, c)

- optimization of sequencing reaction did not produce clean data. Trialed:
 - adding DMSO, changing annealing temperature, decreasing primer and dye concentration

- Amplification** with NEB Q5 Hot Start High-Fidelity DNA polymerase gave clean reproducible sequence data (Figure 2b, d)

DNA quantity/quality from biological fragments

- Developed protocol was tested on ~200 individual insect and plant fragments isolated from various surface soil samples

- Fragments were selected to represent a variety of types (Figure 3)

- DNA quantity was low and protein, RNA and aromatic/phenolic contaminants were detected in extracts (Table 2)

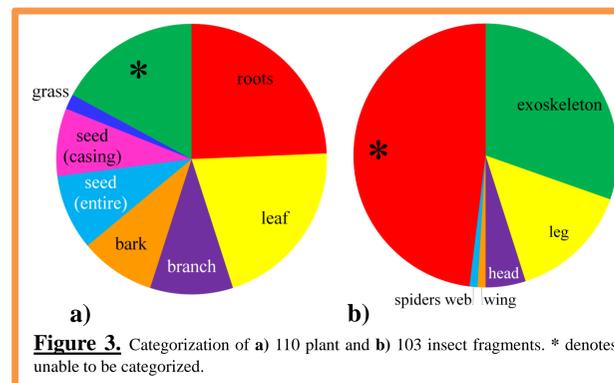


Table 2. Concentration and purity of DNA extracts as determined by Nanodrop ND-1000. Only data from extracts that had a concentration of >1.5 ng/µl are reported.

	ng/µl	260/280	260/230
Insects (n, 96)	9.0 ± 15	1.6 ± 0.32	0.43 ± 0.30
Plants (n, 62)	23 ± 74	0.37 ± 1.3	1.55 ± 0.61

Results and Discussion

Broad protocol success

- Fragments of the expected size were obtained for all primer pairs (Figure 4)
 - relatively free of secondary products, allowing for straight-forward sequencing
- PCR and sequencing success good (Table 3)
- All of the amplicons were from the **expected gene region** after database searching

Table 3. Success of the developed barcoding protocol on a range of samples (total n, 213).

Barcode Region	PCR success	Sequencing success
<i>COI</i> (entire and mini)	80%	63%
<i>matK</i>	71%	71%
<i>rbcL</i> (entire and mini)	92%	85%

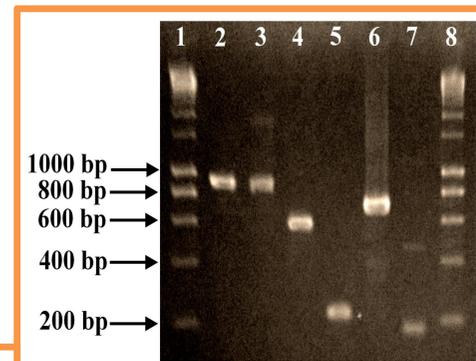


Figure 4. Plant and insect DNA barcoding region amplicons: (1) 1 kbp DNA Ladder (2) ~850 bp *matK* (primers matK-KIM-1R/matK-KIM-3F) (3) ~830 bp *matK* (primers matK4La/matKMALPR1) (4) ~590 bp *rbcL* (primers rbcLa-F/rbcLa-R) (5) ~230 bp *rbcL* (primers rbcL1/rbcLB) (6) ~650 bp *COI* (primers LCO1490-L/HCO2198-L) (7) ~130 bp *COI* (primers uniminibarF1/uniminibarR1) (8) 1 kbp DNA Ladder

Utility of current sequence databases

- Better success with taxonomic identification of plant sequences; insect data was poor (Table 4)

Table 4. Comparison of barcode sequence data to public databases.

	<i>COI</i>		<i>matK</i>		<i>rbcL</i>	
	BOLD	GenBank	BOLD	GenBank	BOLD	GenBank
No match in database*	24%	11%	0%	0%	5%	2%
Good matching statistics^	29%	20%	96%	96%	78%	78%
Database concordance	31%		98%		100%	

* Taxonomic resolution only of order level or higher; ^ percentage similarity of ≥ 90%

DNA barcoding protocol

