



G68 Spatial Autocorrelation Using Mantel Tests and Bioinformatics for the Classification and Provenance of Soil Samples for Intelligence and Forensic Applications

Natalie Damaso, 6151 W 22nd Lane, Hialeah, FL 33016; Maria A. Mendoza, PhD, 11200 SW 8th Street, Biological Sciences, OE 167, Miami, FL 33199; and DeEtta Mills, PhD, OE 167, Biological Sciences, 11200 SW 8th Street, Florida International University, Miami, FL 33199*

After attending this presentation, attendees will understand that the metagenomic information contained within soils is relevant to the geographic origin of the soil sample. In addition, using bioinformatic machine-learning algorithms, these data can provide soil databases that can be queried when an unknown sample is collected and can provide approximate geographic location and provenance of the source.

This presentation will impact the forensic science community by greatly expanding the forensic geology discipline and enhancing the probative value of soil samples for intelligence gathering and forensic cases. These data can supplement and enhance the investigative and probative value of soil samples.

Based on the ecological hypothesis that soil type (e.g., the chemical/physical properties) drives which microbes live in a particular soil, it should therefore be possible, in theory, to use soil metagenome profiling to produce a unique biotic pattern and associate it to the collection site.^{1,2} If this is indeed true, then a soil biotic profile from a particular geographical location should be able to be mathematically correlated to that location.³ Subsequently, if shown to be true, these soil microbial assemblages would be able to predict soil provenance and have forensic applications. Therefore, the basic questions become: (1) "Is there a correlation between the genetic profile of a soil's microbial community and the geographic location?"; and, (2) "If so, can these predicted patterns be used for soil provenance for intelligence data and/or have a forensic application?" The first question was addressed using a Mantel test for spatial autocorrelation which tests the assumption that geographically closer samples will be more similar than those further apart. Metagenomic data from four taxa — bacteria, archaea, fungi, and plant — from $\approx 1,900$ samples collected from primarily undisturbed soils in Miami-Dade County, Florida, was used. Spatial autocorrelations of genetic data to geographic locations within soil types, from multiple transects, and subplots within subplots were modeled. The analyses used the geographic location (Global Positioning System (GPS) coordinates) and metagenomic biotic profiles from the four taxa. At all scales and in almost all samples, the biotic profiles were positively correlated and significant ($p < 0.05$) to the source of the soil sample. For the two samples that were not significant yet still positively correlated, the historical satellite data were analyzed and it was noted that one site had been burned three months prior to sampling while the other was a disturbed site at an abandoned nursery. This indicated that when there was a deviation from the correlation, the soil was impacted or disturbed in some fashion. Question 2 is being answered using a suite of unsupervised and supervised bioinformatics algorithms that can better predict soil biotic associations to the soil type and specific location.³⁻⁶ This will then allow for a searchable database of soil biotic profiles that is linked to its provenance and can be used for either intelligence or forensic purposes.

References:

1. Girvan MS, Bullimore J, Pretty JN, Mark-Osborn A, Ball AS. Soil type is the primary determinant of the composition of the total and active bacterial communities in arable soils. *Appl Environ Microbiol*, 2003: 69:1800-1809.
2. Moreno LI, Mills DK, Entry J, Sautter RT, Mathee K. Microbial metagenome profiling using amplicon length heterogeneity polymerase chain reaction proves more effective than elemental analysis in discriminating soil specimens. *J Forensic Sci* 2006: 51: 1315-
3. Legendre, Pierre, and Marie-Josée Fortin. Comparison of the Mantel Test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. 2010: *Mol Ecol Res*:10:831-844.
4. Larsen, P.E., Field, D., & Gilbert, J.A. Predicting bacterial community assemblages using an artificial neural network approach. *Nature Methods*, 2012:9:621-625.
5. Metzger, KJ, Klaper, R, Thomas, MA. Implications of informatics approaches in ecological research. *Ecol Informatics*, 2011:6:4-12
6. Yang, C, Mills D., Mathee, K, Wang, Y, Jayachandran, K, Sikaroodi, M, Gillevet, P, Entry, J, Narasimhan, G. An ecoinformatics tool for microbial community studies: Supervised classification of amplicon length heterogeneity (ALH) profiles of 16S rRNA, *J Microbiol Methods*, 2006:65:49-62.



Pathology/Biology Section - 2014

Spatial Autocorrelation, Soil Forensics, Bioinformatics