Jones, Oliver A.H.; Sdepanian, Stephanie; Lofts, Steven; Svendsen, Claus; Spurgeon, David J.; Maguire, Mahon L.; Griffin, Julian L.. 2014. **Metabolomic analysis of soil communities can be used for pollution assessment.** *Environmental Toxicology and Chemistry*, 33 (1). 61-64. 10.1002/etc.2418

Environmental Toxicology and Chemistry

Short Communication

METABOLOMIC ANALYSIS OF SOIL COMMUNITIES CAN BE USED FOR POLLUTION ASSESSMENT

OLIVER A.H. JONES, STEPHANIE SDEPANIAN, STEVEN LOFTS, CLAUS SVENDSEN, DAVID J. SPURGEON, MAHON L. MAGUIRE, and JULIAN L. GRIFFIN

Environ Toxicol Chem., **Accepted Article** • DOI: 10.1002/etc.2418

Short Communication **Environmental Toxicology and Chemistry**
DOI 10.1002/etc.2418

METABOLOMIC ANALYSIS OF SOIL COMMUNITIES CAN BE USED FOR POLLUTION ASSESSMENT

OLIVER A.H. JONES,†* STEPHANIE SDEPANIAN,‡ STEVEN LOFTS,‡ CLAUS SVENDSEN,§ DAVID J. SPURGEON,§ MAHON L. MAGUIRE,|| and JULIAN L. GRIFFIN#††

† School of Applied Sciences, RMIT University, Melbourne, Australia.
‡ Centre for Ecology & Hydrology, Lancaster Environment Centre, Library Avenue, Bailrigg, Lancaster, United Kingdom
§ The Maclean Building, Centre for Ecology and Hydrology, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire, United Kingdom
|| BHF Magnetic Resonance Unit, Department of Cardiovascular Medicine, Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, United Kingdom
Department of Biochemistry and the Cambridge Systems Biology Centre, University of Cambridge, Tennis Court Road, Cambridge, United Kingdom
†† The Medical Research Council Human Nutrition Research, Elsie Widdowson Laboratory, Fulbourn Road, Cambridge, United Kingdom

Running title: Metabolomics of soil communities

* Address correspondence to Oliver.Jones@rmit.edu.au

Submitted 5 August 2013; Returned for Revisions 24 September 2013; Accepted 26 September 2013

Abstract: Here we show that metabolic profiling can be used to assess the changes in biochemical profiles of soil communities living in contaminated sites. We propose the term ‘community metabolomics’ for the application of metabolomics techniques to the study of the entire community of a soil sample. We anticipate our study to be a starting point for the use of this technique to assess how communities respond to factors such as pollution and climate change.

Keywords: Metabolomics, Microbes, Invertebrates, Pollution, NMR, Soil

INTRODUCTION

Soil based microbes and invertebrates form a vital component of all terrestrial ecosystems. Such communities are important in many steps in nutrient cycles, such as the fixation of nitrogen from the atmosphere and the degradation of decaying matter [1]. However, since many, soil-dwelling species cannot be cultured within a laboratory setting these communities are mostly unstudied, particularly in terms of genetics and biochemistry. Consequently there has been little insight into how the activities of soil communities as a whole relate to ecosystem functions [2, 3]. Here we have addressed this issue by using principles from the field of metagenomics.

Metabolomics attempts to capture the complexity of metabolic networks via the comprehensive characterization of the small molecule metabolites (such as amino acids, sugars and lipids) in biological systems and how they vary in response to a variety of stimuli [4]. It has a large practical advantage over other 'omic/systems biology based technologies such as transcriptomics and proteomics in that metabolites are similar in the majority of species; thus, a fully annotated genome is not required for analysis and analytical methods are transferable between species. Metagenomics refers to the application of modern genomic techniques to the study of communities of (mostly microbial) organisms directly in their natural environments [5]. It is a rapidly growing area of the genome sciences based on the genomic analysis of DNA extracted directly from entire communities in their native habitats [6]. It allows us to see, in ever increasing detail, the vast diversity that exists in the biosphere. For instance, over 1.2 million previously unknown genes were identified when the procedure was applied to the analysis of samples from the Sargasso sea [7]. The technique has emerged as a powerful tool that can be used to analyze microbial communities, regardless of the ability of individual component members to be cultured in the laboratory [8]. It has been used in studies on the community responses to UV-B radiation [9], phosphate removal in sewage sludge [10] and, more recently, marine pelagic and sediment environments [11, 12]. Fourier Transform Infrared Spectroscopy has also been shown to allow the chemically-based discrimination of microbial genotypes [13].

It is of note however, that no corresponding study of large-scale metabolic analysis has yet been undertaken. While recent systems biology based studies have shown promise [14] much remains to be learnt from

through what might be termed "Meta-Metabolomics" or, perhaps more sensibly, “Community Metabolomics”. We propose the latter term for studies in which the naturally occurring products of metabolism from the entire community of a given sample are analyzed simultaneously. This is in contrast to previous studies in metabolomics which, to date, have primarily focused on single sample types.

In this study we use a modified methanol-chloroform-water extraction coupled with ^{1}H Nuclear Magnetic Resonance Spectroscopy (NMR) and principle component analysis to assesses the metabolic profiles of communities living in soils from a range of former mine sites in the United Kingdom. Specifically we targeted aqueous phase metabolites, which are relatively quick and simple to extract and analyze and thus have the potential to form the basis for a fast, yet detailed assessment of contaminated field sites. To achieve this we developed protocols for the ^{1}H NMR based metabolic analysis of soils. These enabled detailed spectra to be obtained from 500mg of soil. Many metabolites of interest were evident, including amino acids, nucleotides and sugars. Once methods were developed we applied them the analysis of soil communities from eleven, previously established field sites and were able to show that each had a unique and distinct metabolic profile, despite variations in physiochemical characteristics such as soil type and contamination profile.

METHODS

Site description

We utilized 11 previously established mine sites in the United Kingdom. Each was associated with Pb/Zn mines that discontinued production in the period from approximately 1880 to 1920. Of these, four (Bog, Pennerley, Roman Gravels and Snail Beach) were located in Shropshire in Central West England on Arenig rock of Ordovician origin underlain by Sliperstone quartz. Soils at these sites were circumneutral to mildly acidic and scrub and heathland plants provided the dominant vegetation. Five sites (Castell, Wemyss, Cwmystwyth Cottage, Cwmystwyth West Cottage II and Cwmystwyth Stream) were located on three separate mine working within the River Rheidol Valley in Central Wales. These mines were located on Pb/Zn containing loads present in rocks originating from the late Ordovician and early-mid Silurian period. All sites were associated with nutrient-poor upland grassland on relatively base-poor, acidic soil. A further site within this region (Vertigo) was located on a grassland area, away from the main mineral ore worked area at the main Cwmystwyth site and provided a

relatively unpolluted comparator for the more heavily contaminated areas. The last site (Ecton) was located on the South-West fringe of the Peak District National Park in the North Midlands of England. This site is located on a hill of Carboniferous Limestone on near neutral soil, with oak wood as the dominant vegetation at the collection point.

Soil analysis

A 1kg sample of soil was taken on site at the coordinates given in table one and stored at -80^{o}C prior to analysis. Sub-samples were then taken for the physiochemical and metabolomic analysis. All pH readings were measured at 17^{o}C using a combined pH electrode and meterlab pH/ion meter (Radiometer Analytical, Lyon, France). The pH (H_2O) was calculated using a water extraction with the ratio of 10:1 deionized water to dry weight soil. Samples were sieved to 4mm and oven dried before extracting with water. Individual extractions were left to shake overnight, after which the supernatant was extracted and filtered to 0.2□m. The pH solution (soil) was measured using in soil solution samples extracted using Rhizon soil solution samplers (SDEC, Tauxigny, France).

To calculate the percentage loss on ignition a dried and ground soil sample was weighed into a crucible then placed in a furnace at 550^{o}C for 5 hours. Once removed from the furnace the sample was re-weighed and the %LOI calculated. For heavy metal analysis, soil solutions were acidified to a concentration of 1% nitric acid, using Analar grade nitric acid (VWR, Lutterworth UK) and analyzed for the metals outlined in Figure 3 using a Thermo X series inductively coupled plasma mass spectrometer (ICP-MS, Thermo Fisher Scientific, Hemel Hempstead, UK).

Metabolite Extraction

A 500mg subsample of soil from each site was sieved to with a sieved to 4mm then to ground to a fine powder under liquid N_2. Individual samples were then extracted using a standard methanol/chloroform/water method [15]. No further extraction steps, for example dispersion followed by centrifugation, were used. This means that the method did not exclude resident macro-invertebrates or the eggs/ juveniles of other soil-dwelling invertebrates and thus these organisms would have contributed to the metabolite profile from each site. If one wanted to focus entirely on the microbial community it would have been necessary to employ a method that

specifically extracted the soil microbial community from the soil, for example the procedure described by Mayr *et al*. [16].

^{1}H NMR analysis

Aqueous extracts were dried down in a Concentrator 5301 evacuated centrifuge (Eppendorf, Histon, UK). They were then dissolved in 500 µL of D_2O and buffered with the addition of 100 µL 0.24 M sodium phosphate (pH 7.0) containing 1 mM sodium 3 (trimethylsilyl) 2,2,3,3-tetradeuteriopropionate (TSP; Cambridge Isotope Laboratories, Andover, USA) which provided a chemical shift reference (0 ppm) for the resulting spectra. NMR analysis used a Bruker NMR spectrometer at 11.7 T (^{1}H frequency of 500.3 MHz) using a 5 mm ATMA TXI probe and an Avance II+ console (Bruker BioSpin, Rheinstetten, Germany). Spectra were acquired using a 1D NOESY pulse sequence with water presaturation.

Statistical analysis

The NMR data were Pareto scaled and analysed via Principal Components Analysis (PCA) using SIMCA-P software (version 11, Umetrics, Umea, Sweden). All models were further validated by resampling the model 99 times under the null hypothesis (meaning generating models with a randomly permuted Y matrix not related to the factors of interest) [17]. Models that failed validation (i.e. where no difference between the randomly generated and the real data was observed) were not analyzed further.

RESULTS AND DISCUSSION

The pH of the soils and associated pore waters varied from 4.26 to 7.35 and there was a wide variation in the organic carbon content (as determined by the loss on ignition method) which varied from as low as 12% to almost 80%. The moisture content was similar (around 40% in all cases) for all sites despite their wide geographic dispersal. Interestingly two sites (Ecton and Snail Beach) had very similar metabolic and pollutant profiles despite being almost a hundred miles apart and with very different physiochemical and geological conditions (see table 1 and figure 1). Specifically these two sites had lower metal levels overall and much lower levels of Fe than other sites.

Figure two shows a ^{1}H NMR spectra of the soil extract while figure three shows a Principle Component Analysis (PCA) of ^{1}H-NMR data from samples from all sites. The Ecton and Snail Beach sites can be seen to

cluster together, away from other sampling locations. Due to the large geographic and geological differences between the sites it is likely that their metabolic similarity is a response to their similar pollution profiles.

These results suggest that community metabolomics analysis not only enables a survey of the metabolites present in a specific environment, such as water or soil, to be carried out directly but that this information can be used to develop biomarkers indicative of a defined response. In this case anthropogenic pollution but other factors such as land use and climate change could also be assessed. This is similar to previous observations using earthworms which showed that metabolic profile biomarkers of metal contamination were applicable across multiple sites [18]. If changes in microbial community structure could be detected before major outward changes became apparent it would be very useful in preventing damage to a variety of sensitive systems (e.g. eutrophication).

There are also potential industrial uses of community metabolomics to elucidate metabolic pathways. For example, the anaerobic reactors used for full-scale wastewater treatment and biogas production are reliant on complex, multi-species communities but the metabolic roles of the individual members are still largely undetermined. Moreover, the species are strongly connected through syntrophies where the waste products of one provide resources for another. Consequently, to understand, exploit and extend the application of these systems the 'ecophysiological' roles must be determined. Questions, such as how the syntrophic interactions structure the system-level behavior, could then be addressed. The unification of metabolomics and other 'omic approaches – particularly metagenomics – could be used to provide a high-throughput solution to link taxonomy with function. This approach could also enable the construction and validation of 'ecosystems biology' models operating at the level of the whole community. This could potentially have applications in testing of contaminated land prior to redevelopment, for example house building on former industrial sites. Similar research could also identify soil organisms with favorable metabolic activity for bioremediation activities. Community metabolomics therefore has potential applications in a range of areas and its study has the potential to be of great benefit.

Acknowledgment—OAHJ thanks Aaron Raynor and Trevor Rook for useful conversations.

REFERENCES

1. Handelsman J. 2004. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* 68:669-685.

2. Feng X, Simpson AJ, Wilson KP, Dudley Williams D, Simpson MJ. 2008. Increased cuticular carbon sequestration and lignin oxidation in response to soil warming. *Nature Geosci* 1:836-839.

3. Walker JJ, Spear JR, Pace NR. 2005. Geobiology of a microbial endolithic community in the Yellowstone geothermal environment. *Nature* 434:1011-1014.

4. Jones OAH, Maguire ML, Griffin JL, Dias DA, Spurgeon DJ, Svendsen C. In Press. Metabolomics and its use in ecology. *Austral Ecol* 38:713-720.

5. Chen K, Pachter L. 2005. Bioinformatics for whole-genome shotgun sequencing of microbial communities. *PLoS Comp Biol* 1:e24.

6. Eisen JA. 2007. Environmental shotgun sequencing: Its potential and challenges for studying the hidden world of microbes. *PLoS Biology* 5:e82.

7. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers YH, Smith HO. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304:58-60.

8. Vogel TM, Simonet P, Jansson JK, Hirsch PR, Tiedje JM, van Elsas JD, Bailey MJ, Nalin R, Philippot L. 2009. TerraGenome: a consortium for the sequencing of a soil metagenome. *Nat Rev Microbiol* 7:252-252.

9. Johnson D, Campbell CD, Lee JA, Callaghan TV, Gwynn-Jones D. 2002. Arctic microorganisms respond more to elevated UV-B radiation than CO_2. *Nature* 416:82-83.

10. Garcia MH, Ivanova N, Kunin V, Warnecke F, Barry KW, McHardy AC, Yeates C, He S, Salamov AA, Szeto E, Dalin E, Putnam NH, Shapiro HJ, Pangilinan JL, Rigoutsos I, Kyrpides NC, Blackall LL, McMahon KD, Hugenholtz P. 2006. Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nat Biotechnol* 24:1263–1269.

11. Lloyd KG, Schreiber L, Petersen DG, Kjeldsen KU, Lever MA, Steen AD, Stepanauskas R, Richter M, Kleindienst S, Lenk S, Schramm A, Jorgensen BB. 2013. Predominant archaea in marine sediments degrade detrital proteins. *Nature* 496:215-218.

12. Ma Y, Paulsen IT, Palenik B. 2012. Analysis of Two Marine Metagenomes Reveals the Diversity of Plasmids in Oceanic Environments. *Environ Microbiol* 14:453-466.

13. Timmins ÉM, Howell SA, Alsberg BK, Noble WC, Goodacre R. 1998. Rapid differentiation of closely related *Candida* species and strains by pyrolysis mass spectrometry and fourier transform infrared spectroscopy. *J Clinical Microbiol* 36:367-374.

14. Kubicek CP. 2013. Systems biological approaches towards understanding cellulase production by *Trichoderma reesei*. *J Biotechnol* 163:133-142.

15. Le Belle J, Harris N, Williams S, Bhakoo K. 2002. A comparison of cell and tissue extraction techniques using high-resolution ^{1}H-NMR spectroscopy. *NMR Biomed* 15:37-44.

16. Mayr C, Winding A, Hendriksen NB. 1999. Community level physiological profile of soil bacteria unaffected by extraction method. *J Microbiol Meth* 36:29-33.

17. Westerhuis J, Hoefsloot H, Smit S, Vis D, Smilde A, van Velzen E, van Duijnhoven J, van Dorsten F. 2008. Assessment of PLSDA cross validation. *Metabolomics* 4:81-89.

18. Bundy JG, Keun HC, Sidhu JK, Spurgeon DJ, Svendsen C, Kille P, Morgan AJ. 2007. Metabolic profile biomarkers of metal contamination in a sentinel terrestrial species are applicable across multiple sites. *Environ Sci Technol* 41:4458-4464.

Figure 1. Concentrations of selected metal ions at each site. Error bars show standard error of the mean (n=3).

Figure 2. ^{1}H-NMR spectrum of soil extract from the Ecton site.

Figure 3. Principle Component Analysis (PCA) of ^{1}H-NMR data from samples from each site..

Key - ● = Cottage, ○ =West Cottage, ▼ = Vertigo, ▲ = Ecton, △ =Stream, ■ =Snail Beach, □ =Roman Gravel, ◆ = Pennerly, ◇ = Wemyss, + =Castell, | = Bog

Table 1. Geographic and Physiochemical Details of Each Sampling Site

Table 1. Geographic and Physiochemical Details of Each Sampling Site

Site	Location	pH (soil)	pH (H_2O)	% LOI*	Moisture content (%)
Ecton	+53° 7' 7.00", -1° 51' 2.00"	6.5	6.95	54.45	44.36
Wemyss	+52° 20' 59.00", -3° 53' 14.00"	4.4	6.34	12.72	40.41
Roman Gravels	+52° 35' 32.00", -2° 59' 5.00"	6.78	7.25	17.15	34.70
Snail Beach	+52° 36' 52.00", -2° 55' 34.00"	6.27	6.43	78.22	42.74
Stream	+52° 21' 38.00", -3° 45' 46.00"	4.4	6.24	11.97	38.85
Vertigo	+52° 21' 24.00", -3° 45' 23.00"	4.1	5.14	16.97	39.19
Cottage	+52° 21' 26.00", -3° 45' 26.00"	6.78	7	13.28	39.03
Penelee	+52° 35' 29.00", -2° 57' 14.00"	7.35	7.42	16.79	43.88
West cottage (II)	+52° 21' 27.00", -3° 45' 35.00"	7.22	7.17	18.40	42.47
Bog	+52° 34' 25.00", -2° 56' 56.00”	5.12	7.08	53.33	45.73
Castell	+52° 24' 52.00", -3° 48' 4.00"	4.26	6.67	16.48	39.40

* = Loss on Ignition

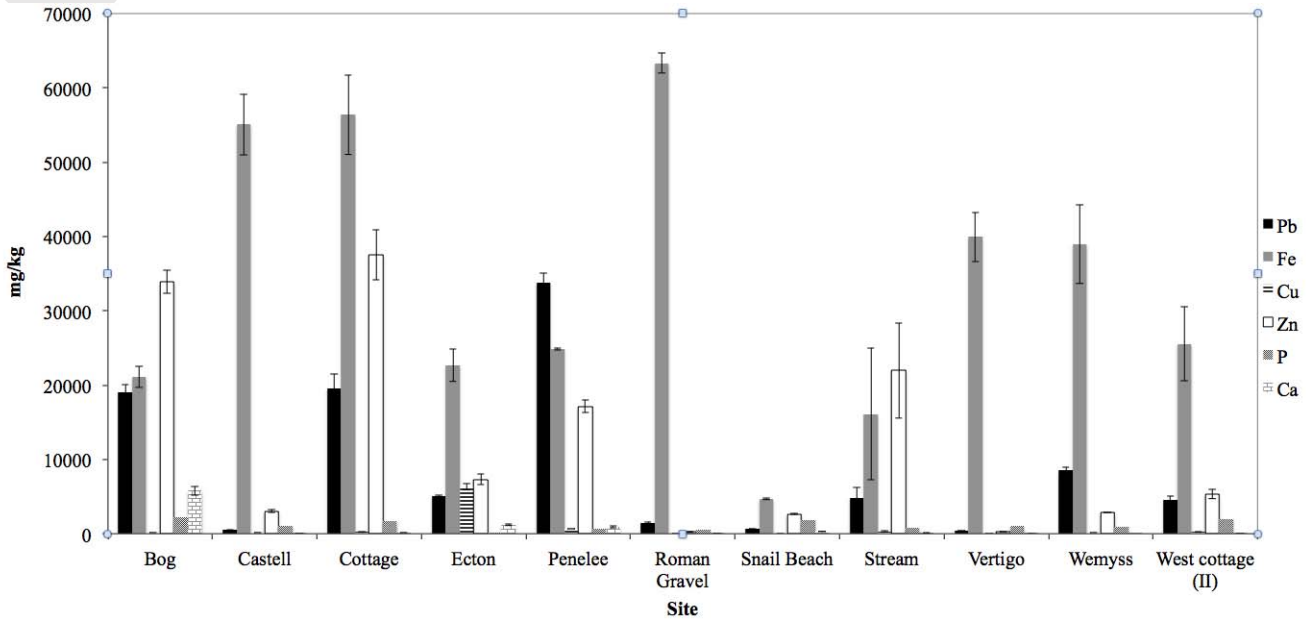

Figure 1

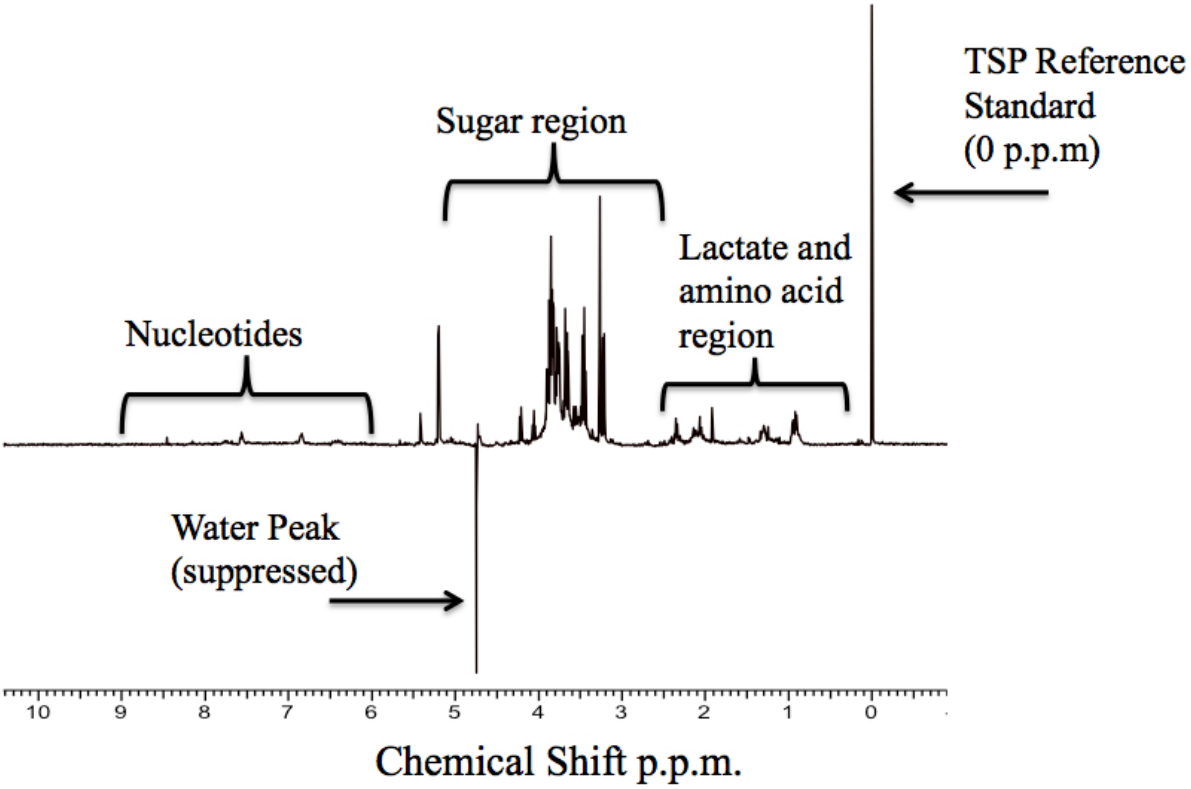

Figure 2

Key - ●=Cottage, ○=West Cottage, ▼= Vertigo, ▲ = Ecton, △=Stream, ■=Snail Beach, □=Roman Gravel, ◆=Pennerly, ◇=Wemyss, +=Castell, | = Bog

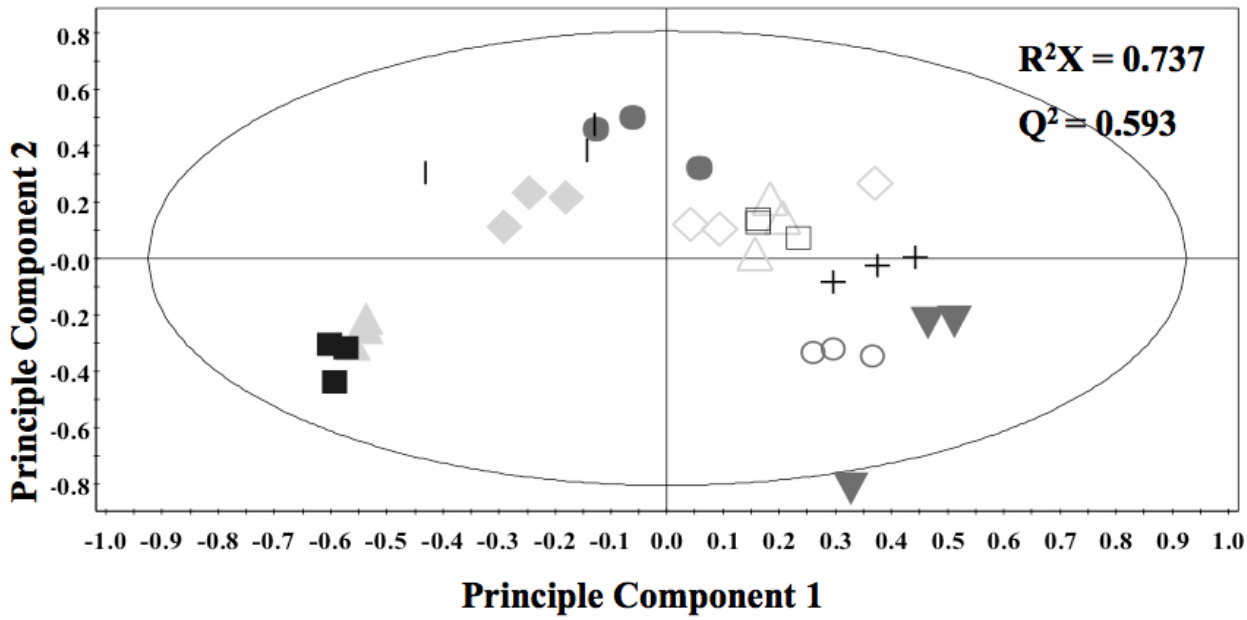

Figure 3