

## CHARACTERISATION OF AND DIFFERENTIATION BETWEEN COMPLEX BIOGEOCHEMICAL MATRICES – A ‘BIG DATA’ LIPIDOMIC APPROACH

I.S. Johari<sup>1</sup>, P.J. Gates<sup>1</sup>, A. MacDonald<sup>2</sup>, S. Freeman<sup>2</sup> and I.D. Bull<sup>1,\*</sup>

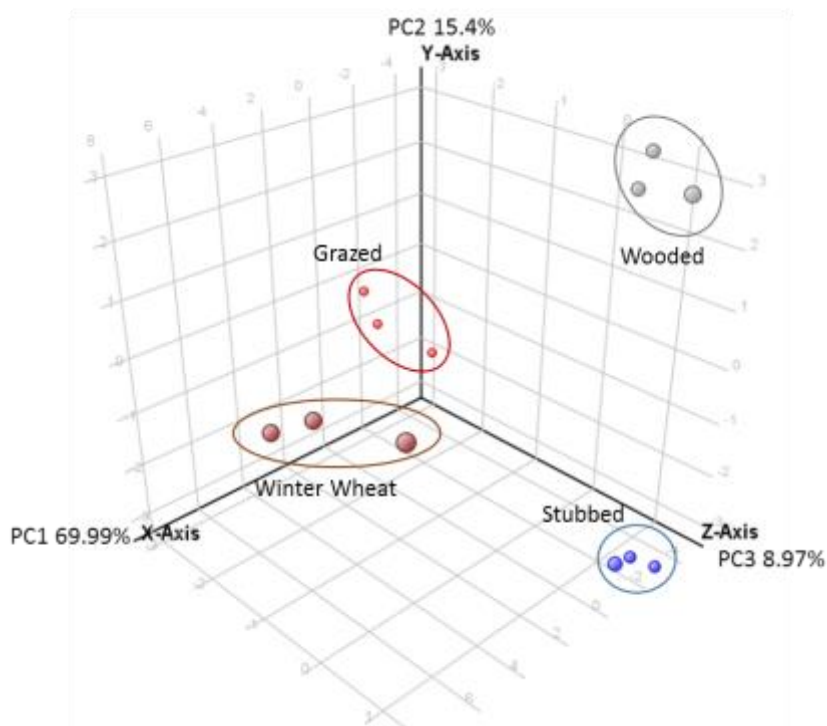
<sup>1</sup>University of Bristol, United Kingdom, <sup>2</sup>Rothamsted Research, United Kingdom,  
\*Author to whom correspondence should be addressed (ian.d.bull@bristol.ac.uk)

To date, lipidomics has been concerned with the comprehensive study of lipids in order to provide a better understanding of physiology and pathology at the cellular level (Fahy et al., 2011). The vast majority of work following this approach falls within the realms of the biomedical sciences although there is a burgeoning number of environmental forensic studies. Analogous investigations of soil (and other biogeochemically relevant) matrices have been relatively scarce with the vast majority of work focussing on either specific biomarker proxies or the comparison of homologous series of lipids largely derived from vegetation. In order to better deal with the complex datasets generated by the analysis of soil lipids using techniques such as GC-MS, a comprehensive classification of lipids together with informatics requirements has been introduced and developed (Fahy et al., 2005). Multivariate statistical analysis enables simultaneous observation and analysis of more than one variable (Dunn et al., 2011) and therefore provides a more holistic approach to the characterisation and comparison of complex analytical datasets. The development of advanced, high-throughput work-flows, based on mass spectrometry combined with an appropriate statistical analysis, will potentially improve data interpretation and thereby provide more reliable information to aid many fields of research. Unique lipid fingerprints of different soils offer less ambiguous characterisation compared with more traditional assessments and will likely enable more secure determination of possible OM sources. Moreover, a ‘big data’ lipidomic approach also facilitates a rapid and systematic identification of compounds most associated with specific inputs to and processes active in biogeochemical matrices thereby potentially providing a more efficient means of biomarker discovery.

The overarching hypotheses being tested by this research are: (H1) soils from different wide-scale geographical locations can be explicitly differentiated based on their lipidomic profile; (H2) soils from the same geographical locations but existing under different local conditions, e.g. different overlying vegetation, can be explicitly differentiated based on their lipidomic profile, (H3) a lipidomic approach facilitates better discrimination of soil compared with less comprehensive approaches, i.e. comparison of lipid compound class distributions, analysis of inorganic compounds, elemental analysis etc.; (H4) specific inputs of extraneous organic matter to a soil will result in a characteristic lipidomic profile which may be used to identify the source of the organic matter; and (H5) predictive models are a viable alternative to traditional biomarker mixing models.

Soils were obtained from a variety of sites and contexts in order to test the above hypotheses. Preliminary work focussed on soils sampled from around Bristol and the West/Southwest region of the UK, each being representative of a different environmental context, e.g. deciduous woodland, coniferous woodland, moorland etc. A high-throughput and consistent experimental workflow was devised and applied to yield a total lipid fraction from each soil. This was followed by analysis using accurate mass GC-MS (Agilent Technologies 7200B GC-QTOF), deconvolution (Masshunter, Agilent Technologies) and alignment/statistical analysis (Mass Profiler Pro, Agilent Technologies). Soils from every environmental context were successfully

differentiated thereby supporting H1. Subsequent work focussed on a set of experimental soils obtained from Rothamsted Research (Harpenden, UK; see <http://www.rothamsted.ac.uk/long-term-experiments-national-capability>) comprising: (1) Broadbalk Wilderness and Broadbalk Winter Wheat (contemporary soils); (2) Hoosfield Spring Barley (contemporary and archived soils); and (3) Park Grass (contemporary soils). Initial work focused on testing H2 through the analysis of four sets of geographically close soils (1) representative of the same parent soil but subjected to very different conditions, namely: wooded, grazed, stubbed and cultivated. Using the same experimental and data workflows developed in the preliminary study all four sets of soils were clearly grouped and differentiated between using principal components analysis (PCA, Figure 1) and hierarchical clustering analysis (HCA, not shown). These results lend strong support to H2 and emphasise the profoundly discriminating effect that localised vegetation has on the lipid profile of a soil.



**Figure 1** A three dimensional PCA score plot showing clear differentiation between the wooded, grazed and stubbed soils from Broadbalk Wilderness and cultivated soil from the Broadbalk Winter Wheat experiment.

Ongoing work is testing the remaining hypotheses utilising results obtained from soil sets (2) and (3) as well as a further series of artificial peat mixes (sedge vs *Sphagnum*).

Dunn, W.B., Erban, A., Weber, R.J.M., Creek, D.J., Brown, M., Breitling, R., Hankemeier, T., Goodacre, R., Neumann, S., Kopka, J. and Viant, M.R. (2013). Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics* 9. 44-66.

Fahy, E., Subramaniam, S., Brown, H.A., Glass, C.K., Merrill, A.H.J., Murphy, R.C., Raetz, C.R.H., Russell, D.W., Seyama, Y., Shaw, W., Shimizu, T., Spener, F., van Meer, G., VanNieuwenhze, M.S., White, S.H., Witztum, J.L. and Dennis, E.A. (2005). A comprehensive classification system for lipids. *Journal of Lipid Research* 46 839-861.

Fahy, E., Cotter, D., Sud, M., Subramaniam, S. (2011). Lipid classification, structures and tools. *Biochimica et Biophysica Acta* 1811 637-647.