

# Development of models for the analysis of metabolome variability in crops

Area of research interest: [Novel and non-traditional foods, additives and processes](#)

Study duration: 2005-12-01

Project code: G03012

Conducted by: University of Wales, Aberystwyth

[Back to top](#)

## Background

A pilot study in the G02 programme project G02006 showed that differences in data structure confounded statistical analyses aimed at producing a common understanding of metabolite baseline data. The G02 programme involved twelve collaborating sites and six projects which produced metabolomics data for potato, wheat, barley, Arabidopsis and tomato. In the emerging field of metabolomics, no widely applied standards for experimental procedure are available and a range of analytical equipment has been used. Consequently, the large amount of data collected is not necessarily directly comparable.

This project aimed to turn these disparate data, developed on different instruments using a range of standard operating procedures (SOPs), into a resource that would provide a meaningful and durable description of GM food/feed crops for any future safety assessment.

This study concentrated on developing a standardised data format for each analysis method. It would then be possible to convert existing data into data tables that can be compared meaningfully in order to develop an understanding of the baseline metabolite content and variability in traditional crop varieties and cultivars. This resource would further aid safety assessments of any proposed GM crops. Additionally, it would provide a starting point for any assessment of quality, provenance, agronomic regime or post-harvest treatment associated with a food raw material.

[Back to top](#)

## Research Approach

This project had five major interacting objectives:

- Collation of metabolite profiling/fingerprinting data from the selected G02 projects based on potato, tomato, external sources and the modification of ArMet (an internet accessible metabolomics database) to accommodate new data structures.
- Development of unified data models for G02 metabolomics data and external data.
- Development of common SOPs for data table pre-processing to allow comparison and validation of metabolite profiles and fingerprints generated in different laboratories.
- Use of standardised statistical and machine learning methods for comparing metabolite composition and generation of a baseline description of metabolite composition in crop species.
- The production of a network-accessible database on the G02 ArMet structure with a flexible interface for technical users to allow access to both metabolomics analytical data and

metadata related to quality assurance.

[Back to top](#)

## Results

This project developed standardised methods for chemical fingerprinting based on Mass Spectrometry (MS) as well as standardised guidelines for the generation and processing of Nuclear Magnetic Resonance (NMR) fingerprinting data. MS profiling was used to develop a 'checklist' of metabolite peaks for potato tubers, together with the development of the Aberystwyth Repository of Metabolite Characteristics (ARMeC) database to hold information on each chemical.

A substantial assessment of external literature and food-related databases revealed that information on the chemical composition of potato tubers was extremely disparate and biological materials and analytical techniques were often inadequately described, therefore data unification was not possible.

A major achievement of this work involved the conversion of all data analysis software into a common package alongside the development of tutorials that will allow biologists to become proficient in data mining. The development of validated procedures for the standardised generation and interrogation of metabolomics data can be used to provide an overview of the chemical composition of genetically modified (GM) plants in comparison with their non-GM comparator; which could potentially be of utility as an additional tool in the risk assessment of GM plants.

[Back to top](#)

## Additional Info

The analytical SOPs and software resources generated by the G03012 project are freely available as web resources and supported by comprehensive tutorials and example analyses which can be used 'off the shelf' to support and inform the safety assessment of GM crops in the future. The ARMeC database, is a repository for metabolite and compositional data.

The G03012 project has demonstrated that:

- The use of 'first pass' screening by metabolite fingerprinting (ESI-MS or NMR) is a cost-effective and reproducible method to rapidly compare GM/novel foods to their progenitors, as long as signal intensity/resolution thresholds are similar in any instrument used to make measurements
- After exhaustive analysis, annotation in metabolite GC-MS profiles was only successful for around 40% of peaks (i.e. approximately 60% remained unknowns); however a sub-set of expected peaks could be identified using any instruments which should be used as the basis of future rational comparisons
- Food raw materials can indeed be compared in a standardised way (in terms of both analytical chemistry technology and statistical analysis) as long as there has been a previous comprehensive analysis of the food crop species in question

A final key contribution of the G03012 project to food safety assessment procedures was the validation of a strategy to determine specific thresholds of 'similarity' between samples, below which the food raw material should be considered effectively identical when examined by the specific analytical technique in use. The project further demonstrated that only by promoting the use of specific validated quantitative measures of similarity (such as model margins, AUC or Eigen values) can the relative scale of any compositional differences between two types of food

material be assessed rationally.

In addition to food safety, in particular the safety assessment of GM crops, the technologies validated in this project are also expected to have great utility for the development of food composition databases in future studies investigating quality and provenance aspects of food raw materials.

[Back to top](#)

## Published Papers

1. Assunta Sansone, S., Schober, D., Atherton, H.J., Fiehn, O., Jenkins, H., Rocca-Serra, P., Rubtsov, D.V., Spasic, I., Soldatova, L., Taylor, C., Tseng, A. & Viant, M.R. (2007) Metabolomics Standards Initiative – Ontology Working Group – Work in Progress. *Metabolomics*, 3, 249-256.
2. Beckmann, M., Enot, D.P., Overy, D.P. & Draper, J. (2007) Representation, comparison and interpretation of metabolome fingerprint data for total composition analysis and quality trait investigation in potato cultivars. *Journal of Agricultural and Food Chemistry* 55, 3444-3451.
3. Beckmann, M., Parker, D., Enot, D.P., Duval, E. & Draper, J. (2008) High throughput, non-targeted metabolite fingerprinting using nominal mass Flow Injection Electrospray Mass Spectrometry. *Nature Protocols*, 3, 486-504.
4. Enot, D., Lin, W., Beckmann, M., Parker, D. Overy, D., & Draper, J. (2008) Pre-processing, classification modelling and feature selection using Flow Injection Electrospray Mass Spectrometry (FIE-MS) metabolite fingerprint data. *Nature Protocols*, 3, 446-470.
5. Enot, D.P. & Draper, J. (2007) Statistical measures for testing substantial equivalence of GM plant genotypes in a multivariate context. *Metabolomics* 3, 349-355. 1, 2
6. Enot, D.P., Beckmann, M. & Draper, J. (2007) Detecting a difference - assessing generalisability when modelling metabolome fingerprint data in longer term studies of genetically modified plants. *Metabolomics* 3, 335-347.
7. Enot, D.P., Beckmann, M., Overy, D. & Draper, J. (2006) Predicting interpretability of metabolome models based on behaviour, putative identity, and biological relevance of explanatory signals. *Proc Natl Acad Sci USA* 103, 14865-14870.
8. Hardy, N.W. & Jenkins, H. (2007) Reporting Standards in Metabolomics, Jens Nielsen and Michael C Jewett (Eds), In: *Topics in Current Genetics (Series: Volume 18)*, Springer, pp 53-73.
9. Hardy, N.W. & Taylor, C.F. (2007) A roadmap for the establishment of standard data exchange structures for metabolomics. *Metabolomics*, 3, 243-248. 3
10. Jenkins, J., Beckmann, M., Draper, J. & Hardy, N. (2007): GC-MS Peak labelling under ArMet. In: *Concepts in Plant Metabolomics*, Nikolau, Basil J.; Wurtele, Eve Syrkin (Eds.), Springer, ISBN: 978-1-4020-5607-9.
11. Overy, D.P., Enot, D.P., Taillart, K., Jenkins, H., Parker, D., Beckmann, M. & Draper, J. (2008) Explanatory signal interpretation and metabolite identification strategies for nominal mass FIE-MS metabolite fingerprints. *Nature Protocols*, 3, 471-485.
12. Parker, D., Beckmann, M., Enot D.P., Overy D.P., Rios, Z.C., Gilbert, M., Talbot, N. & Draper, J. (2008) Rice blast infection of *Brachypodium distachyon* as a model system to study dynamic host/pathogen interactions. *Nature Protocols* 3, 435-445.
13. Rubtsov, D.V., Jenkins, H., Ludwig, C., Easton, J., Viant, M.R., Günther, U., Griffin, J.L. & Hardy, N. (2007) Proposed reporting requirements for the description of NMR-based metabolomics experiments. *Metabolomics*, 3, 223-229.

Research report

**England, Northern Ireland and Wales**

PDF

[View FSA G03012 Final report as PDF\(Open in a new window\)](#) (8.3 MB)

## **England, Northern Ireland and Wales**

EXCEL

[View ARMeC ID sheet appendix 9 as Excel\(Open in a new window\)](#) (252.5 KB)