A.J. Márquez (Editorial Director). 2005. *Lotus japonicus* Handbook. pp. 165-174.
http://www.springer.com/life+sci/plant+sciences/book/978-1-4020-3734-4

Chapter 4.6

METABOLOME ANALYSIS USING GC-MS

Guilhem Desbrosses, Dirk Steinhauser, Joachim Kopka, and Michael Udvardi*

*Max-Planck-Institute of Molecular Plant Physiology, Plant Nutrition Group, Am Mühlenberg 1, 14476 Golm, GERMANY; *Corresponding author*

Email: *udvardi@mpimp-golm.mpg.de*
Phone: +49 331 567 8149 Fax: +49 331 567 8134

Keywords: Lotus, omics technologies, metabolism, metabolites, metabolome, GC-MS, expression profiling, systems biology, bioinformatics

Global changes in gene transcription (the transcriptome), and protein amount/activity (the proteome) ultimately effect metabolism and the metabolite content (the metabolome) of cells, tissues, and organs of plants. In the past, studies of plant metabolism focussed on one or a few genes, proteins, and/or metabolites. New tools have been developed to measure levels of thousands of gene transcripts and proteins in parallel (see Chapters 4.2-4.4), facilitating non-biased, systems-wide investigations. To complement these 'OMICS' technologies, we use gas chromatography coupled to mass-spectroscopy together with bioinformatics tools to monitor changes in the levels of hundreds of metabolites in different organs. The methods described in this chapter provide qualitative and quantitative information about the Lotus metabolome, which together with transcriptome and proteome analyses, enable new insights into the links between genotype and phenotype at the whole-system level.

INTRODUCTION

The metabolome of an organism is the complete set of metabolites produced by the organism during its life cycle under all possible conditions. Unlike the transcriptome and proteome, which can be defined quite precisely by reference to a whole genome sequence, the metabolome is difficult to come to grips with. The chemistry encompassed by the metabolome is orders of magnitude more complex than the relatively simple chemistry of nucleic acids and proteins. Thus, comprehensive analysis of an organism's metabolome is a daunting challenge, and one that has not been fully realised in any organism. Individual plant species contain tens of thousands of different metabolites, and the plant kingdom is believed to produce hundreds of thousands of distinct metabolites (Pichersky and Gang, 2000). No single analytical platform has yet been devised that can measure all

metabolites in an organism. The quest for the Holy Grail that is the metabolome has resulted in an almost bewildering array of extraction, separation, and detections systems, many of which have been optimised for the analysis of certain classes of compounds. These include separation systems like gas chromatography (GC) and liquid chromatography (LC), and numerous detection systems, including mass spectrometry (MS), nuclear magnetic resonance spectroscopy (NMR), UV, and visible light spectroscopy, and enzyme based assays (Tretheway et al., 1999; Fiehn et al., 2000b; Sumner et al., 2003; Weckwerth, 2003; Kopka et al., 2004). Coupling of powerful separation and detection systems, such as GC-MS, enables relatively non-biased analysis of thousands of known and unknown metabolites, which can provide novel insight into the biological system under study. This type of analysis has come to be known as metabolite profiling, as it falls short of complete analysis of the metabolome, called metabolomics. In this chapter, we describe methods that have been developed for routine metabolite profiling of Lotus and other plant species, using GC-MS.

Metabolite profiling with GC-MS involves six steps:

1. **Extraction** of metabolites from the biological sample, which should be as comprehensive as possible, while at the same time avoiding degradation or modification of metabolites.
2. **Derivatisation** of metabolites to make them volatile and amenable to gas chromatography. Trimethylsilylation is the most favoured of the many possible derivatisation treatments that are possible.
3. **Separation by GC,** which generally involves highly standardised conditions of gas-flow, temperature programming, and standardised column material to achieve reproducible results.
4. **Ionisation** of compounds as they are eluted from the GC, which is required for subsequent mass spectrometry. Electron impact (EI) ionisation is most widely used, as it is robust and produces reproducible fragmentation patterns and molecular ions, which simplifies the evaluation of resulting mass spectra.
5. **Detection** of molecular ions, which can be achieved with different mass-detection devices, including single quadrupole detectors (QUAD), ion trap technology (TRAP), and time-of-flight detectors (TOF). Depending on the detection system, 10-40 samples per day can be processed with these different systems, with GC-TOF-MS providing the highest throughput.
6. **Evaluation** of data begins by matching chromatographic retention times and mass-spectral fragmentation patterns to reference data in local and/or public databases. Software provided with GC-MS equipment facilitates this to a greater or lesser extent, depending on the platform used. The best software programmes support automated, comprehensive extraction of all mass spectra from a chromatogram, correction for co-eluting metabolites, calculation of GC retention time indices, and automated selection of suitable mass fragments for quantification.

These six steps are described in more detail by Kopka et al. (2004), and in the sections below.

To date, GC-MS profiling of metabolites in plants has largely been confined to hydrophilic compounds, recovered in the methanol-water phase after methanol-water/chloroform extraction of tissues. Although not all hydrophilic compounds can be volatilised by derivatisation, the following classes of compounds are detected routinely: amino-, organic-, and aromatic-acids, sugars up to trisaccharides, alcohols and polyols, sterols, diglycerides, and some mono-phosphorylated metabolites. GC-MS metabolite profiling of potato (Roessner et al., 2000; Roessner et al., 2001b; Roessner et al., 2001a), tomato (Roessner-Tunali et al., 2003), Arabidopsis (Fiehn et al., 2000b; Taylor et al., 2002), pumpkin (Fiehn, 2003), and alfalfa (Huhman and Sumner, 2002; Chen et al., 2003) have provided insight into the effects of genetic manipulation on plants and highlighted metabolic diversity amongst natural populations. We have combined GC-MS with transcriptome analysis to reveal aspects of metabolic differentiation that accompany global changes in gene expression during nodulation and symbiotic nitrogen fixation in Lotus (Colebatch et al., 2004).

MATERIALS AND METHODS

Plant metabolite extraction and derivatisation

Plant material that is harvested for GC-MS analysis is frozen immediately in liquid nitrogen and stored at -80°C until used. For each sample for metabolite analysis, between 25-50 mg (FW) frozen tissue is pulverized in a 2 ml Eppendorf tube containing a clean stainless steel metal ball (5 mm diameter) in a mixer-mill grinder (MM200, Restch, Haan, Germany) for 2 min at 30 cycles per sec. Grinding components of the mill are cooled with liquid nitrogen to keep the sample as cold as possible. Metabolites are extracted according to (Roessner *et al.*, 2000) with the following modifications: Ground samples are mixed with 360 µl methanol (-20°C) plus 30 µl ribitol in methanol (0.2 mg / ml) and 30 µl deuterated d_4-alanine in water (1mg/ml), which are included as internal standards. Samples are shaken for 15 min at 70°C before addition of 200 µl chloroform and further shaking at 37°C for 5 min. After addition of 400 µl water, samples are vortexed then centrifuged at 14.000 rpm for 5 min at RT. Two 80 µl aliquots of the aqueous phase are transferred to Eppendorf tubes and dried at RT by vacuum centrifugation (Centrivac, Hereaus, Hanau, Germany).

Metabolite samples are derivatized by methoxyamination, using a 20 mg/ ml solution of methoxyamine hydrochloride in pyridine, and subsequent trimethylsilylation, with N-methyl-N-(trimethylsilyl)-trifluoroacetamide, MSTFA, (Figure 1; Roessner et al., 2000; Fiehn et al., 2000b).

GC-MS metabolite profiles

In our institute, GC-MS spectra are typically obtained with a GC8000 gas chromatograph coupled to a Voyager quadrapol-type mass spectrometer, operated by MassLab software (ThermoQuest, Manchester, UK). Modifications to the initial GC-MS profiling method (Fiehn *et al.*, 2000b; Fiehn *et al.*, 2000a) include: injection of 1 µl sample in split-less mode; use of a 5°C min^{-1} temperature ramp with final

temperature set to 320°C on a 30 m x 0.25 mm inner diameter Rtx-5Sil MS capillary column with an integrated guard column (Restek GmbH, Bad Homburg, Germany); and use of the C_{12}, C_{15}, C_{19}, C_{22}, C_{28}, C_{32}, C_{36} n-alkane mixtures for the determination of retention time indices (RI). Typical GC-MS profiles of different organs of Lotus are shown in Figure 2.

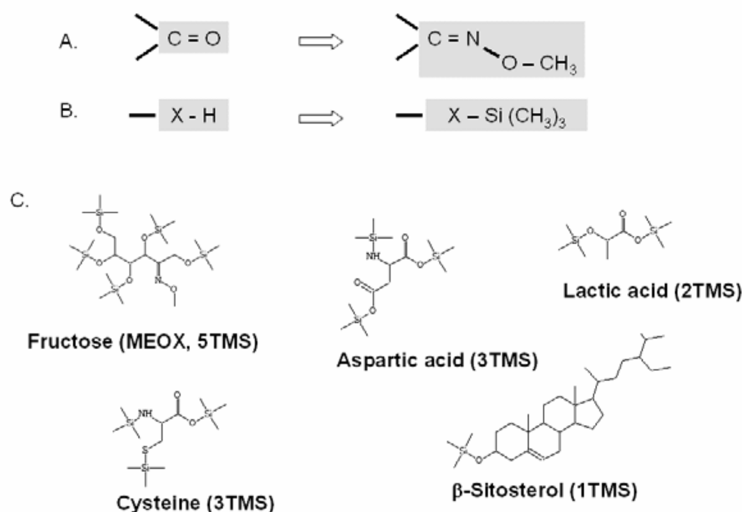

Figure 1. Chemical derivatisation and some typical metabolite derivatives. (A) Methoxyamination of carbonyl moieties. (B) Trimethylsilylation of acidic protons such as are present in hydroxyl-, sulfhydryl- carboxyl-, phosphate- or amino-groups. (C) Five typical metabolite derivates.

Mass-spectral tags for metabolites

Gas chromatography separates complex mixtures of metabolite derivatives into a series of compounds that enter the mass spectrometer and are subsequently ionized, fragmented, and detected. Each metabolite is, therefore, represented by one or more ionic fragments of precise mass, which together can serve as a 'tag' for that metabolite. We have termed these mass-spectral tags (MST), by analogy to expressed sequence tags (ESTs) of genes. Each MST has properties that facilitate unequivocal identification of the parent metabolite, following comparison to the pure reference compound (Wagner et al., 2003). The properties are (1) gas chromatographic retention, which is best characterised by a retention time index (RI), and (2) a specific composition of fragments, which are each characterized by a mass to charge ratios (m/z). A compendium library of MSTs has been collected from a set of Lotus organs using AMDIS software (Automated Mass Spectral Deconvolution and Identification System).

Fragments that belong to one MST have the same retention time index and occur in fixed relative abundance, independent of metabolite concentration. Therefore, any

single fragment or set of fragments with identical retention time index can be used for the quantification of metabolites. As a rule, choice of fragments for quantitative purposes must be selective, i.e. only those fragments that are unique to a MST can be used. Fragments that are common to co-eluting MSTs or MSTs with similar retention time indices must be avoided for quantification purposes.

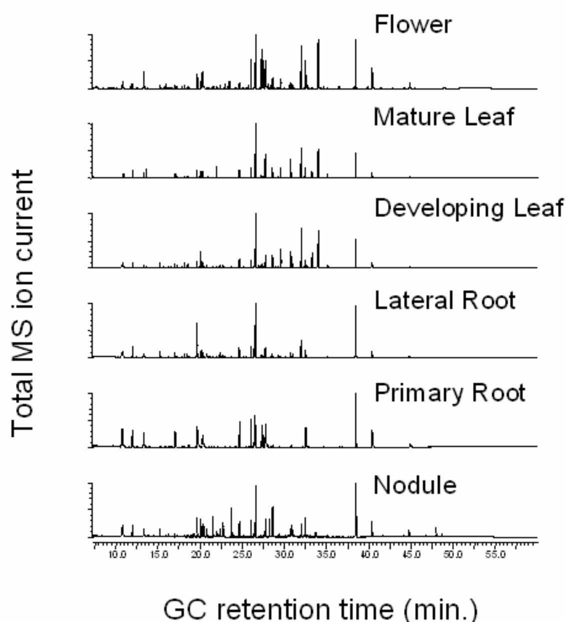

Figure 2. Typical GC-MS profiles of Lotus organs.

Fragments that are used for quantification are identified by mass to charge ratio (m/z), RI, and metabolite derivative. We use the following nomenclature for GC-MS fragments: mass to charge ratio followed by retention time index and name both separated by an underlined character, for example fragment 333_2014 Glucaric acid (6TMS) (Figure 3). MSTs that remain unidentified are classified tentatively by best matching mass spectra from a custom and a commercial NIST02 library (Institute of Standards and Technology, Gaithersburg, USA). A tentative match required a score >600 on a scale of 0-1000. We use the nomenclature, match value, and substance name of best fit, separated by a semicolon character and set into square brackets, for example 243_1930_[802; Methylcitric acid (4TMS)].

Compendium library of mass spectral metabolite tags

Mass spectral metabolite tags (MSTs) were obtained by automated de-convolution of GC-MS chromatograms using publicly available software (AMDIS, http://chemdata.nist.gov/mass-spc/amdis/; National Institute of Standards and

Technology, Gaithersburg, USA) (Stein, 1999). Mass spectra were collected above a threshold of 0.001% of total signal. Two independent GC-MS metabolite profiles of each sample type were processed. RIs of all MSTs were determined by AMDIS software. RI-annotated MSTs were uploaded into a non-supervised custom NIST02 mass spectral library (NIST02 mass spectral search program, http://chemdata.nist.gov/mass-spc/Srch_v1.7/index.html; National Institute of Standards and Technology, Gaithersburg, USA) (Ausloos *et al.*, 1999). Mass spectra of low relative amount were rejected. The resulting MST library of *Lotus japonicus* root and shoot organs is available as Supplementary Material, NIST02 custom library file Q_LJA_NS (Figure 4; see also: http://csbdb.mpimp-golm.mpg.de/gmd.html).

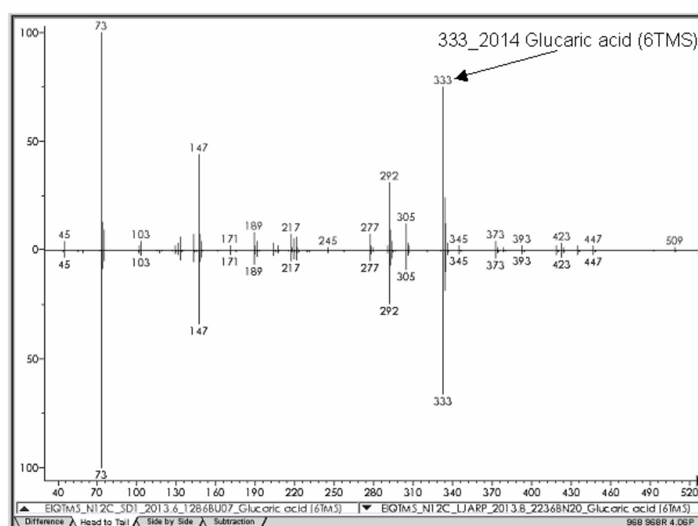

Figure 3. Head to Tail matching of glucaric acid mass spectra from the pure reference compound (above) and from a preparation of Lotus primary roots (below). NIST02 software and customized user libraries of mass spectra were used.

Identification of MSTs

MSTs are identified by manual comparison with MSTs derived from commercially available pure standard compounds (Figure 2). Standard compounds are processed through standard addition experiments in order to obtain respective MEOX- and MSTFA-derivative mass spectra and corresponding retention time indices. Required criteria for MST identification are chromatographic co-elution within an RI window of ± 2.5, and high mass spectral similarity, with matching value >700, on a scale of 0-1000. MST identification is supported by comparison with mass spectral entries from the commercial NIST02 library (Institute of Standards and Technology, Gaithersburg, USA). The current supervised and non-redundant MST Lotus library is available as Supplementary Material, NIST02 custom library file Q_LJA_ID.

A.J. Márquez (Editorial Director). 2005. *Lotus japonicus* Handbook. pp. 165-174.
http://www.springer.com/life+sci/plant+sciences/book/978-1-4020-3734-4

Identified metabolites and additional information can be found at the Golm Metabolome Database (GMD; http://csbdb.mpimp-golm.mpg.de/gmd.html).

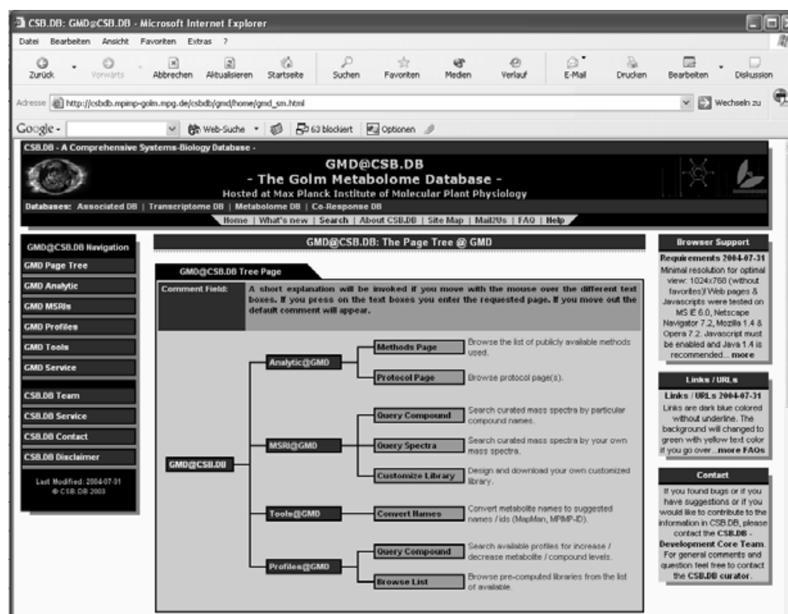

Figure 4. Resources for GC-MS Mass Spectral Metabolite Identification – The Golm Metabolome Database (GMD) http://csbdb.mpimp-golm.mpg.de/gmd.html

Lactic acid and benzoic acid are typical laboratory contaminants. Oligomethyl-cyclosiloxanes are also common contaminants, and accumulate with time in the MSTFA reagent used for derivatisation.

Generation of a metabolite response matrix

For each metabolite that is identified in the non-supervised MST library, a specific mass fragment and corresponding retention-time window is selected. The find algorithm of the MassLab software (ThermoQuest, Manchester, UK) is then used to automatically retrieve peak areas from GC-MS metabolite profiles. Correct peak integration is monitored manually. Peak areas with low intensity are rejected.

Numerical analysis

Peak areas, X, are equivalent to what we call fragment responses (X_i of fragment i). Fragment responses are normalized to the fresh weight of the sample and response of the internal standard, ribitol, ($N_i = X_i * X_{ribitol}^{-1} * \text{fresh weight}^{-1}$). This procedure corrects pipette errors and slight differences in sample amount. In a further step, the relative response of a fragment, R_i, is divided by the average normalised responses of all samples in one experiment or, alternatively of all non-treated control samples

of an experiment ($R_i = N_i * avgN^{-1}$) to facilitate comparisons between samples. All response ratios are combined into a single matrix that described the complete set of metabolite response ratios of all samples in an experiment or group of experiments.

Principal components analysis

Principal components analysis (PCA) is typically performed after log_{10} transformation of data in a metabolite response matrix. Missing values are defined as 0 after log_{10} transformation. This substitution procedure assumes that missing values are unchanged compared to the co-analysed standard samples. The S-Plus 2000 software package standard edition release 3 (Insightful, Berlin Germany) is used for PCA and visualization. PCA analysis can reveal differences between the metabolic phenotype of different organs of wild type Lotus plants (Figure 5), or between the same organs of different plant genotypes (Taylor et al., 2002).

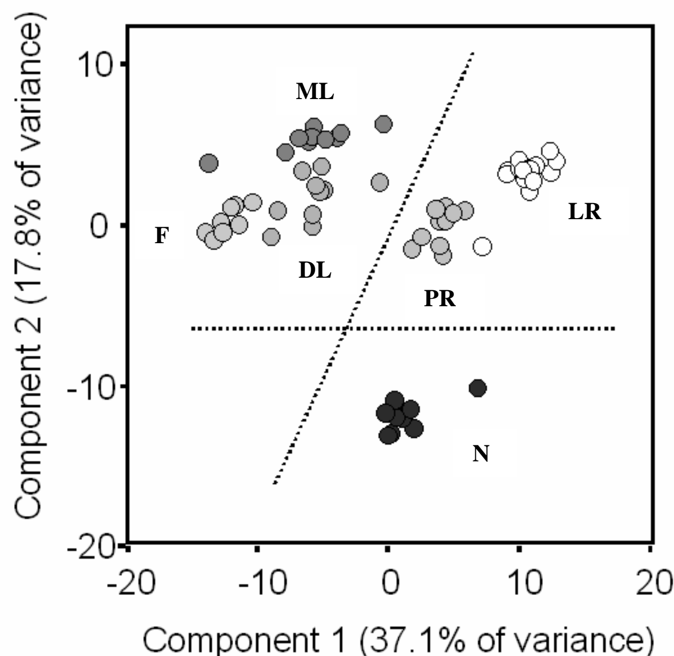

Figure 5. PCA analysis of GC-MS data from different Lotus organs. Filled circles represent individual samples from different organs: F, flowers; DL, developing leaves; ML mature leaves; LR, lateral roots; PR, primary roots; N, nodules. Clustering of samples from the same organ and separation of samples from different organs are obvious.

A.J. Márquez (Editorial Director). 2005. *Lotus japonicus* Handbook. pp. 165-174.
http://www.springer.com/life+sci/plant+sciences/book/978-1-4020-3734-4

Hierarchical cluster analysis (HCA)

Hierarchical cluster analysis (HCA) is used to classify samples and MSTs, which represented identified metabolites and a selection of non-identified metabolites, according to their relative abundance in different samples or sample types. For this purpose, average normalized responses ($avgN_i$) are calculated for each MST and organ. Missing data are substituted by the normalized response at detection limit. HCA is performed after range normalization using Euclidian distance and average linkage. All procedures including analysis of variance and visualisation are performed with EXCEL software and the S-Plus 2000 software package standard edition release 3 (Insightful, Berlin Germany).

REFERENCES

Castrillo JI, Hayes A, Mohammed S, Gaskell S.J, and Oliver SG. (2003). **An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry**. *Phytochemistry* 62, 929-937.

Chen F, Duran AL, Blount J.W, Sumner LW, Dixon RA, and Huhman D.V. (2003). **Profiling phenolic metabolites in transgenic alfalfa modified in lignin biosynthesis**. *Phytochemistry* 64, 1013-1021.

Colebatch G, Desbrosses G, Ott T, Krusell L, Kloska S, Kopka J, and Udvardi MK. (2004) **Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus***. *The Plant Journal* 39, 487-512.

Fiehn O. (2003). **Metabolic networks of *Cucurbita maxima* phloem**. *Phytochemistry* 62, 875-886.

Fiehn O, Kopka J, Trethewey RN, and Willmitzer L. (2000a). **Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry**. *Analytical Chemistry* 72, 3573-3580.

Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, and Willmitzer L. (2000b). **Metabolite profiling for plant functional genomics**. Nat Biotechnol 18, 1157-1161.

Griffin JL, Cemal CK, and Pook MA. (2004). **Defining a metabolic phenotype in the brain of a transgenic mouse model of spinocerebellar ataxia 3**. *Physiological Genomics* 16, 334-340.

Huhman DV, and Sumner LW. (2002). **Metabolic profiling of saponins in *Medicago sativa* and *Medicago truncatula* using HPLC-coupled to an electrospray ion-trap mass spectrometer**. *Phytochemistry* 59, 347-360.

Kopka J, Fernie A, Weckwerth W, Gibon Y, and Stitt M. (2004). **Metabolite profiling in plant biology: platforms and destinations**. Genome Biology 5, 109.

Pichersky E. and Gang D. (2000). **Genetics and biochemistry of secondary metabolites: An evolutionary perspective**. *Trends in Plant Sciences* 5, 439-445.

Roessner U, Wagner C, Kopka J, Trethewey RN, and Willmitzer L. (2000). **Technical advance: simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry**. *Plant Journal* 23, 131-142.

Roessner U, Willmitzer L, and Fernie AR. (2001a). **High-resolution metabolic phenotyping of genetically and environmentally diverse potato tuber systems**. **Identification of phenocopies**. *Plant Physiology* 127, 749-764.

A.J. Márquez (Editorial Director). 2005. *Lotus japonicus* Handbook. pp. 165-174.
http://www.springer.com/life+sci/plant+sciences/book/978-1-4020-3734-4

Roessner U, Luedemann A, Brust D, Fiehn O, Linke T, Willmitzer L, and Fernie A. (2001b). **Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems**. *Plant Cell* 13, 11-29.

Roessner-Tunali U, Hegemann B, Lytovchenko A Carrari F, Bruedigam C, Granot D, and Fernie AR. (2003). **Metabolic profiling of transgenic tomato plants overexpressing hexokinase reveals that the influence of hexose phosphorylation diminishes during fruit development**. *Plant Physiology* 133, 84-99.

Sumner LW, Mendes P, and Dixon RA. (2003). **Plant metabolomics: large-scale phytochemistry in the functional genomics era**. *Phytochemistry 6*26, 817-36.

Taylor J, King RD, Altmann T, and Fiehn O. (2002). **Application of metabolomics to plant genotype discrimination using statistics and machine learning**. *Bioinformatics* 18, S241-248.

Trethewey RN, Krotzky AJ, and Willmitzer, *L. (1999)* **Metabolic profiling : a Rosetta stone for genomics?** *Current Opinion in Plant Biology* 2, 83-85.

Wagner C, Sefkow M, and Kopka J. (2003). **Construction and application of a mass spectral and retention time index database generated from plant GC/EI-TOF-MS metabolite profiles***.* *Phytochemistry* 62, 887-900.

Weckwerth W. (2003). **Metabolomics in Systems Biology**. *Annual Review of Plant Biology* 54, 669-689.