Available online at www.sciencedirect.com
SCIENCE DIRECT®
PERGAMON
Phytochemistry 62 (2003) 949–957
PHYTOCHEMISTRY
www.elsevier.com/locate/phytochem

Assessment of ^{1}H NMR spectroscopy and multivariate analysis as a technique for metabolite fingerprinting of *Arabidopsis thaliana*

Jane L. Ward, Cassandra Harris, Jennie Lewis, Michael H. Beale*

IACR-Long Ashton Research Station, Department of Agricultural Sciences, University of Bristol, Long Ashton, Bristol BS41 9AF, UK

Received 26 August 2002; received in revised form 14 November 2002

Abstract

An approach to metabolite fingerprinting of crude plant extracts that utilizes ^{1}H nuclear magnetic resonance (NMR) spectroscopy and multivariate statistics has been tested. Using ecotypes of *Arabidopsis thaliana* as experimental material, a method has been developed for the rapid analysis of unfractionated polar plant extracts, enabling the creation of reproducible metabolite fingerprints. These fingerprints could be readily stored and compared by a variety of chemometric methods. Comparison by principal component analysis using SIMCA-P allowed the generation of residual NMR spectra of the compounds that contributed significantly to the differences between samples. From these plots, conclusions were drawn with respect to the identity and relative levels of metabolites differing between samples.

Keywords: *Arabidopsis thaliana*; NMR spectroscopy; Metabolomics; Multivariate analysis; Principal component analysis; Metabolite fingerprints

1. Introduction

Arabidopsis thaliana (*Arabidopsis*) is well known as a model system in plant research due to its relatively small genome, rapid life cycle, easy cultivation and high level of seed production. The completion of the genome sequence of *Arabidopsis* has provided the impetus for understanding the function of all the genes in this model plant (The *Arabidopsis* Genome Initiative, 2000; Wixon, 2001). Techniques such as proteomics and metabolomics may provide the necessary data to link gene sequence to function via the metabolic network (Hall et al., 2002; Fiehn, 2002) and thus high-throughput metabolomic analysis in *Arabidopsis* is an important goal in plant functional genomics (Trethewey, 2001). The "metabolome" has been defined, in a microbial context, as the total complement of metabolites in a cell (Tweeddale et al., 1998). For plants, examination of the metabolome is a more complex problem due to the larger number of potential metabolites and the presence of differentiated tissue, including specialist storage organs, with different metabolite complements. It is unlikely that a single analytical method will yield information about all the metabolites in a plant system. Differences due to volatility, polarity, solubility and chromatographic behaviour mean that multiple methods will need to be deployed to analyse different subsets of metabolites. In this context coupled gas chromatography–mass spectrometry (GC–MS) has already been successfully applied to plant metabolite profiling (Roessner et al., 2001), including *Arabidopsis* (Fiehn et al., 2000), where 326 distinct compounds from leaf extracts were quantified. Another potentially powerful tool for plant metabolite analysis is high-resolution nuclear magnetic resonance spectroscopy (NMR), in particular ^{1}H NMR. This technology has been utilized extensively to profile metabolites in clinical samples (e.g. Nicholson and Wilson, 1989; Holmes et al., 2000; Beckwith-Hall et al., 2002) and has been applied to complex mixtures of compounds exuded from cereal roots (Fan et al., 2001). Unlike GC–MS, which detects only those compounds that can be volatilized (usually achieved by derivatization), ^{1}H NMR can simultaneously detect all proton-bearing compounds in a sample. This covers most of the "organic" compounds such as carbohydrates, amino acids, organic and fatty acids, amines, esters, ethers and

* Corresponding author. Tel.: +44-1275-549289; fax: +44-1275-394281.
E-mail address: mike.beale@bbsrc.ac.uk (M.H. Beale).

0031-9422/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S0031-9422(02)00705-7

lipids, which are present in plant tissues. Thus, NMR spectra of unpurified solvent extracts of plants has the potential to provide a relatively unbiased fingerprint, containing overlapping signals of the majority of the metabolites present in the solution.

In this paper we report the development of one-dimensional ^{1}H NMR spectroscopy methods, coupled with multivariate statistical analysis (Antti et al., 2002), for the analysis of crude extracts of *Arabidopsis*. A small set of ecotypes was used as suitable experimental material to develop the method, as a previous GC–MS study had shown that the metabolite profiles of two ecotypes, (Col-2 and C24) showed significant differences (Fiehn et al., 2000).

2. Results and discussion

2.1. Extraction and analysis of plants

The *Arabidopsis* ecotypes employed (Table 1) were grown under identical long-day controlled environment conditions in trays containing 24 individual plants. Plants were harvested at growth stage 6.1–6.5 (just bolted, first flower present) as described by Boyes et al. (2001). In order to smooth out plant to plant variability, all aerial plant material from each tray was combined. Enzyme activity was stopped by immediately immersing the harvested material in liquid nitrogen, before freeze-drying. The extraction method developed is relatively simple, requiring a suspension of weighed aliquots of the powdered, freeze-dried plant material in deuterated NMR solvent (80:20 D_2O:CD_3OD), a short period of moderate heating, followed by micro-centrifugation. An aliquot of the supernatant was then analysed directly by ^{1}H NMR. An obvious advantage of this method is the use of deuterated solvents for tissue extraction. This eliminated the need for evaporation and re-dissolution of extracts, which has the associated potential problem of loss of material. Three sample replicates were taken in each case to assess the robustness of the sample preparation method.

2.2. Features of ^{1}H NMR spectra of polar extracts of *Arabidopsis*

In general the NMR spectra obtained showed a dominance of signals in the carbohydrate region of the spectrum. In addition to these signals, well-defined signals in both the aromatic and aliphatic regions of the spectra were present (Fig. 1). The sharp singlet at δ 6.5 was identified as fumaric acid. Similarly, other signals in relatively clear areas of the trace could be assigned to

Table 1
Arabidopsis ecotypes used in assessment of multivariate analysis by ^{1}H NMR spectroscopy

Name of ecotype	Code	NASC code	Country of origin
Columbia	COL-0	N1092	USA
Landsberg	LER-1	N1642	Germany
Dijon	Di-0	CS1106	France
Estland	Est-0	N1148	Russia
Nossen	No-0	N3081	Germany
Wassilewskija	WS-0	N1602	Russia
Wassilewskija	WS-2	N1601	Russia
C24	C24	N906	USA
Rschew	Rld-2	N1641	Russia

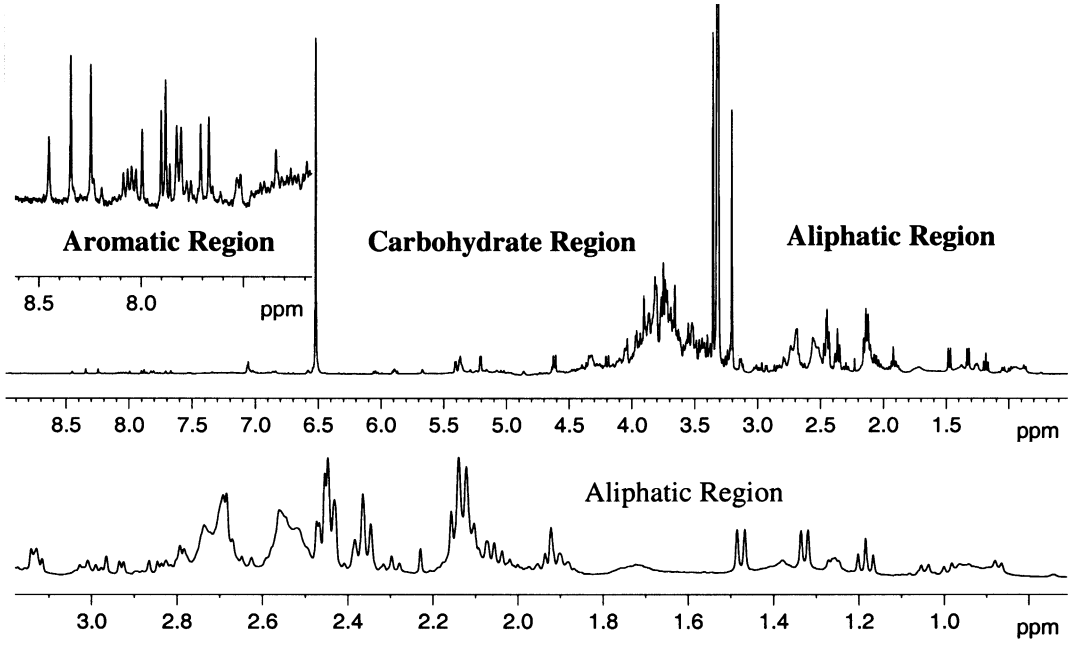

Fig. 1. ^{1}H NMR spectrum of a typical *Arabidopsis* (Landsberg) polar extract in D_2O:CD_3OD.

particular amino acids (e.g. alanine doublet at δ 1.47 and threonine doublet at δ 1.32), and particular carbohydrates (e.g. α- and β-glucose anomeric hydrogens at δ 5.20 and 4.60). From visual analysis of spectra from the nine ecotypes, clear differences were evident. For example, by comparison to the other eight ecotypes in the set, WS-0 had significantly increased intensity in many of the carbohydrate signals (Fig. 2). In addition, the ecotype Dijon possessed completely new signals in the region δ 6.0–4.90 (Fig. 2). Other differences included a variation in intensities of the same signals in different ecotypes (e.g. the fumaric acid signal).

2.3. Standardization and processing of the data for export

For electronic comparison of the data sets by multivariate methods it was important to ensure that there is as little experimental variation as possible in the sample set. The spectra were all Fourier-transformed, in automation, using the same processing parameters, an exponential window and a line-broadening factor of 0.5 Hz. Each data set was automatically scaled to trimethylsilylpropionate-d_4 internal standard, phased and baseline corrected. After importation into AMIX (Analysis of MIXtures software, Bruker, Germany) the negative peaks were removed and a compressed form of the data was stored in a spectral database for future reference. Data from standards were collected and processed in an identical fashion and stored in the AMIX spectral database. Before analysis by multivariate methods, data sets, selected from the database, were reduced in complexity by using the "bucketing" function to generate a set number of integrated regions or "bins" of the data set. This table of 'binned' data from those spectra selected could then be exported as a spreadsheet suitable for importation into statistical analysis software, such as SIMCA-P (Umetrics). The ability to batch process datasets from any number of samples, held in the database, as described represents a further benefit of using ^{1}H NMR to collect metabolite fingerprints. Currently some other methodologies for large-scale metabolite fingerprinting, for example GC–MS, the ability to database aligned and normalised data

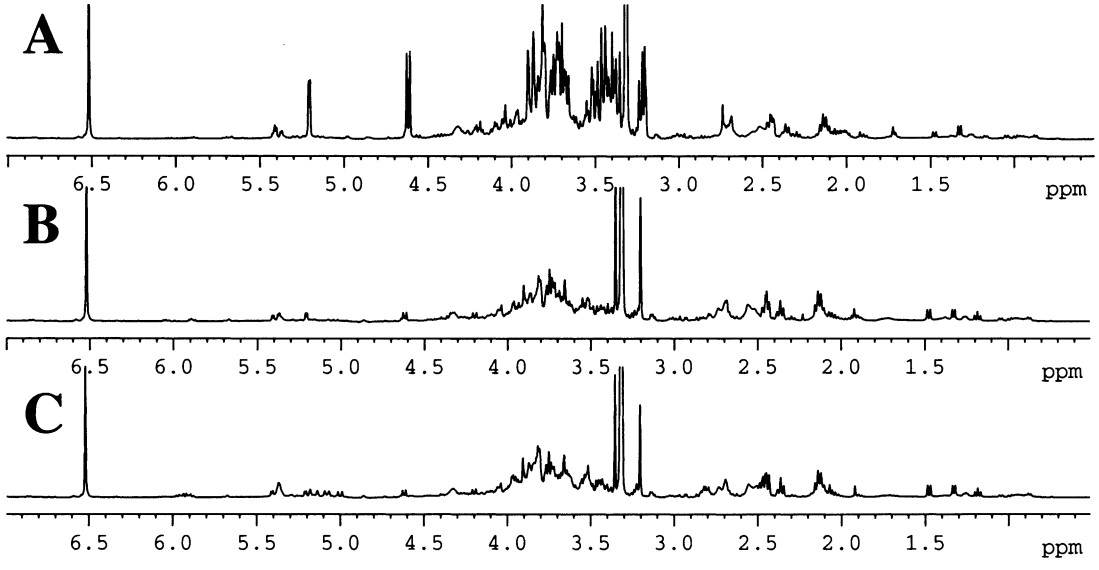

Fig. 2. ^{1}H NMR spectra of three *Arabidopsis* ecotypes. A: WS-0, B: Landsberg, C: Dijon.

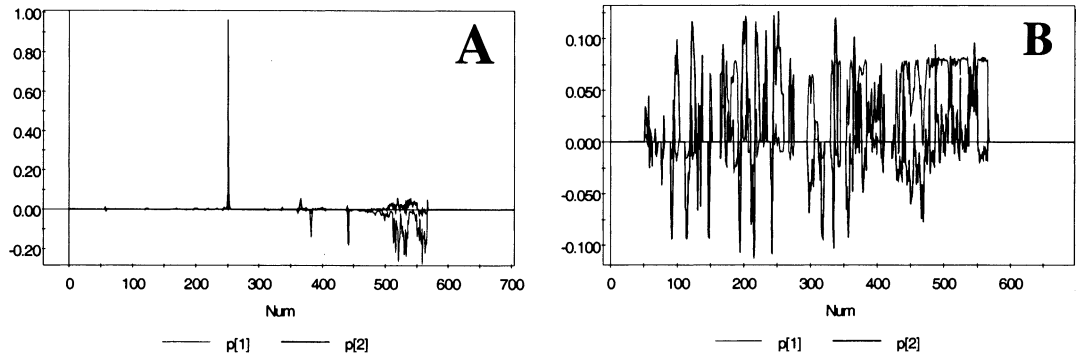

Fig. 3. Comparison of typical loadings plots generated using A: the covariance matrix and B: the correlation matrix.

sets *in automation* is not easily achieved within current spectrometer operating software. The inherent problems of retention time drift, and column and source variability, mean that peak alignment and data export methods from GC–MS require operator quality control to ensure accurate peak alignment to prepare each data set for alignment, storage and multivariate analysis.

2.4. Principal component analysis (PCA) of Arabidopsis ecotype data sets

PCA is a data visualization method that is useful for observing groupings within multivariate data. Data is represented in n dimensional space, where n is the number of variables, and is reduced into a few principal components, which are descriptive dimensions that describe the maximum variation within the data. The principal components can be displayed in a graphical fashion as a "scores" plot. This plot is useful for observing any groupings in the data set and in addition will highlight outliers that may be due to errors in sample preparation or instrumentation parameters. PCA models are constructed using all the samples in the study. Coefficients by which the original variables must be multiplied to obtain the PC are called "loadings." The numerical value of a loading of a given variable on a PC shows how much the variable has in common with that component (Massart et al., 1988). Thus for NMR data, "loading plots" can be used to detect the spectral areas responsible for the separation in the data.

The data for PCA can be scaled in different ways. If the data is mean-centred with no scaling then a covariance matrix is produced, but if the data mean-centred and the columns of the data matrix scaled to unit variance, a correlation matrix is produced. An advantage of the covariance matrix is that the loadings retain the scale of the original data. In the case of the data reported here, the loadings plots, when viewed as line plots, resemble NMR spectra and can be interpreted as such (Fig. 3). In contrast, the correlation matrix produces loadings plots which are unfamiliar in appearance (Fig. 3). For the purposes of this work, a covariance matrix was used to allow for a more useful interpretation of the loadings plots. Contribution plots allow further interpretation of the differences observed in the scores diagram, and depict the changes in variables (e.g. chemical shift) between two observations (samples) or between a selected observation and the average. When plotted as line diagrams these also resemble NMR spectra and in that sense depict spectra of compounds responsible for the differences between chosen samples.

For the data set obtained from replicate analysis of the ecotypes, a nine-component model explained 99% of the variance, with the first two components explaining

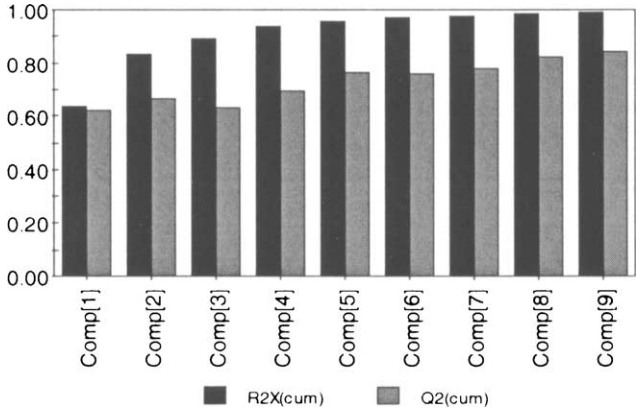

Fig. 4. Model overview illustrating the number of components and explained variances used in PCA analysis of *Arabidopsis* ecotypes.

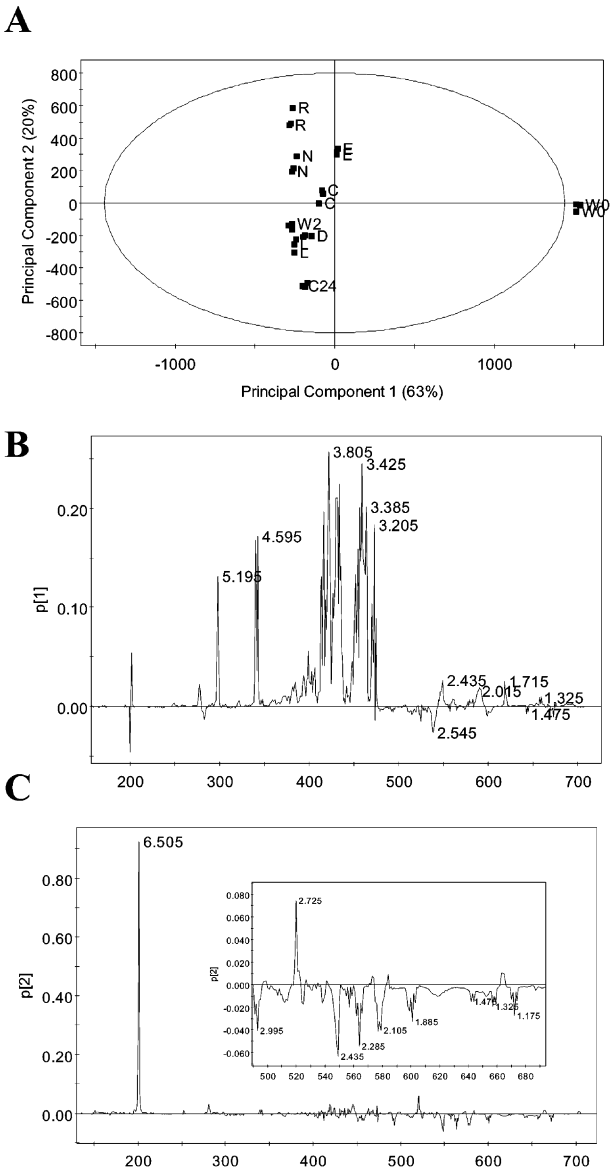

Fig. 5. Scores and loadings plots generated from PCA of *Arabidopsis* ecotypes. A: scores plot of PC1 vs. PC2. B: loadings plot of PC1. C: loadings plot of PC2. W0 = WS-0, W2 = WS-2, E = Estland, D = Dijon, R = Rschew, N = Nossen, C = Columbia-0, L = Landsberg, C24 = C24.

83% of the variability (Fig. 4). Examination of the scores and loadings plots for PC1 vs. PC2 (Fig. 5) showed good experimental replication since tight clustering of replicate samples could be seen with several of them clustering on top of each other. Examination of the scores plot (Fig. 5A) demonstrated that WS-0 was separate from the rest of the group. Examination of the loadings plot of PC1 (Fig. 5B) showed that the first component explained the variance in carbohydrate levels since high loadings values were observed for peaks in the carbohydrate region of the NMR spectrum. In addition, the loadings plot of PC1 illustrated some small positive and some small negative regions of the spectrum between δ 2.5 and δ 1.25 (Fig. 5C). This region contained many peaks attributable to amino acids and this information may give clues as to the variance of amino acids between ecotypes. It is evident that WS-0 is separated mainly by virtue of its increase in carbohydrate relative to the rest of the group. Examination of the scores plot (Fig. 5A) also indicated that the rest of the set of ecotypes had fairly similar levels of carbohydrate. In order to correctly determine the nature of this increased carbohydrate we examined, in AMIX, the spectrum of WS-0 against a library of spectra of carbohydrates run in the same solvent under the same conditions. As can be seen in the contribution plot, Fig. 6A, the increased peaks were due to glucose (approx. 1:1 anomeric mixture). Thus it would appear that WS-0 has elevated levels of glucose relative to all of the other ecotypes examined here. This observation was confirmed by the quantitative GC–MS analysis of methoxyamine-terimethylsilylated samples relative to added ribitol internal standard (Roessner et al., 2001). The results indicated that glucose levels in WS-0 were four times higher than in the other ecotypes, while other simple carbohydrates, such as fructose, mannose and galactose, which are of lower abundance than glucose,

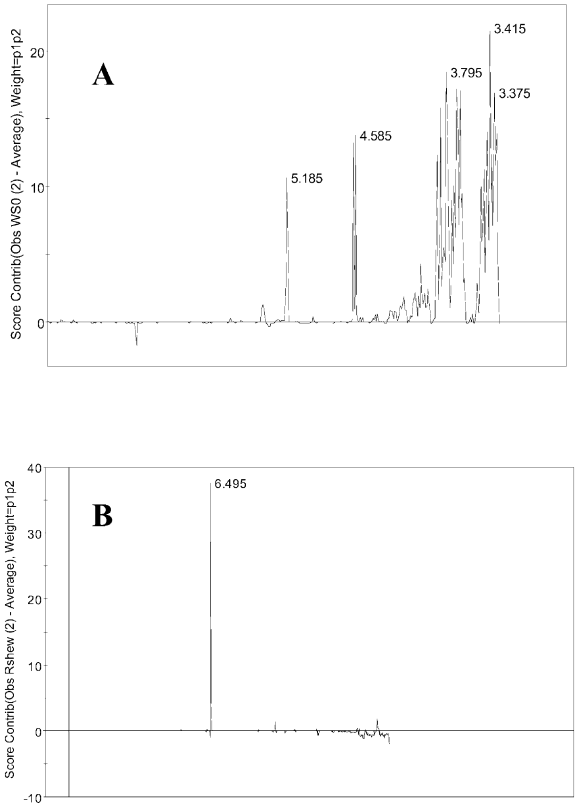

Fig. 6. Contribution plots of A: WS-0 *minus* average and B: C24 *minus* average, generated from PC1 vs. PC2 scores plot from PCA of *Arabidopsis* ecotypes.

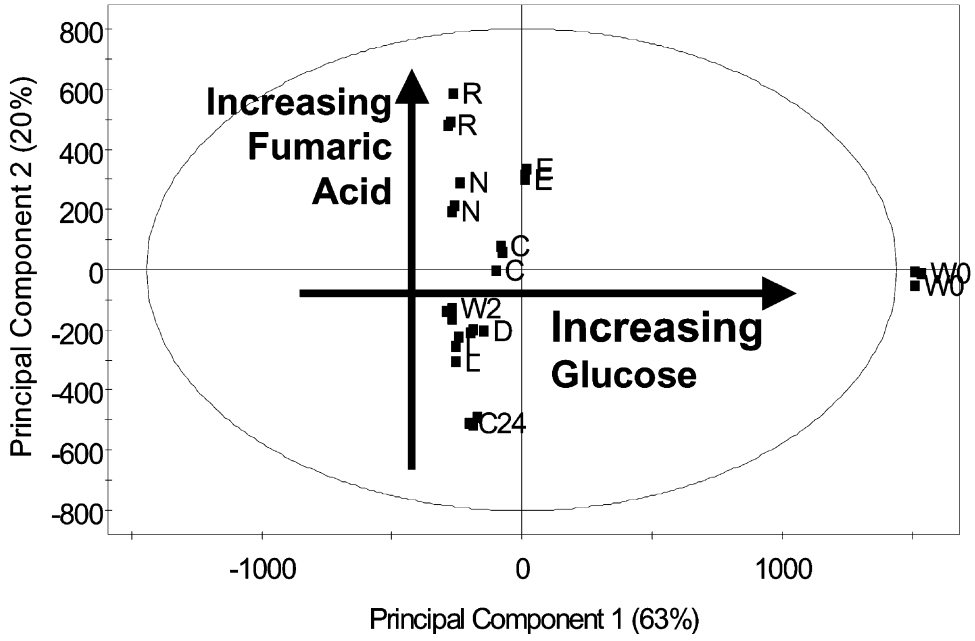

Fig. 7. Analysis and summary of the scores plot of PC1 vs. PC2, after examination of the associated loadings plots, indicating the metabolites responsible for the greatest variance.

were also elevated to 2–3 times those of the other ecotypes. The origin of the highly elevated monosaccharides in WS-0 is unclear at present. Ecologically, the origin of WS-0 resembles WS-2. It is possible that the high sugar levels are a result of increased polysaccharide hydrolytic activity in this ecotype. The possibility that this kind of enzyme activity may be manifested during sample processing was investigated by repeating the extraction procedure several times and by re-running the NMR spectra after storage of the samples. No indication of such post-harvest degradation was found, but further experiments are necessary to fully investigate this. The loadings plot (Fig. 5C) and the contribution plot (Fig. 6B) of the second principal component PC2 was relatively simple with a large peak at δ 6.5. This peak has been identified as fumaric acid by the comparison of a set of organic acid standards run in the same solvent system and confirmed by addition of fumaric acid to an *Arabidopsis* NMR sample. The scores plot of PC1 vs. PC2 can now be summarized according to Fig. 7, which shows that the level of fumaric acid in the set of ecotypes varies with Rschew having the highest amount and C24 possessing the least. Columbia fumarate levels are intermediate between these.

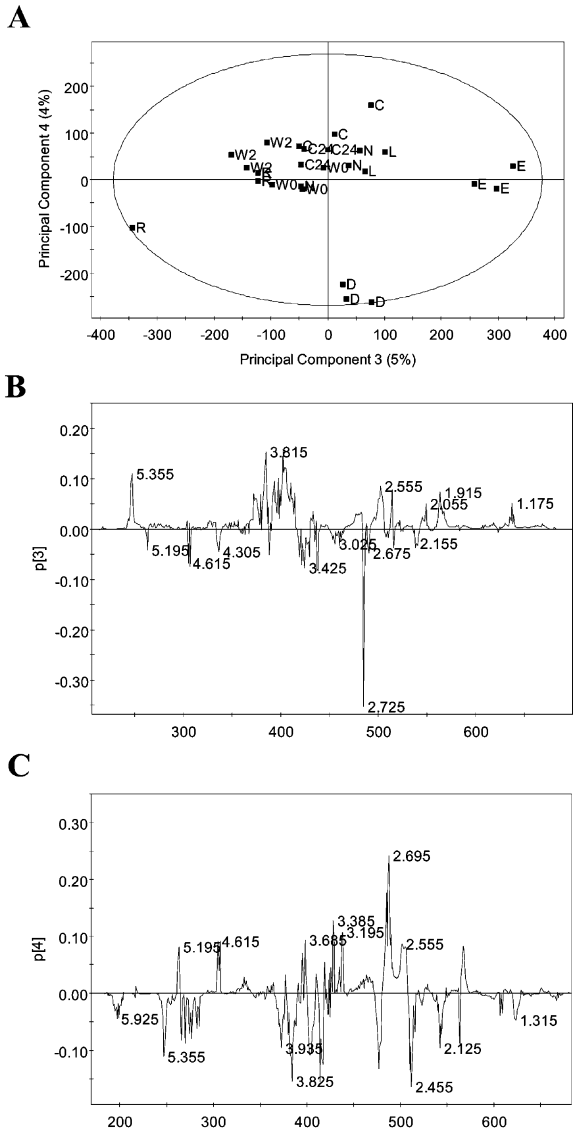

Fig. 8. Scores and loadings plots generated from PCA of *Arabidopsis* ecotypes. A: scores plot of PC3 vs. PC4. B: loadings plot of PC3. C: loadings plot of PC4. W0 = WS-0, W2 = WS-2, E = Estland, D = Dijon, R = Rschew, N = Nossen, C = Columbia-0, L = Landsberg, C24 = C24.

PC3 and PC4 accounted for the next 9% of variability within the sample set and demonstrated a separation of Estland and Dijon from the rest of the group (Fig. 8). Estland was separated by virtue of PC3. Examination of the loadings plot for PC3 (Fig. 8B) shows positive loadings for some (non-glucose) signals in the carbohydrate region. Examination of the original NMR spectrum and by comparison of standards in AMIX, this was identified as the disaccharide maltose. The ecotype Dijon separated from the rest of the group according to PC4. The loadings plot for PC4 (Fig. 8C) is more difficult to interpret since there are both positive and negative loadings. Since Dijon was found in the lower half of the scores plot we can infer that it is the negative loadings that are associated with Dijon. Examination of the contribution plots for Estland and Dijon (Fig. 9) revealed approximate NMR spectra corresponding to increased metabolites that these two ecotypes possess over the rest of the group. New signals in the ^{1}H NMR spectrum for Dijon were observed, and through the analysis of the contribution plot for ''Dijon minus average'', clues to the identity of this compound(s) are revealed. The compound appears to be olefinic or contain an unsaturated heterocyclic ring. So far the comparison of Dijon with metabolite standards and further investigation by two-dimensional NMR, and GC–MS, has yet to reveal the identity of this metabolite. On the other hand the contribution plot of Estland confirmed the presence of elevated levels of maltose and several amino acids, including lysine (δ 1.92).

Examination of the higher PCs (PC5–PC9) highlighted further differences in the sample set (data not shown). For example, PC5 vs. PC6 separated WS-2, and analysis of PC7 vs. PC8 separated the Landsberg ecotype. Differences in amino acids such as valine, isoleucine and threonine could be detected by examination of these PCs and further inspection of the original NMR spectra confirmed these minor differences.

2.5. Reproducibility in the method

It can be seen from the scores plot that the experimental variability is acceptable since tight distinct clusters form corresponding to each ecotype. The whole procedure, from extraction to data analysis, was repeated on the same set of freeze-dried plant samples in

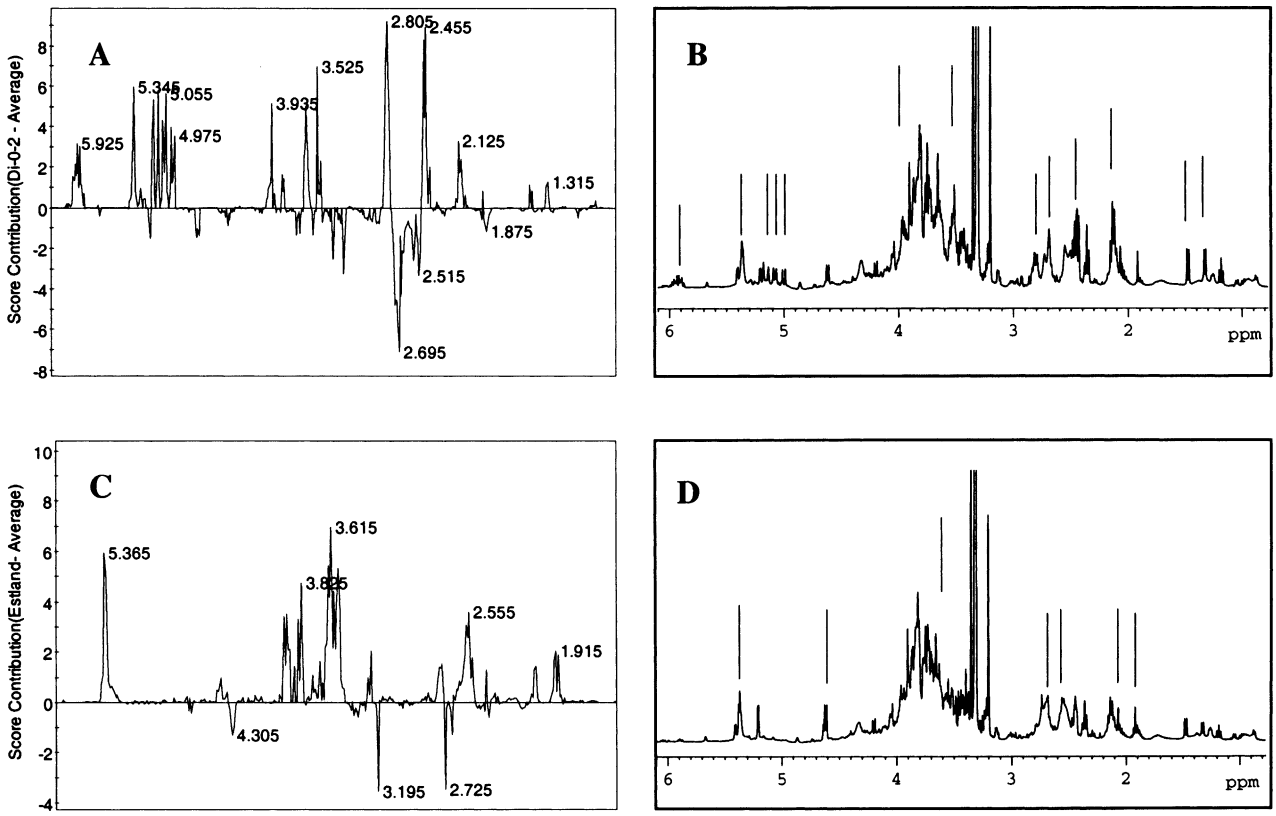

Fig. 9. Contribution plots generated from PCA of *Arabidopsis* ecotypes. A: contribution plot of Dijon *minus* average. B: ^{1}H NMR spectrum of Dijon. C: contribution plot of Estland *minus* average. D: ^{1}H NMR spectrum of Estland.

order to determine the reproducibility of the method as a whole. When the data were modelled using PCA in the same way, clustering of each ecotype was observed in a similar fashion to that seen previously. There was no separation of the individual clusters, again indicating the method was reproducible. Relative standard deviations were calculated for each observation of each ecotype, in the ^{1}H NMR spectrum. The mean of these deviations was $12\pm4\%$.

The earlier experiments utilized aliquots of combined material from trays of plants, all grown at the same time in controlled environment. In these experiments plant to plant biological variability was not assessed. However, examination of extracts from single Landsberg plants by the method above (data not shown) indicated that plant-to-plant variability was quite large. The mean relative standard deviation was $52\pm7\%$. In a another experiment aliquots of combined freeze-dried tissue from replicate trays were analysed. In this case the mean relative standard deviation was calculated to be $28\pm3\%$. These results indicate that pooling of plants grown together can reduce differences in data sets due to biological variability. However variability due to effects such as position of trays in the growth chamber is still significant.

3. Conclusions

^{1}H NMR spectroscopy has proved to be a valuable tool for unbiased metabolite fingerprinting of *Arabidopsis*. Principal component analysis highlighted genuine differences between ecotypes with loadings plots giving clues as to the nature of these differences. Comparison of the spectra of highlighted ecotypes with a library of NMR spectra of standards run under identical conditions, in AMIX, allowed us to identify compounds responsible for differences between spectra of different ecotypes. Differences could be detected in both the carbohydrate region and the aliphatic region, with sugars, organic acids and amino acids contributing to the differences in the sample set. The work has demonstrated how ^{1}H NMR analysis may be used in the future as a first pass screen to rapidly determine and characterize differences in molecular composition of plant samples. The technique serves as a rapid fingerprinting method that compares favourably with FT-IR (Goodacre and Anklam, 2001) with respect to reproducibility and extent of metabolome coverage. NMR, however, has the advantage over FT-IR in that the identities of many of the major metabolites can be deduced from the spectra. Coupled-MS techniques have advantages both

in terms of numbers of metabolites that can be quantified and the dynamic range of the concentrations that can be measured, but suffer from the disadvantage that chromatography selects subsets of the total metabolites. An integrated approach where differences are thrown up by NMR screening and then further investigated and accurately quantified by more targeted (chromatography-linked) methods such as GC–MS with appropriate internal standards seems to be a reasonable way forward to initiate high-throughput screens of plants. In this respect we foresee many uses of the NMR technique described, from large-scale analysis of natural variation, through mutant collections to transgenic plants.

4. Experimental

4.1. Plant material

Arabidopsis thaliana seeds were obtained from Nottingham *Arabidopsis* Seed Centre (NASC) and were germinated on agar containing Gamberg's B-5 basal medium containing 3% sucrose at 22 °C in continuous light. Plants were transferred to soil at the 2–4 leaf stage and grown in a controlled environment under long day (16 h) conditions, at a temperature of 23 °C and 75% humidity during the day and 18 °C and 80% humidity at night. Plants were harvested at growth stage 6.1–6.5 (Boyes et al., 2001) and immediately plunged into liquid nitrogen before freeze drying and grinding to a fine powder in a pestle and mortar. Samples were then stored until required at −80 °C.

4.2. Extraction and ^1H NMR spectroscopy

Freeze-dried plant material (15 mg) was weighed into an autoclaved 2 ml Eppendorf tube. D_2O:CD_3OD (1 ml, 80:20) containing 0.05% w/v TSP-d_4 (sodium salt of trimethylsilylpropionic acid) was added to each sample. The contents of the tube were mixed thoroughly and then heated at 50 °C in a water bath for 10 min. After cooling, the samples were spun down in a micro-centrifuge for 5 min. Of the supernatant 750 μl were added to a 5 mm NMR tube. All spectra were acquired under automation at a temperature of 300 K on a Bruker Avance spectrometer operating at 399.752 MHz ^1H observation frequency using the multinuclear broadband BBO 5 mm probe, and a water suppression pulse sequence with a relaxation delay of 5 s. Each spectrum consisted of 2048 scans of 32 k data points with a spectral width of 4845 Hz. The spectra were automatically Fourier transformed using an exponential window with a line broadening value of 0.5 Hz, phased and baseline corrected within the automation programme. ^1H NMR chemical shifts in the spectra were referenced to TSP-d_4 at δ 0.00.

4.3. Data reduction of the NMR spectra and multivariate analysis

The ^1H NMR spectra were automatically reduced to ASCII files using AMIX (Analysis of MIXtures software v.3.0, Bruker Biospin). Spectral intensities were scaled to TSP-d_4 and reduced to integrated regions or "buckets" of equal width (0.01 ppm) corresponding to the region of δ 9.0 to δ −0.5. The regions between δ 4.90 and δ 4.76 were removed prior to statistical analyses thus eliminating any variability in suppression of the water sample. The residual proton signals corresponding to methanol–d_4 (δ 3.365–3.285) and TSP–d_4 (δ 0.00) were also removed at this stage. The generated ASCII file was imported into Microsoft EXCEL for the addition of labels and then imported into SIMCA-P 9.0 (Umetrics, Umea, Sweden) for PCA analysis.

Acknowledgements

IACR-Long Ashton Research Station receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the United Kingdom. We thank the BBSRC-GARNet project (http://www.york.ac.uk/res/garnet/garnet.html) for support (C.H. and J.L.).

References

Antti, H., Bollard, M.E., Ebbels, T., Keun, H., Lindon, J.C., Nicholson, J.K., Holmes, E., 2002. Batch statistical processing of ^1H NMR-derived urinary spectral data. J. Chemometrics 16, 461–468.

Beckwith-Hall, B.M., Holmes, E., Lindon, J.C., Gounarides, J., Vickers, A., Shapiro, M., Nicholoson, J.K., 2002. NMR-based metabonomic studies on the biochemical effects of commonly used drug carrier vehicles in the rat. Chem. Res. Toxicol. 15, 1136–1141.

Boyes, D.C., Zayed, A.M., Ascenzi, R., McCaskill, A.J., Hoffman, N.E., Davis, K.R., Görlach, J., 2001. Growth stage-based phenotypic analysis of *Arabidopsis*: a model for high throughput functional genomics in plants. Plant Cell 13, 1499–1510.

Fan, T.W.-M., Lane, A., Shenker, M., Bartley, J.P., Crowley, D., Higashi, R.M., 2001. Comprehensive chemical profiling of gramineous plant root exudates using high-resolution NMR and MS. Phytochemistry 57, 209–221.

Fiehn, O., Altmann, T., Trethewey, R.N., Willmitzer, L., 2000. Metabolite profiling for plant functional genomics. Nature Biotech. 18, 1157–1161.

Fiehn, O., 2002. Metabolomics: the link between genotypes and phenotypes. Plant Mol. Biol. 48, 155–171.

Goodacre, R., Anklam, E., 2001. Fourier transform infrared spectroscopy and chemometrics as a tool for the rapid detection of other vegetable fats mixed in cocoa butter. J. Am. Oil Chem. Soc. 78, 993–1000.

Hall, R., Beale, M.H., Fiehn, O., Hardy, N., Sumner, L., Bino, R., 2002. Plant metabolomics: the missing link in functional genomics strategies. Plant Cell 14, 1437–1440.

Holmes, E., Nicholl, A.W., Lindon, J.C., Connor, S.C., Conollelly, J.C., Haseldon, J.N., Damment, S.J.P., Spraul, M., Neidig, P.,

Nicholson, J.K., 2000. Chemometric methods for toxicity classification based on NMR spectra of biofluids. Chem. Res. Toxicol. 13, 471–478.

Massart, D.L., Vandeginste, B.G.M., Deming, S.N., Michotte, Y., Kauffman, L., 1988. Chemometrics: A Textbook. Elsevier, New York.

Nicholson, J.K., Wilson, I.D., 1989. High-resolution proton magnetic resonance spectroscopy of biological fluids. Prog. Nucl. Magn. Reson. Spectrosc. 21, 449–501.

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

The *Arabidopsis* Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408, 796–815.

Trethewey, R.N., 2001. Gene discovery via metabolite profiling. Curr. Opin. Biotechnol. 12, 135–138.

Tweeddale, H., Notley-McRobb, L., Ferenci, T., 1998. Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool (''metabolome'') analysis. J. Bacteriol. 180, 5109–5116.

Wixon, J., 2001. Featured organism: *Arabidopsis thaliana*. Comp. Funct. Genom. 2, 91–98.