Unsupervised Multivariate Analyses for NMR-based Metabolic Profiling

Product used: Nuclear Magnetic Resonance (NMR)

Recently, NMR metabolic profiling (NMR-MP) techniques, which are useful for systematically analyzing metabolites in an organism (metabolome), have been widely used in post-genome studies and quality controls for agricultural and biotechnological products. NMR-MP is characterized as a non-target metabolome analysis. It is difficult to interpret results directly from data in non-target analysis; therefore, multivariate analyses are applied to explore characteristic information from multivariate data. This is referred to as “data-mining”.

It is important to choose appropriate multivariate analyses that correspond to the goal of your study. A few example analyses are dimension reduction, discrimination analysis, regression, modeling, and prediction. In this note, NMR-MP with principal component analysis (PCA) and hierarchical cluster analysis (HCA) that are unsupervised multivariate analyses* are introduced.

Materials and Methods

Polar metabolites in six commercially available sprouts (radish, broccoli, mustard, alfalfa, pea, and okra) were extracted by a modified Bligh-Dyer method. The extracts were then dried with a centrifugal evaporator. Dried metabolite mixtures were resuspended with the NMR solvent (a 100 mM deuterated potassium phosphate buffer, pH = 7.0). ^{1}H homonuclear 2D *J*-resolved spectra were recorded and their projections were used for multivariate analyses (**Figure 1**). There were less spectral overlapping in X (F_2) projections of the 2D *J*-resolved spectra than conventional ^{1}H single pulse spectra because there were no splitting from J_{HH} [1]. Three sample replicates were performed.

Bucket integrations were performed on the series of spectra (0.02 ppm width each, between 0.5-9.0 ppm). At the time, variables integrals of which are equivalent to those of noise were removed before multivariate analyses. As a result, the matrix (multivariate for multivariate analyses) consisted of 100 variables * 18 samples were obtained. PCA and HCA were performed with free statistical functions in “R”.

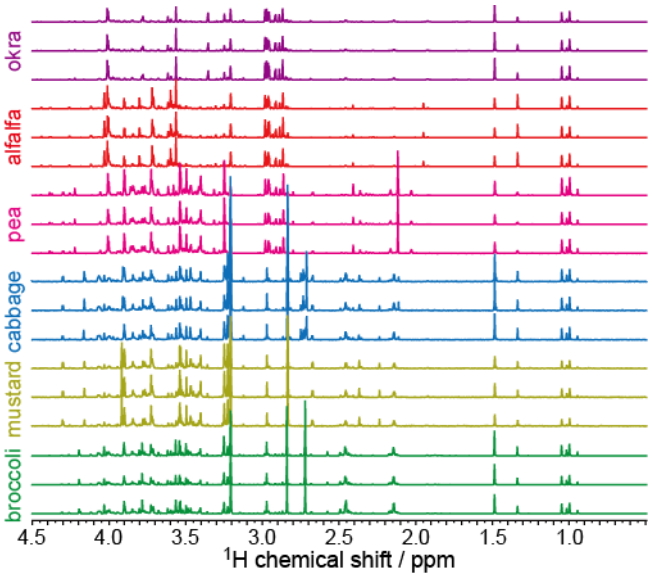

Figure 1. Series of ^{1}H spectra of polar metabolites in sprouts. X (F_2) projection of 2D *J*-resolved spectra were used. NMR spectra were recorded continuously on a JNM-ECZ400S spectrometer equipped with a ROYAL probe and ASC30 auto sample changer.

Applying PCA

Principal components (PCs) are linear combinations of each variable in a data set that result in a maximizing of the variances. The result of this process is that almost all data can be explained using far less dimensions of data. Formula 1 shows a relation between scores, loadings, and original variables. Normally, a centering and scaling of each variable are performed before PCA.

original variables (centering and scaling)

$$t_{ij} = \overline{x}_{i1} w_{j1} + \overline{x}_{i2} w_{j2} + \cdots + \overline{x}_{in} w_{jn} \quad (1)$$

score loadings

Here, subscripts *i* represents *i*th sample, and *j* represents *j*th variable or PC. We can interpret interesting information (ex. characteristic metabolites in particular sample group) from scores and loadings. In this case, eigenvectors were used as loadings.

The score and loading plots in PCA are shown in **Figure 2**. Brassicaceous sprouts including radish, broccoli, and mustard had negative scores in PC1, and others had positive scores (**Figure 2(a)**). In loading plot, asparagine had positive value, on the other hand, glutamine had negative (**Figure 2(b)**).

It is known that amino acids which have an amino group on their side chain such as asparagine and glutamine have an important role in nitrogen metabolism in young plants [2]. The ratio of these amino acids are different by plant species and growing environment. For example, it was reported that brassicaceous and leguminous plants used glutamine and asparagine as the major amino acid has amino group on side chain, respectively [3]. It was also reported that plants synthesize asparagine more than glutamine when they cannot photosynthesize because C/N ratio is smaller in the former than the latter [4].

Applying HCA

HCA is able to summarize relations between each sample and also variable as a dendrogram. Dendrograms are generated according to distance between data and clusters. HCA with heat map visually shows relations between samples and variables (*i.e.* metabolites). While Euclidean distance is one of the most common definition of distance (formula 2), correlation distance (formula 3) which is based on Pearson's correlation coefficient is used in the analysis in this note.

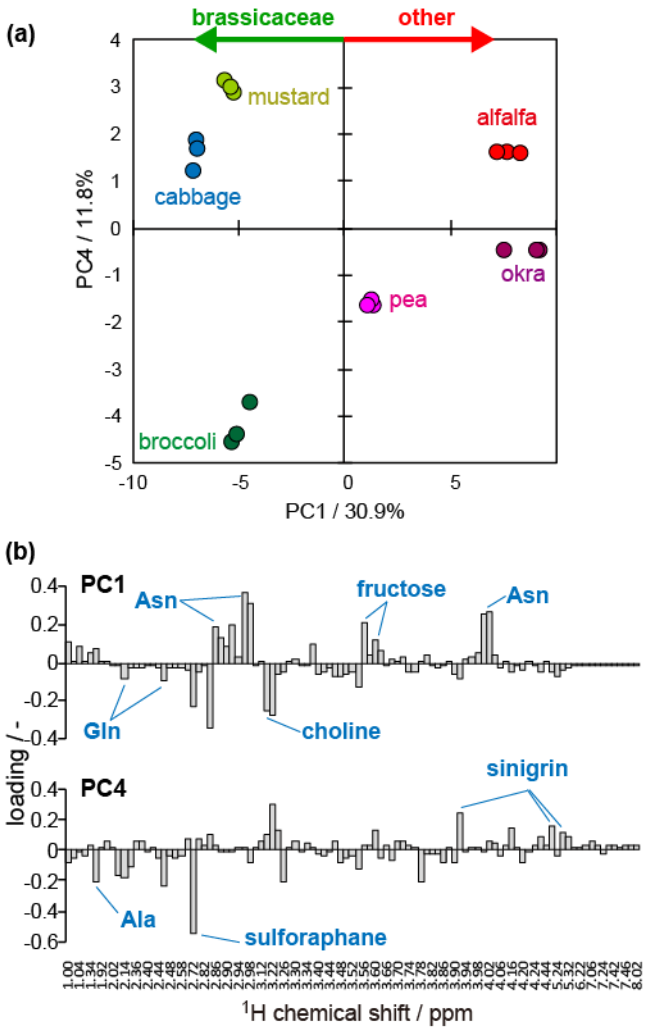

Figure 2. PCA of polar metabolites in sprouts. (a) score plot. (b) loading plot. PCA was performed after the matrix was normalized by the constant sum method and centering was performed.

$$dist_{\text{euclidean}}(A,B) = \left[\sum_{i=1}^{n}(a_i - b_i)^2\right]^{1/2} \qquad (2)$$

$$dist_{\text{corr}}(A,B) = (1 - corr(A,B))/2 \qquad (3)$$

Here, a_i and b_i are *i* th variable in A and B, respectively, and *corr*(A,B) is Pearson's correlation coefficient.

HCA with the heat map is shown in **Figure 3**. Columns and rows in the heat map represent variables and samples, respectively. Dendrograms on left- and upper-side show distance between variables and samples hierarchically. In the dendrogram of sample, brassicaceous sprouts were discriminated from the others. In the dendrogram of variable, variables from the same metabolite and variables from different metabolites which had similar behavior built clusters. The heat map was replaced according to dengrograms; Thus, we can visualize which metabolites were present in a particular sample group. For example, sulforaphane, which is an isothiocyanate, is concentrated more heavily in broccoli, and sinigrin, which is a glucosinolate, is found in higher concentrations in mustard [5].

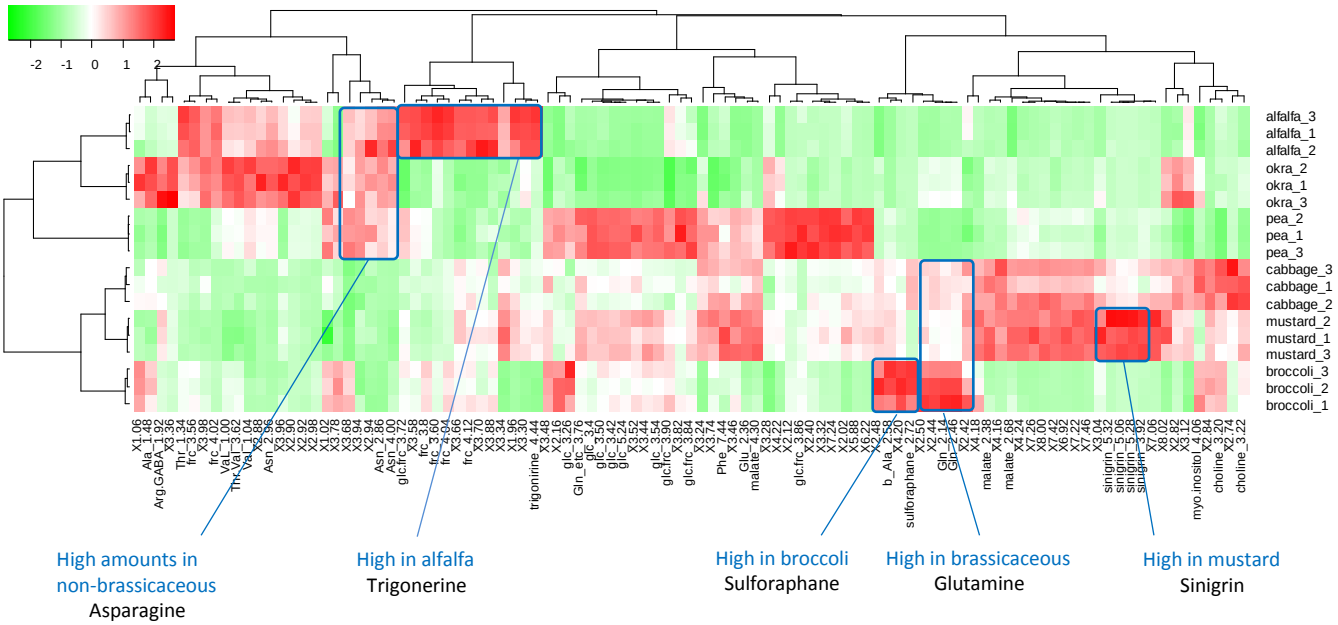

Figure 3. HCA with heat map of ^{1}H NMR spectra of polar metabolites in sprouts. Columns and rows of the heat map represent variables and samples, respectively. Correlation distance and complete linkage method is used to build dendrograms. Auto-scaled variables are used for the heat map. Therefore metabolites with relatively high and low concentrations in the sample are shown as contrasting density of red and green, respectively. The map visualizes which metabolites are a lot or a little in each sample or sample cluster. glc: glucose, frc: frctose, amino acids are shown as 3 letter abbreviations.

NMR-MP is fruitful because it has excellent experimental reproducibility and doesn't require sample separation. Metabolome analyses by mass spectrometry with chromatography require careful selections of appropriate solid phase, mobile phase, ionization method, extraction method, and considerations of "carry-over". The total number of experimental considerations are considerably less in NMR-MP. If specificity and sensitivity is important in the analysis, mass spectrometry with various chromatography can prove to be superior methods. However, NMR-MP is useful as a fast and first screening in terms of data-driven science especially if you have no hypotheses on your dataset.

* Multivariate analyses do not use any hypotheses are referred to as "unsupervised analysis". On the other hand, multivariate analyses use hypotheses (=use external variables as objectives) are referred to as "supervised analysis".

References

[1] Application note "*Simplifying ^{1}H-NMR Spectra in NMR Metabolic Profiling*" (NM170014).
[2] Komatsu; T. et al. *Metabolites* (2014) **4**, 1018.
[3] Nord; F. F. and Street; H. E., *"Nitrogen Metabolism of Higher Plants"* in Advances in Enzymology (1949) **9**, John Wiley & Sons, Inc, 391.
[4] McGrath; R. B. et al., *Plant J.* (1991) **1,** 275.
[5] Rangkadilok; N. et al., *Sci. Hort.* (2002) **96**, 11.

JEOL Ltd.
3-1-2 Musashino Akishima Tokyo 196-8558 Japan Sales Division Tel. +81-3-6262-3560 Fax. +81-3-6262-3577
www.jeol.com ISO 9001 • ISO 14001 Certified

• **AUSTRALIA & NEW ZEALAND** /JEOL(AUSTRALASIA) Pty.Ltd. Suite 1, L2 18 Aquatic Drive - Frenchs Forest NSW 2086 Australia • **BELGIUM** /JEOL (EUROPE) B.V. Planet II, Gebouw B Leuvensesteenweg 542, B-1930 Zaventem Belgium • **BRAZIL** /JEOL Brasil Instrumentos Cientificos Ltda. Av. Jabaquara, 2958 5° andar conjunto 52 ; 04046-500 Sao Paulo, SP Brazil • **CANADA** /JEOL CANADA, INC. 3275 1ere Rue, Local #8 St-Hubert, QC J3Y-8Y6, Canada • **CHINA** /JEOL(BEIJING) CO., LTD. Zhongkeziyuan Building South Tower 2F, Zhongguancun Nansanjie Street No. 6, Haidian District, Beijing, P.R.China • **EGYPT** /JEOL SERVICE BUREAU 3rd Fl. Nile Center Bldg., Nawal Street, Dokki, (Cairo), Egypt • **FRANCE** /JEOL (EUROPE) SAS Espace Claude Monet, 1 Allee de Giverny 78290, Croissy-sur-Seine, France • **GERMANY** /JEOL (GERMANY) GmbH Gute Aenger 30 85356 Freising, Germany • **GREAT BRITAIN & IRELAND** /JEOL (U.K.) LTD. JEOL House, Silver Court, Watchmead, Welwyn Garden City, Herts AL7 1LT, U.K. • **ITALY** /JEOL (ITALIA) S.p.A. Palazzo Pacinotti - Milano 3 City, Via Ludovico il Moro, 6/A 20080 Basiglio(MI) Italy • **KOREA** /JEOL KOREA LTD. Dongwoo Bldg. 7F, 1443, Yangjae Daero, Gangdong-Gu, Seoul, 05355, Korea • **MALAYSIA** /JEOL(MALAYSIA) SDN.BHD. 508, Block A, Level 5, Kelana Business Center, 97, Jalan SS 7/2, Kelana Jaya, 47301 Petaling Jaya, Selangor, Malaysia • **MEXICO** /JEOL DE MEXICO S.A. DE C.V. Arkansas 11 Piso 2 Colonia Napoles Delegacion Benito Juarez, C.P. 03810 Mexico D.F., Mexico • **RUSSIA** /JEOL (RUS) LLC. Krasnoproletarskaya Street, 16, Bld. 2, 127473, Moscow, Russian Federation • **SCANDINAVIA** /SWEDEN JEOL (Nordic) AB Hammarbacken 6A, Box 716, 191 27 Sollentuna Sweden • **SINGAPORE** /JEOL ASIA PTE.LTD. 2 Corporation Road #01-12 Corporation Place Singapore 618494 • **TAIWAN** /JIE DONG CO., LTD. 7F, 112, Chung Hsiao East Road, Section 1, Taipei, Taiwan 10023 (R.O.C.) • **THE NETHERLANDS** /JEOL (EUROPE) B.V. Lireweg 4, NL-2153 PH Nieuw-Vennep, The Netherlands • **USA** /JEOL USA, INC. 11 Dearborn Road, Peabody, MA 01960, U.S.A.