

Supporting Information

A novel strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS

Pär Jonsson,^{‡1} Jonas Gullberg,^{†1} Anders Nordström,[†] Miyako Kusano,[†] Mariusz Kowalszyk,[†] Michael Sjöström,[‡] and Thomas Moritz^{*†}

[†]Umeå Plant Science Center, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, SE-901 87 Umeå, Sweden

[‡]Research Group for Chemometrics, Organic chemistry, Department of Chemistry, Umeå University, SE-901 87 Umeå, Sweden

Corresponding author. E-mail: thomas.moritz@genfys.slu.se

¹P.J and J.G. contributed equally to the work

ALIGNMENT

Alignment is done by finding the maximum covariance between the Total Ion Current Chromatograms (TIC).

- 1) Set one sample as master. The shift for this sample is set to zero.
- 2) Set the maximum shift
- 3) Select sample to analyse.
- 4) Set shift to minus maximum shift
- 5) Calculate covariance between the master TIC and the selected samples TIC.
- 6) Set shift to shift +1 scan number. If shift < max shift go to 5
- 7) Locate the maximum covariance. The shift corresponding to the maximum covariance is the shift to use.
- 8) If more samples to align go to 3.
- 9) Use all calculated shifts to align the samples.

Linear interpolation

For each sample a straight line is made for each m/z from one zero concentration region to the next this line is subtracted from the signal, negative values is set to zero.

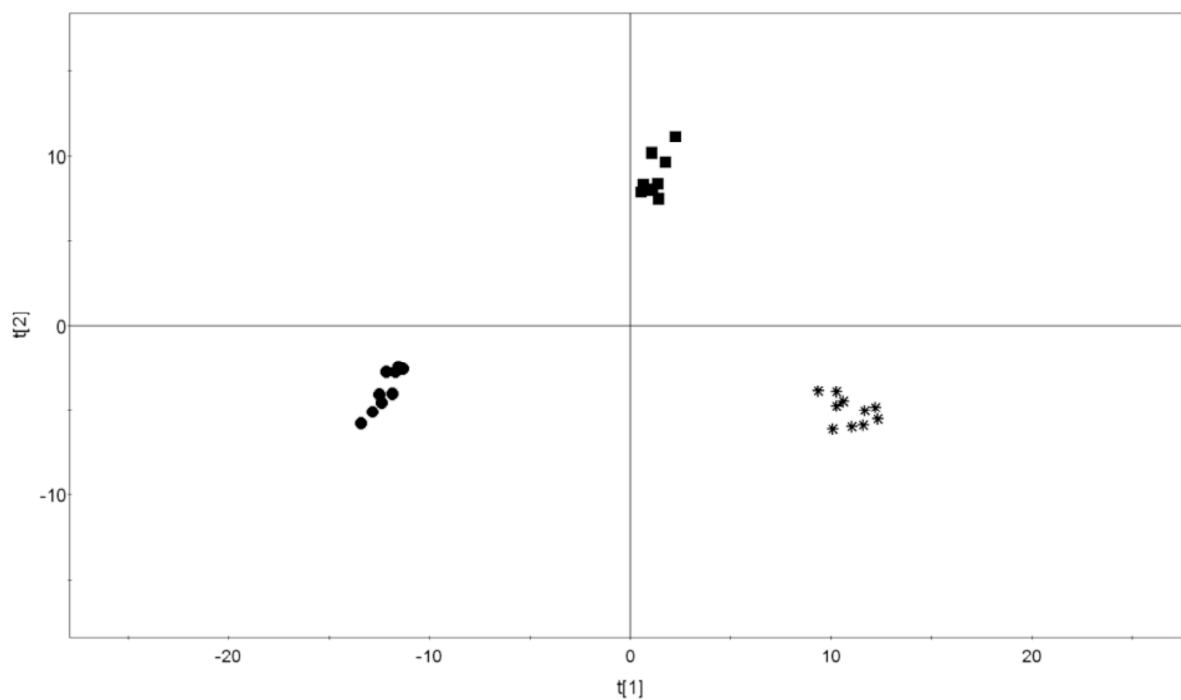
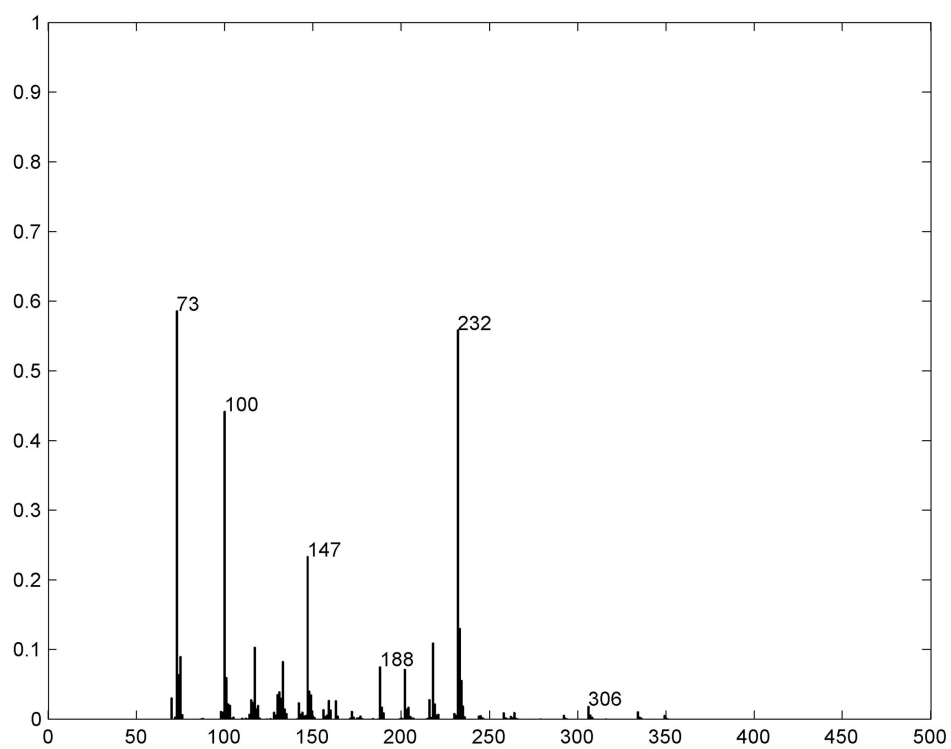
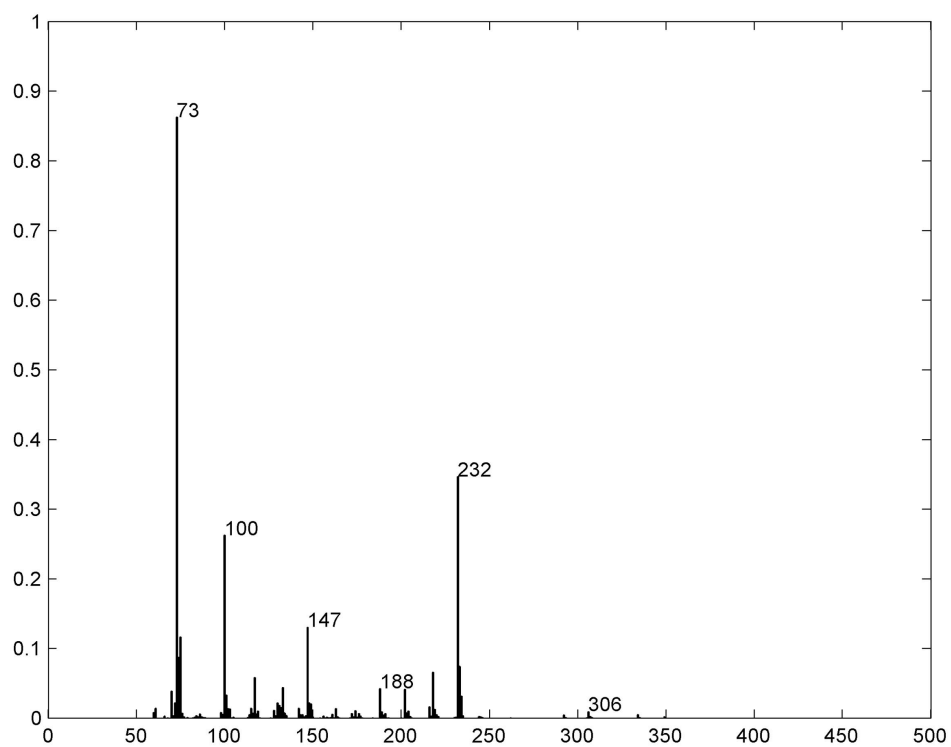


Figure S1 PCA-score plot from non-processed MS-files derived by described method, with data down-loaded from www.noble.org/plantbio/MS/downloads.html. The data-set is from the GC/MS analysis of *Medicago truncatula* root, leaf and stem tissue. The PCA-analysis did not take longer than 2.5 h to generate. We acknowledge the Sumner group at Noble Foundation for the data.



A.



B.

Figure S2 (A) An Alternating Regression (AR) mass profile from one significant component in a time window that shows difference between LD_{0/2} and SD₆ samples. (B) Deconvoluted mass spectrum from corresponding time window from where the AR mass profile was obtained. The compound was identified as aspartic acid (methoxime-TMS derivative). The significance in levels of aspartic acid between LD and SD samples was further tested by student t-test.