

High quality plant metabolomics by GC-TOF mass spectrometry and automated database annotation: 1000 runs * 500 metabolites

Oliver Fiehn, Tobias Kind, Gert Wohlgemuth
UC Davis Genome Center, Davis, CA

Overview

Purpose

Published metabolomic studies so far have involved less than 200 individual samples. Such surveys are accomplished within a week without too much drift of instrument sensitivity or retention time shifts. We show how quality control charts and standard operating procedures maintain high data quality even in larger programs. Equally important are the detection, alignment and annotation of individual deconvoluted peaks for over 1,000 chromatograms. If each deconvoluted 'signal' is taken into consideration, lists of target peaks expand to thousands of entries. We here demonstrate how severe constraints on mass spectral purity, s/n ratios and detection frequency manage unbiased metabolomics data in large scale experiments.

Methods

A Leco corp. Pegasus III GC-TOF was equipped with a dual arm CTC robotic derivatization station and an ATAS automatic liner exchange and direct thermodesorption injection device. Quality control was achieved by seven standard operating procedures. Daily calibration series of reference compound mixtures, reagent and method blank controls and sample master mixtures allowed monitoring of upper and lower intervention limits in quality control charts. Deconvoluted mass spectra were fed into the self-programmed data base BinBase for restricting the final data set to highly valid annotated metabolite peaks.

Results

A novel method for high quality metabolite profiling by automated derivatization and liner exchange is presented. This procedure enables the analysis of unfractionated plant leaf extracts with a throughput of 600 samples per month and instrument. Importantly, no carry over effects were observed for lipophilic compounds such as free fatty acids. Up to 1,000 peaks were detected from *Arabidopsis thaliana* leaf extracts, however, the actual number and identity of peaks varies between samples of different origin. Use of mass spectral metadata like 'peak purity', 'signal/noise' and 'unique ion' filtered noisy spectra to a set of 500 unique and valid metabolites. A comparison of light, cold, and nitrogen perturbations from >1,000 samples of *Arabidopsis* ecotypes and F1 crosses highlighted the importance of raffinose biosynthesis and C/N balancing in response to environmental stress.

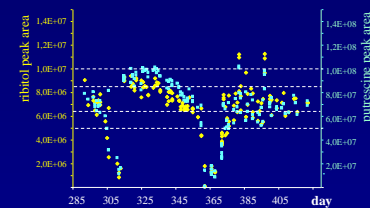


Setup of a GC/TOF instrument with automatic two-step derivatization (red ovens) and liner exchange. Vials are crimped with magnetic caps for transport.

Quality Control

Standard Operating Procedures (SOPs)

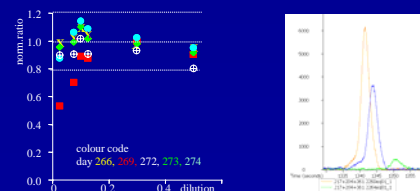
Seven SOPs have been implemented for routine metabolomics: (1) pipette calibration, (2) GCTOF operation, (3) plant tissue extraction, (4) blood plasma extraction, (5) quality control mix, (6) robotic derivatization and liner exchange for GCTOF, (7) LC-QIT mass spectrometry by HILIC and monolithic RP. A former SOP on 'injection sequence generation' has been replaced on June 01, 2005, by an online web form (<http://fiehnlab.ucdavis.edu>). For high throughput metabolomics, the most important SOP has been identified as the daily generation of s/sl calibration curves for a 37-compound quality control mixture. This mix serves for determining in-control and out-of-control situations with respect to instrument sensitivity drifts, injector discrimination for high and low boiling compounds, silylation efficiency for N-TMS and O-TMS derivatives, peak deformations and column aging.



Control of GC-TOF sensitivity and selectivity

Upper panel: Quality control chart for instrument sensitivity drift. Identical amounts of a QC mix are injected as daily routines. Two compounds, ribitol and putrescine, serve as control markers for absolute GCTOF performance. If lower intervention limits are violated (at day 300 and 360), sample analyses are stopped and maintenance measures are taken to bring the sensitivity back into control (e.g. by filament change).

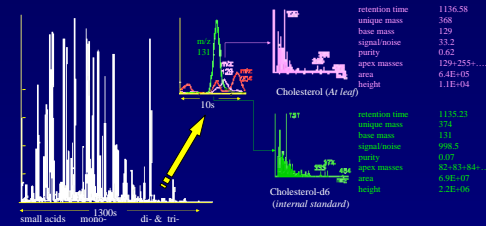
Lower panel, left: the instrument selectivity with respect to silylation of amines is controlled by analysis of the ratios of putrescine/ribitol. Problems may occur with improper liner qualities or catalytic spots in the injector body, e.g. at day 269. **Right panel:** abundance of high boiling maltotriose is a direct indicator for matrix deposition at the front of the separation column.



Database algorithms

Mass spectral metadata

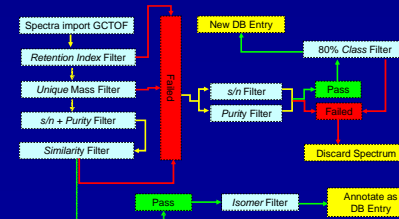
All chromatograms are subjected to mass spectral deconvolution for peak finding and purification. Each mass spectrum is qualified by metadata such as 'peak purity' or 'apexing masses' that are used within the BinBase database.



In unfractionated chromatograms of *Arabidopsis* leaf extracts, up to 1,000 peaks can be detected. Most of these peaks comprise genuine plant metabolites. However, for each chromatogram, the number and the identity of peaks will be different. Moreover, deconvolution errors may cause false positive peak detections or insufficient mass spectral purifications.

BinBase

In order to annotate all detected peaks and add novel compounds, a filtering and matching algorithm has been programmed: BinBase. After import, spectra are validated and matched against known compounds by retention index, unique masses and spectral similarity (with thresholds depending on peak purity and abundance). Isomers with MS similarity differences <100 are annotated by RI.



Peaks that do not match existing database entries may qualify as novel metabolites if they pass purity and abundance thresholds. In order to validate the metabolic origin of peaks, potential new BIN spectra must be detectable in 80% of all chromatograms of a *Class*. *Classes* are defined by the experimental design metadata by the users (biologists) within a separate LIMS module.

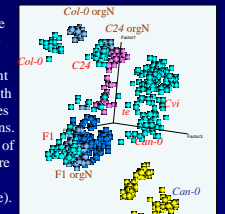
Application

Environmental impact on hybrid vigor in *A. thaliana* crosses



The F1 crosses between accessions of *Arabidopsis thaliana* show hybrid vigor after six days of development. The degree of this effect is also dependent on light intensity.

Genetic causes such as epistasis or overdominance may result in release of metabolic bottlenecks. To test this hypothesis, 1,000 samples were profiled. Col-0, C24 and F1 crosses were markedly different in carbon/nitrogen balancing under standard growth conditions (red labels). Other *Arabidopsis* ecotypes were clearly separable from the Col/C24 accessions. Under high light conditions (orange), metabolism of C24 and Col-0 self-crossed parental lines was more similar to the F1 progenies than under N-overfertilization (brown) or cold acclimation (blue).



Network regulation

Many metabolite levels show strong correlation with other metabolites. The investigation of corresponding comprehensive correlation networks revealed a far higher connectivity of metabolites such as fumarate, erythrose, galactinol and raffinose in F1 crosses than in parental lines. This finding points to differential control of carbon partitioning in biochemical networks.



Conclusions

Monitoring instrument performance and automated chromatogram evaluations enabled a high quality, high throughput study of differential regulation of metabolism in *Arabidopsis thaliana* crosses. Metabolic phenotypes enable interpreting the complex phenomenon of hybrid vigor, which at least partially seems due to differences in carbon partitioning.

Acknowledgments

Plant crosses: T. Altmann and R. Meyer, Univ. Potsdam, Germany. GC-TOF data evaluation: S. Gatzek, formerly MPI-MP, Potsdam, Germany. Cold acclimation experiments: D. Hinch, MPI-MP, Potsdam, Germany.