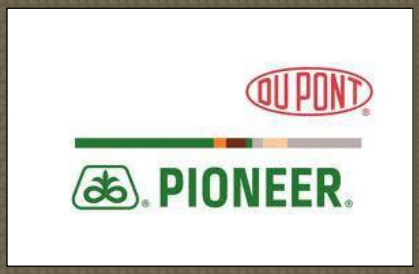
Differential Metabolomic Profiling of Maize Genotypes under Drought-Stressed Conditions using IROA® (Isotopic Ratio Outlier Analysis)

Chris Vlahakis1, Jan Hazebroek1, Felice de Jong2, Chris Beecher2

[1]*DuPont Pioneer, Johnston, Iowa;* [2] *IROA Technologies LLC, Ann Arbor, MI*



Application Note: IT - AG 001

Abstract

The IROA protocol has been applied in a phenotypic analysis of field grown maize (*Zea mays*) to understand the biochemical differences across selected genotypes when exposed to drought conditions. In this IROA phenotypic analysis, field-grown leaves containing carbon at natural abundance were compared to a standard maize leaf that was grown to contain universally-distributed ~97% ^{13}C; resulting in a targeted analysis using a biologically-relevant internal standard. At 97% ^{13}C the IROA patterns were sufficient to find isotopically labeled peaks, identify their ^{12}C isotopomers, and remove artifacts, noise and extraneous peaks using the IROA ClusterFinder software. With accurate mass and IROA, the identification of observed component peaks to chemical formula is unambiguous. The benefit of IROA is it takes into account variances introduced during sample-preparation and analysis, including ion suppression.

Introduction

Basic IROA Application

The greatest challenge of most metabolic profiling experiments is the ability to differentiate peaks of biological origin from artifactual peaks, and to accurately identify and quantitate the peaks of interest. The "Basic" IROA labeling protocol (see Figure 1), utilizes isotopically-defined media in which all nutrients are labeled with either 5% ^{13}C, "13C IROA media" (experimental), or 95%^{13}C, "13C IROA media" (control), so that all biological compounds carry a distinct molecular signature and molecules can be distinguished from each sample set, as they have differing masses. Control and experimental samples can be analyzed as a single composite sample by LC-MS with all biological peaks uniquely paired, and unlabeled natural abundance artifacts may be identified and discarded.

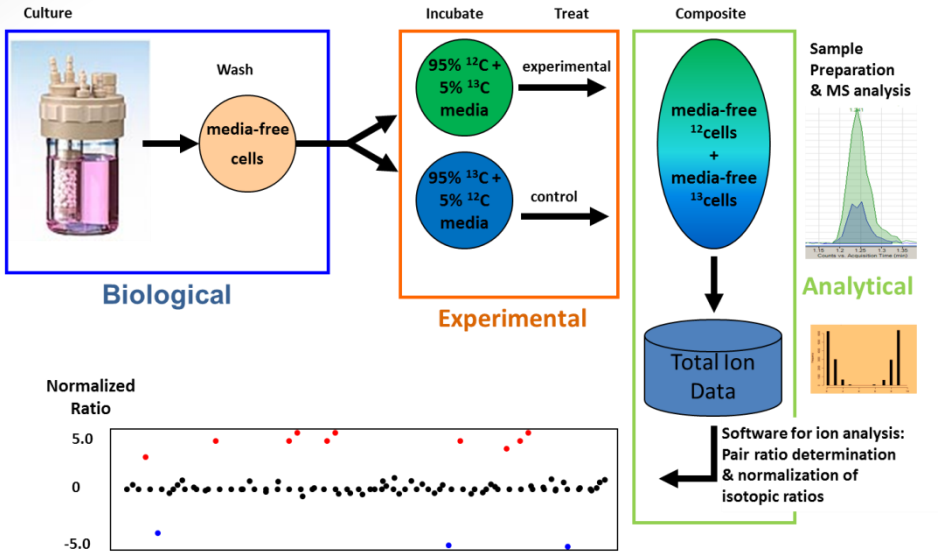


Figure 1. The Basic IROA Protocol Overview.

All biological compounds will have two paired peaks; the peak from the 12C-media is mirrored by a second peak from the 13C-media (Figure 2). The distance between these peaks readily identifies the number of carbons in the compound. In addition there are corresponding M+1 and M-1 peaks (and M+2 and M-2 etc. peaks) which are a mass difference of 1.00335 amu (the mass difference between a ^{12}C and ^{13}C isotope), giving the IROA peaks a characteristic U-shape “smile” pattern. Accurate mass together with the knowledge of the number of carbons in a molecule greatly facilitates metabolite identification.

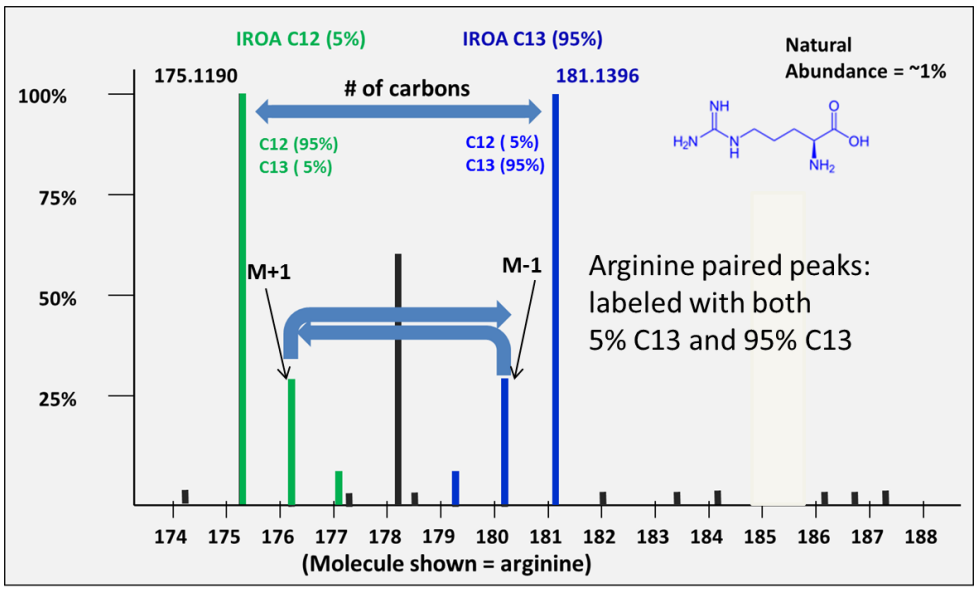


Figure 2. The IROA peaks. In the case of arginine, the ^{12}C M+ located at 175.1190 and its ^{13}C mate at 181.1396 clearly indicate a 6 carbon molecule. The corresponding M+1 and M-1 peaks are plus or minus the mass of a neutron. Natural abundance peaks do not have IROA patterns.

Phenotypic IROA Protocol

Where it is not possible to label the biological sample, the "Phenotypic" IROA Protocol (Figure 3) is applied. Here the sample is collected at natural abundance and mixed with a fully predefined "Standard" that has been isotopically labeled using IROA 13C media. An ideal Standard would be one that represented the entire metabolome of the sample under study. All peaks of the IROA-labeled Standard may be easily identified according to their characteristic M-1 peak. The natural abundance metabolite peaks (paired to each Standard peak) may be readily identified as their exact mass and position are established relative to the Standard. As the Standard that is used is always the same for a given sample, the compounds that are present in it have already been characterized. Artifacts have no match in the Standard and can be discarded. Whereas in a basic IROA dataset the ratio of the peak areas represents the relative deviation of the metabolic pool sizes brought about by the experimental condition, in a Phenotypic IROA experiment the overall pattern of deviations from the Standard will define phenotype by difference from the Standard.

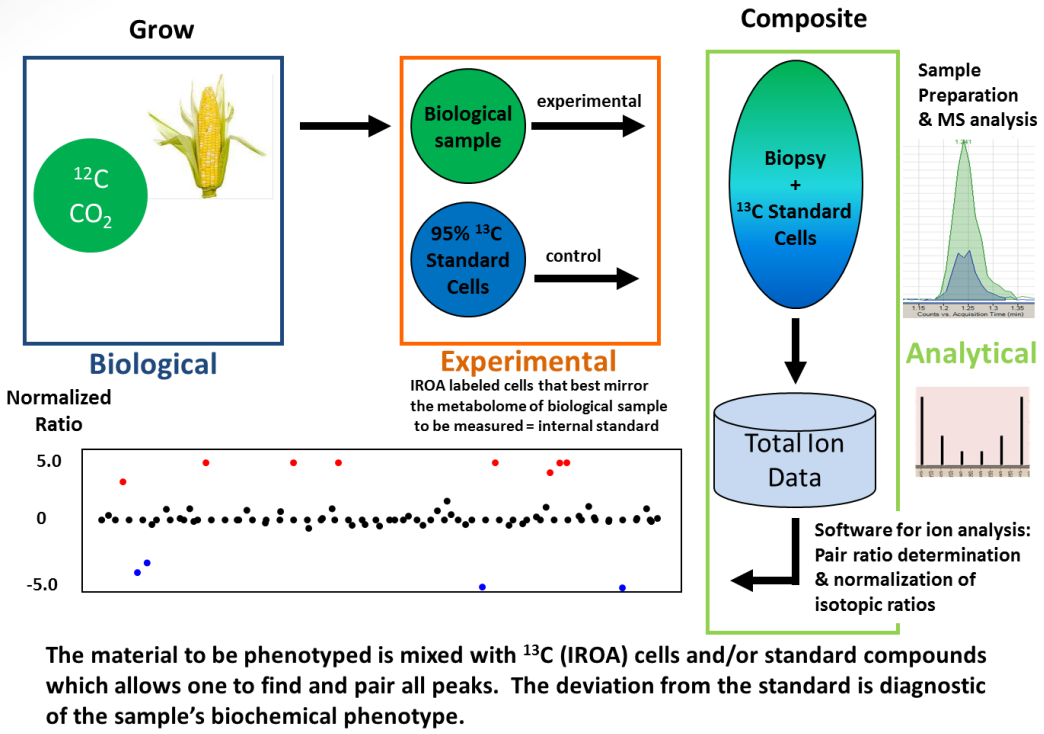


Figure 3. Phenotypic IROA protocol. Maize leaves are pooled with an isotopically-labeled maize standard (95-97% ^{13}C-enriched, IsoLife), processed, and analyzed by mass spectrometry. Software algorithms are used to sort the peaks, locate the M-1/^{13}C ratio of the labeled metabolites and match the experimental M natural abundance peaks. The ratio of the paired peak areas are measured and normalized and outliers determined. The Phenotypic experiment is considered a very complex targeted analysis.

Theory & Discussion

Using the Phenotypic IROA Protocol, both control and experimental samples are analyzed as a pooled sample so that variability and ion suppression are accounted for, and fewer samples are required to be analyzed. Since the control Standard is labeled and because even though the experimental

samples do not carry any isotopic labeling, their exact mass and position are established relative to the Standard and noise and artifacts can be removed by software algorithms allowing for a very dramatic reduction in data size. The number of carbons for each IROA peak and accurate mass for either of the monoisotopic peaks make it possible to determine the empirical formula, unambiguously for masses below 400. Representative peaks are shown in Figure 4.

In this experiment a total of 985 IROA peaks were identified over the 12 samples analyzed in both positive and negative mode. The 12 samples consisted of three groups of four samples; four of each well watered, and two different drought conditions.

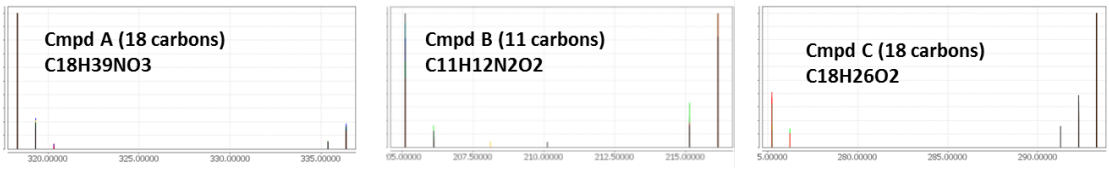
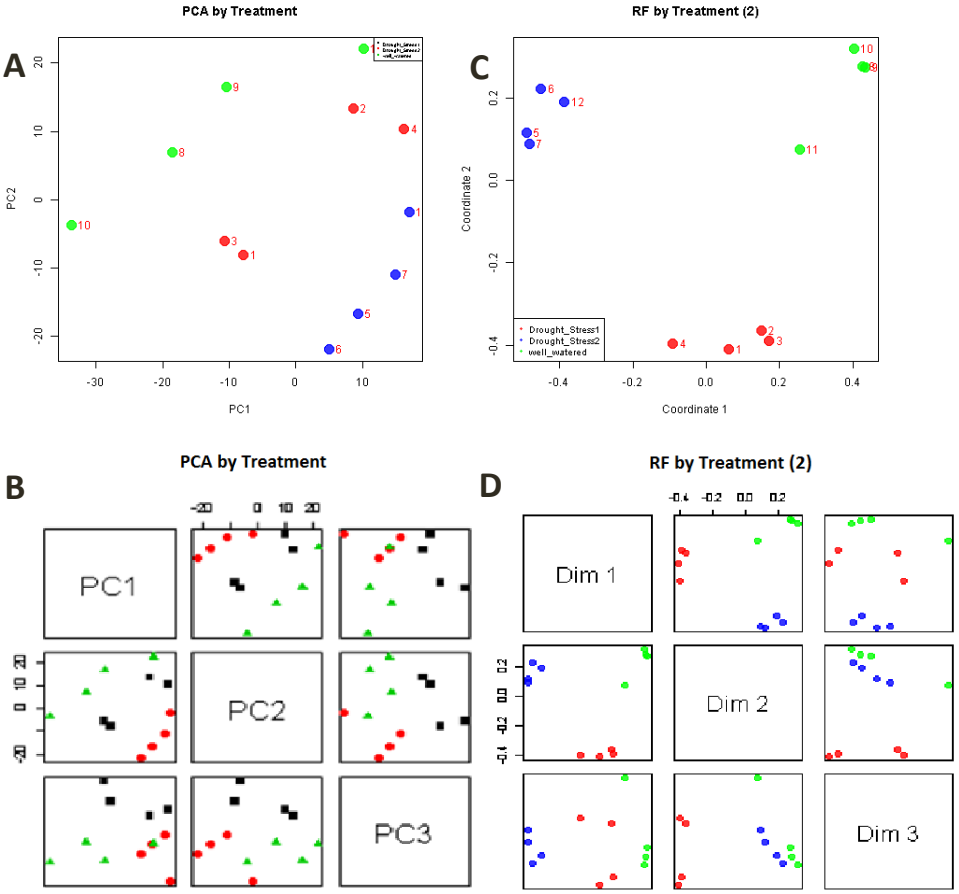


Figure 4. Representative IROA peaks. These peaks are representative of the 985 IROA peaks.

Results

This study describes the use of an analytical and bioinformatic metabolomics technology, "Phenotypic IROA Protocol" (Figure 3) applied to the understanding of trait profiles of genetically diverse DuPont Pioneer maize that were field grown under watered or one of two drought conditions. The IROA-labeled Standard used was maize leaves derived from 7 week old Zea *mays* seedlings grown using $^{13}CO_2$ as the only carbon source. 985 features were tentatively identified using IROA ClusterFinder™ software. The natural abundance compounds in the field grown plants were compared to the same compound abundance seen in the IROA standard. The resulting dataset was analyzed as discussed in the Methods. PCA (Figure 5 A & B), Random Forest (RF - C & D), and Non-negative Matrix Factorization (NMF – Figure 5 E & F) were all clearly able to separate the three classes of samples. A correlation

analysis was run on the 32 most significant compounds in the RF analysis (Figure 5 G) to group these compounds. As an example, metabolites selected by RF (Figure 5 H & J) demonstrated that drought had a strong effect on their biochemistry. At k= 6 NMF identified three distinct patterns associated with each of the three states (Figure 5 I).



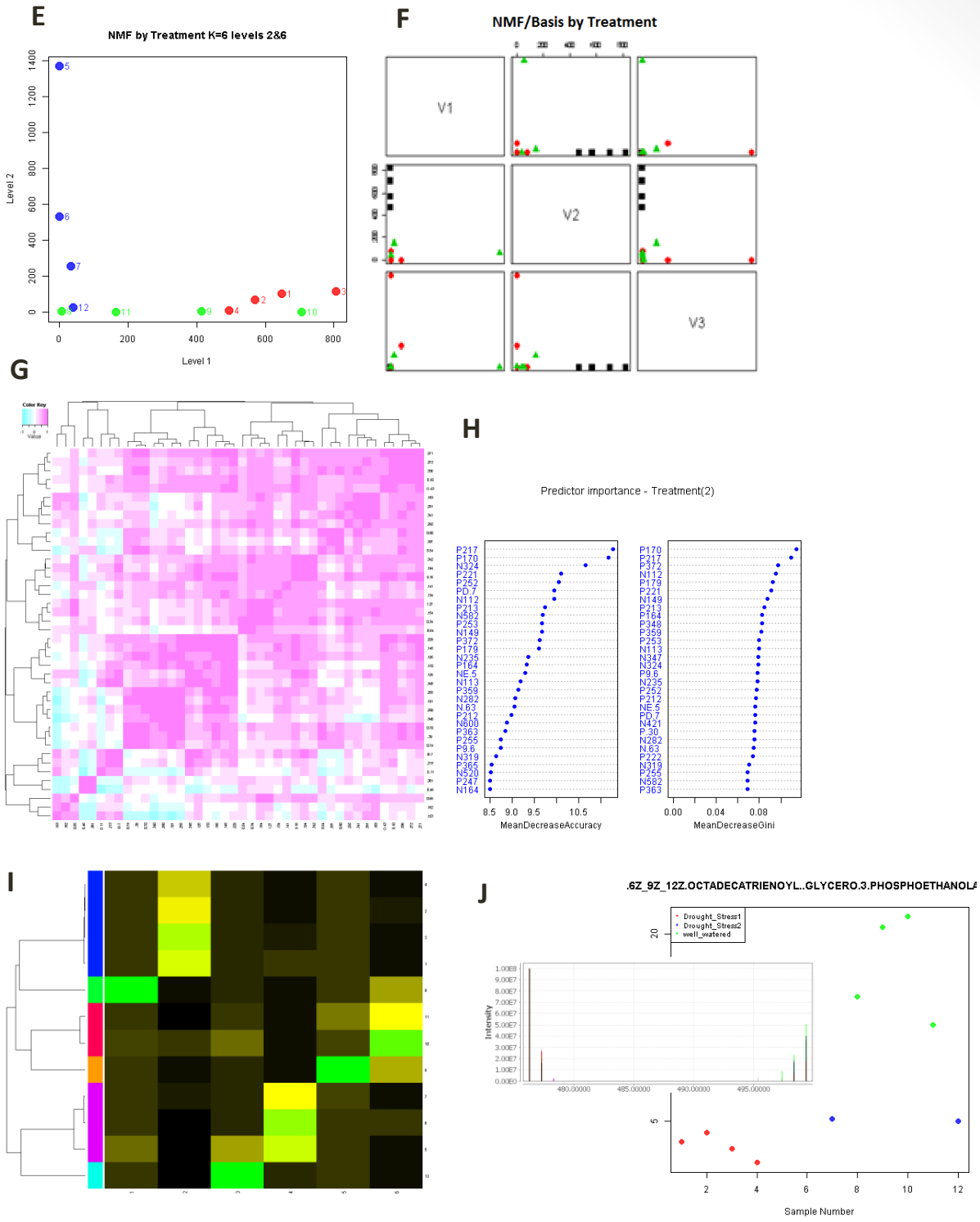


Figure 5: IROA Portal Statistical Analysis: Principle Components Analysis (A&B); Random Forest (C&D); Non-negative Matrix Factorization (E&F); Correlation analysis on 32 most significant compounds selected by RF (H&J); Drought stress #1 ●; Drought stress #2 ●: well watered ●

Methods

Experiment design

Leaf tissue from four DuPont Pioneer maize genotypes were collected from plants field grown under either one of two different drought stresses or one well-watered. Samples were lyophilized and ground. Prior to analysis, individual samples (5-6 mg dry weight) were mixed with an equal quantity of ground maize leaf that had been grown in an environmental chamber where the CO_2 levels were controlled to maintain 97.5% ^{13}C and 2.5% ^{12}C (standard). The combined maize leaf samples were held at -80°C until preparation.

Sample Preparation

Cold extraction solvent (0.5mL) consisting of MeOH:Water (80:20) and containing two internal standards was added and samples were homogenized using a Genogrinder. Samples were placed on a rotator for 1.5hr @ 4°C. Samples were centrifuged and transferred to 0.2µm spin filters and centrifuged for 5min. Extracts were transferred to limited volume autosampler vials and crimp sealed.

LC-MS

Samples were stored in freezer and at 4°C on the instrument autosampler. Twenty five microliter aliquots were submitted to reversed phase HPLC coupled to a high resolution Thermo uPLC/LTQ-FTICR-MS. The FT was operated at 100,000 resolution.

Data processing and analysis

The dataset was analyzed by the IROA ClusterFinder software (Figure 6). The 985 peaks identified tentatively by ClusterFinder were compared against libraries of compounds for their IROA characteristics; ^{12}C base peak, ^{12}C M+1, ^{13}C base peak, ^{13}C M-1, and intervening peaks. The metabolic phenotype of each experimental sample consists of the relative concentration of a number of metabolites. In the IROA method, all compound measurements are made relative to a ^{13}C standard; in this case, an isotopically-labeled maize standard (97% ^{13}C-enriched, IsoLife). Therefore, these measurements represent the

deviation of each metabolite relative to the isotopically-labeled standard. The resulting dataset was analyzed in the IROA portal to assure that the samples were statistically similar. The compounds found in the samples were analyzed by Principal Component Analysis, Random Forest, Hierarchical Clustering, Non-negative Matrix Factorization and Correlational analysis.

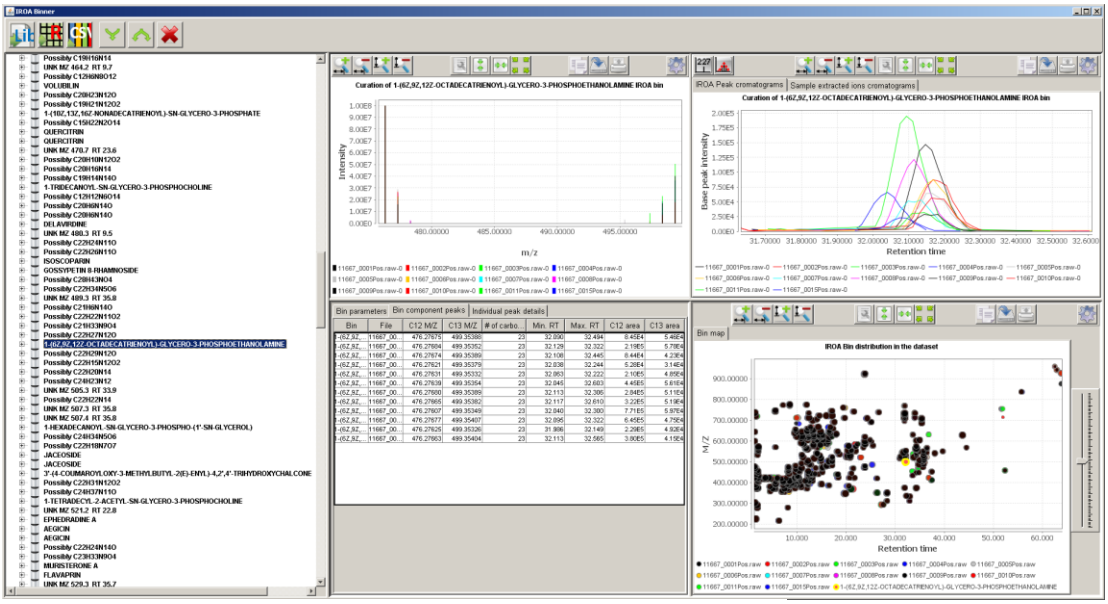


Figure 6: Screen shot of ClusterFinder software analysis.

Conclusions

The additional ability to understand structural aspects of compounds given through the use of distinct isotopic signatures, as seen in the mass spectra, made it possible to assign names and understand structure to a degree not possible if only natural abundance peaks are present. All experimental biological peaks exhibited a ^{13}C-labeled pair and therefore artifacts were removed from consideration.

Summary

This IROA Phenotypic study was performed in maize to understand key growth processes that improve drought tolerance, mediating selection of advantageous genotypes. The IROA results demonstrated clearly that the genetically distinct maize hybrids under drought conditions exhibited diverse metabolic profiles. The significantly enhanced quantitation of having a complex internal IROA Standard present in every experimental sample is a major benefit of the IROA protocol. All experimental biological peaks will have a ^{13}C-labeled pair and therefore artifacts may be removed from consideration and prevented from becoming false positives. The application of the IROA standard as a “recovery”-type standard, i.e. put in prior to sample prep and carried throughout the rest of the sample preparation, is a means of further enhancing data quality and reducing sample-to-sample variation (correcting both in-sample and between sample variations).