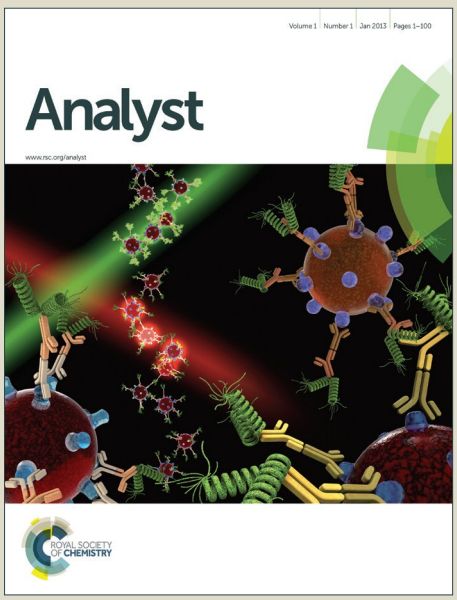
Analyst

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/analyst

Multi-Matrix, Dual Polarity, Tandem Mass Spectrometry Imaging Strategy Applied to a Germinated Maize Seed: Toward Mass Spectrometry Imaging of an Untargeted Metabolome

Adam Feenstra[1,2], Rebecca Hansen[1,2], Young Jin Lee[1,2,*]

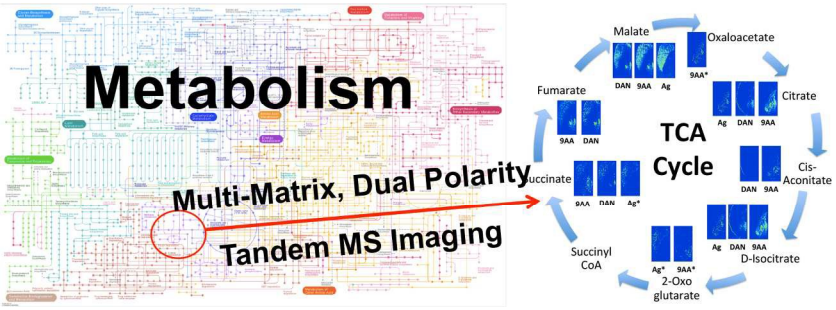
[1]*Department of Chemistry, Iowa State University, Ames, IA 50011*

[2]*Ames Laboratory-US DOE, Ames, IA 50011*

* Correspondence to: Young Jin Lee, Department of Chemistry, Iowa State University, Ames, IA 50011, USA, E-mail: yjlee@iastate.edu

Table of Contents

Mass spectrometry imaging strategy to allow for visualization and identification of compounds on tissue to help understand plant metabolism



Abstract

Mass spectrometry imaging (MSI) provides high spatial resolution information that is unprecedented in traditional metabolomics analyses; however, the molecular coverage is often limited to a handful of compounds and is insufficient to understand overall metabolomic changes of a biological system. Here, we propose an MSI methodology to increase the diversity of chemical compounds that can be imaged and identified, in order to eventually perform untargeted metabolomic analysis using MSI. In this approach, we use the desorption/ionization bias of various matrixes for different metabolite classes along with dual polarities and a tandem MSI strategy. The use of multiple matrixes and dual polarities allows us to visualize various classes of compounds, while data-dependent MS/MS spectra acquired in the same MSI scans allow us to identify the compounds directly on the tissue. In a proof of concept application to a germinated corn seed, a total of 166 unique ions were determined to have high-quality MS/MS spectra, without counting structural isomers, of which 52 were identified as unique compounds. According to an estimation based on precursor MSI datasets, we expect over five hundred metabolites

could be potentially identified and visualized once all experimental conditions are optimized and an MS/MS library is available. Lastly, metabolites involved in the glycolysis pathway and tricarboxylic acid cycle were imaged to demonstrate the potential of this technology to better understand metabolic biology.

Introduction

Since its development nearly two decades ago, molecular mass spectrometry imaging (MSI) has become a useful and increasingly common tool for tissue analysis.[1, 2] Several desorption/ionization techniques have been adopted for MSI, but by far matrix-assisted laser desorption/ionization (MALDI)-MSI has been most widely used.[3-6] Applications demonstrated thus far include imaging various classes of compounds such as lipids, peptides, proteins, drugs, and small molecules in tissue samples as diverse as animal organs, plant tissues, and bacterial colonies. The benefit of MSI is that it combines the versatility and high specificity of mass spectrometry with high spatial resolution localization information. While MSI cannot reach the spatial resolution of optical imaging techniques, it provides valuable chemical information and the ability to simultaneously image hundreds of chemical compounds in one sample.

MALDI-MSI is often used for targeted imaging of a specific compound or class of compounds (e.g., drugs or drug metabolites[7]); however, its application for complex biological phenomena (e.g., nitrogen fixation[8]) is relatively limited. The lack of such "discovery" type experiments originates from 1) the limited number of compounds that can be characterized with a particular matrix and 2) the difficulty in confidently identifying metabolites directly on tissue. Most discovery type experiments in metabolomics studies,

so called 'untargeted metabolomics', are being performed with traditional gas chromatography-electron ionization-mass spectrometry (GC-EI-MS) or liquid chromatography-tandem mass spectrometry (LC-MS/MS) experiments, which require sample homogenization, resulting in the loss of spatial localization information of metabolites.

Untargeted metabolomic profiling is limited even in traditional analyses by the difficulty of confidently identifying metabolites due to the lack of EI-MS or MS/MS databases. This issue is especially significant in plant metabolomic analysis because over 200,000 metabolites are expected to be present in the plant kingdom and most of them are unknown.[9, 10] For example, the NIST EI-MS library has over 200,000 entries, but most of them are from man-made synthetic compounds or abundant natural products. In typical untargeted GC-MS profiling of a plant metabolome, one-half to two-thirds of the metabolites cannot be confidently identified.[11-13] Unknown compound identification in LC-MS/MS is even more difficult, as its database is much smaller.

Recently there have been some efforts to build widely applicable metabolite MS/MS databases. The Suizdak group has accumulated MS/MS spectra of 12,834 metabolites (as of May 11, 2015) in their Metlin database (metlin.scripps.edu), but that number is only about 5% of the number of known metabolites they list (240,588), not to speak of unknown metabolites. Another difficulty arises from the fact that MS/MS spectral patterns strongly depend on the instrument type and collision energy used. Many of the current MS/MS databases, such as Metlin, are being constructed for a quadrupole time-of-flight mass spectrometer (Q-TOF) and may not be as useful for other instruments, such as an ion trap.

We have previously developed an MSI strategy[14] in which each raster step is split into multiple spiral steps composed of a precursor MS scan followed by several MS/MS or MSn scans, which we called 'multiplex MSI'. We also demonstrated that a wider variety of lipid classes could be visualized by incorporating polarity switching into multiplex MSI.[15] In this work, we attempt to expand the multiplex MSI approach to a large metabolomics scale by using the inherent bias of matrixes for desorbing/ionizing different classes of compounds to analyze as wide of a range of compounds as possible. Following accurate mass precursor scans, data-dependent MS/MS spectral acquisition is performed so that we can obtain hundreds or thousands of MS/MS spectra directly on tissue. To overcome the lack of chromatographic separation in MSI, we propose to use multiple serial tissue sections and selectively ionize various metabolite classes by utilizing different matrixes and polarities. By improving the two bottlenecks in MSI, the number of compounds that can be visualized and the confidence in their identifications, we propose MSI of an untargeted metabolome is possible.

In the present study, we demonstrate this strategy as applied to a germinated maize seed. Seed germination is a critical stage in the life cycle of a plant when very active metabolic changes occur to break down storage starch and proteins and convert them to developing tissues and organs[16]. Various metabolic pathways are turned on at specialized locations, making it an ideal system to demonstrate the proposed MSI strategy.

Experimental

Materials

Organic matrixes were purchased from Sigma-Aldrich (St. Louis, MO, USA): 1,5-diaminonaphthalene (DAN, 97%), 2,5-dihydroxybenzoic acid (DHB, 98%), and 9-aminoacridine (9AA, 98%) hydrochloride hydrate. 9AA was converted to the free base form upon arrival and before use by dissolving in boiling water and adding an equimolar amount of sodium hydroxide. The precipitate was then filtered, washed three times with cold water, and dried. Iron oxide nanoparticles (NPs) (Fe_3O_4, 11nm, no organic capping) were synthesized and provided by Dr. Javier Vela's research group at Iowa State University.[17] Tungsten oxide (WO_3, 23-65nm, no organic capping) and silver (20nm, no organic capping) NPs were obtained from US Research Nanomaterials, Inc. (Houston, TX, USA) and suspended in 2-propanol (Sigma Aldrich, LC-MS Chromasolv) upon arrival. Gelatin from porcine skin (300 bloom) was purchased from Electron Microscopy Sciences (Hatfield, PA, USA). B73 inbreed corn seeds were obtained from Dr. Marna Yandeau-Nelson at Iowa State University.

Corn Seed Growth

A B73 inbreed corn seed was imbibed by placing it in a glass scintillation vial with deionized water and shaking the vial at 300rpm for 10 minutes. The seed was then placed on a moist piece of filter paper in a plastic petri dish with the embryo side of the corn seed down. The petri dish was covered with a transparent, plastic lid and placed in a climate-controlled greenhouse (27°C at day time, 24°C at night time, 15 hours of day light each day) to grow. The seed was allowed to germinate for 29 hours with water periodically added to keep the filter paper moist but without standing water.

After 29 hours, the seed was removed from the greenhouse. Using a razor blade, the seed was cut longitudinally approximately halfway between the edge and center. Immediately after cutting, the seed was flash-frozen in liquid nitrogen to prevent metabolite turnover. Once the seed was frozen, it was stored at -80°C until cryosectioning and transported on dry ice to prevent thawing.

Sample Preparation for Imaging

Sample preparation procedure for MSI of plant tissues is described in detail in our recent protocol paper[18] and is briefly described below. The procedure consists of cryosectioning, drying, and matrix application. For cryosectioning, the frozen seed was placed in a cryo-mold with the flat, cut face down. A warm, aqueous, 10% w/v gelatin solution was poured into the mold to completely cover the seed. The mold was immediately floated on liquid nitrogen until the gelatin in the mold had become mostly opaque, then transferred to a cryostat pre-chilled to -22°C. Once the gelatin medium was completely frozen and thermally equilibrated (~30 min), the block was mounted onto the cryostat chuck and sections of the region of interest were cut at 10μm thickness and captured intact using adhesive tape windows (Leica Biosystems, Buffalo Grove, IL, USA). The tape windows with tissue sections were taped face-up onto chilled glass slides to prevent thawing, then transported on dry ice and stored at -80°C until further processing.

Adjacent serial tissue sections were selected for imaging. The slides with the attached seed sections were placed onto a chilled aluminum block, and then lyophilized under moderate vacuum (~250mtorr) while slowly equilibrating to room temperature.

Matrix was then either applied by sublimation-vapor deposition as previously described (for DHB)[19], or by a homemade oscillating capillary nebulizer (for 9AA, DAN, Fe_3O_4 NPs, WO_3 NPs, Ag NPs).[20] The matrix-coated tissue section on the tape window was transferred to a stainless steel slide, which was put into the MALDI plate holder and inserted into the mass spectrometer.

MALDI Mass Spectrometry Imaging

Optical images of each tissue were obtained by scanning and combining camera shots in the instrument (image resolution of ~6 μm/pixel). Imaging data was acquired using a MALDI-linear ion trap (LIT)-Orbitrap mass spectrometer (MALDI-LTQ-Orbitrap Discovery; Thermo Finnigan, San Jose, CA) that has been modified to incorporate an external Nd:YAG laser (UVFQ; Elforlight, Ltd., Daventry, UK). Laser pulse energy and number of shots were optimized individually for each matrix.

A multiplex MSI strategy was adopted for data acquisition using a four spiral step pattern for each pixel (see Figure 1) with a 100μm raster step (pixel) size and 50μm spiral step size. The laser spot size is 25μm as estimated from the burn mark on thin film of DHB. In the first scan event, a full Orbitrap mass spectrum was acquired for *m/z* 50-800 with a mass resolution of 30,000 at m/z 400. Next, a data-dependent ion trap MS/MS scan was performed from the previous scan. In the third scan event, another full Orbitrap scan was acquired for *m/z* 600-1600, followed by another data-dependent ion trap MS/MS scan from the third scan event.

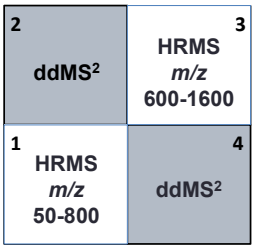


Figure 1. Four step multiplex MSI setup used in this work. Steps 1 and 3 are high resolution Orbitrap scans in low and high mass range. Steps 2 and 4 are data dependent ion trap MS/MS scans.

To ensure MS/MS would be performed for known metabolite compounds, two separate precursor mass lists (136 masses in positive mode, 93 in negative mode) consisting of previously observed metabolites were imported to the Excalibur data acquisition program (Thermo) for positive and negative ion mode. It should be noted that MS/MS have not been previously obtained for many of these compounds, and certainly not directly on maize embryos. Collision energies were broadly assigned depending on the precursor mass range for optimal fragmentations (*m/z* 50-300, 125%; *m/z* 300-600, 75%; above *m/z* 600, 35%). If no ion was found among the precursor masses within a 0.05 Da mass tolerance, the instrument was set up to perform MS/MS of the most abundant ion with a default setting of 35% collision energy. Collision duration was 30ms in all cases. Dynamic exclusion was also adopted to obtain as many useful MS/MS scans as possible using a repeat count of one and an exclusion duration of 50 seconds.

All the images shown in this manuscript were generated using ImageQuest (Thermo) with a ±0.01Da mass tolerance and in absolute ion intensity scale without any normalization. Normalization with total ion count was also performed but displayed almost no difference.

MS/MS Data Analysis

Manual interpretation was performed for high quality MS/MS spectra. To reduce the number of MS/MS spectra for manual interpretation, we performed three levels of filtering to extract only those MS/MS spectra that are most likely to be meaningful. The first filtering was performed at the MS/MS level. All MS/MS spectra were extracted from each raw data file as separate ascii files using ExtractMSN.exe (Thermo). Filtering parameters were optimized by trial and error and imposed on these ascii files in order to minimize the loss of MS/MS spectra from real metabolites while removing most of the MS/MS from contaminants. This procedure removed most of the low quality MS/MS spectra. The following filtering parameters were used for the MS/MS data sets. For negative mode, compounds with *m/z* 100-300 were filtered with a total ion count (TIC) threshold of 50 and a minimum of 2 fragment ions, whereas compounds of *m/z* 300-1600 were filtered with a TIC threshold of 75 and a minimum of 3 fragment ions. For positive mode, compounds with *m/z* 100-400 were filtered with a TIC threshold of 50 and a minimum of 2 fragment ions while compounds of *m/z* 400-1600 were filtered with a TIC threshold of 1000 and a minimum of 10 fragment ions. In the precursor mass list that passed the above MS/MS filtering, we ignored structural isomers and combined all the same precursor masses by treating all masses within ±0.01Da as if they were the same compound.

The second filtering was at the precursor MS spectral level. Precursor mass spectra were averaged over the entire sample and low intensity precursor ions were removed if their ion abundance was below 0.1% of the base peaks. Then, 13 carbon isotopes were

removed from the remaining precursor masses with the mass difference of 1.00335Da (i.e., mass difference of ^{13}C-^{12}C).

The final filtering was at the MS image level. The precursor mass list that passed both MS/MS and MS level filtering above was subjected to image inspection and the ions from the non-tissue area were removed. The images were generated in ImageQuest with a ±0.01Da mass tolerance and manually inspected to determine if the signal was localized only on the tissue. This left a final mass list composed of high-abundance, potentially identifiable metabolite ions.

All MS/MS spectra that passed through the above filtering were then subjected to manual interpretation. Based on the image localization of each precursor mass, MS/MS spectra were averaged where signal was present using QualBrowser (Thermo). The averaged MS/MS were interpreted in a multi-step process. First, the accurate precursor mass was searched against metabolite, lipid, or chemical databases (Metlin, LipidMaps (www.lipidmaps.org), Chemspider (www.chemspider.com)) with a mass tolerance of 5ppm. If a matching compound was found, Metlin and MassBank (www.massbank.jp) were checked for the presence of MS/MS spectra for comparison. If a compound had at least two major fragments matching and no unusual fragments, the assignment was presumed to be correct. If no MS/MS was available for comparison, manual MS/MS spectral interpretation was attempted to explain the possible fragmentation of matching precursor compounds.

Estimation of Potential Metabolites Based on Accurate Mass

The total number of identifiable metabolites was estimated from the precursor mass spectra based solely on accurate mass. MSiReader[21] was used for automatic MS image

generation for both low and high mass scans of the Orbitrap spectra. The Peak Finder tool in MSiReader was used to select the entire tissue sample and generate a peak list for the masses found only on tissue. The parameters used in the Peak Finder tool were a minimum of 0.1% presence in the interrogated zone and a maximum presence of 5% (or maximum S/N of two) in the reference zone.

After the peak list was generated, the Batch Processing feature of MSiReader was used to automatically generate images for all the masses with ±0.005Da mass tolerance (none of these images are used in the text; rather, this served as a quick and efficient way of determining if peaks were localized on tissue). Any peaks whose images showed localization anywhere off the tissue area were removed. Images that were indeterminate were also generated in ImageQuest to check their localization. For masses that were separated by less than 0.02Da, their images were compared to see if they showed different localizations. For masses with the same localization, the precursor mass spectrum was carefully inspected for the image area to determine if multiple masses were actually present. If not, the mass of the only peak in the spectrum for that area was used.

After image filtering, 13C isotope peaks were manually removed from this mass list in both positive and negative modes. Additionally, potential multiple alkali metal adducts (+Na, +K) were removed in positive mode data while potential water loss peaks were removed in negative mode data. All removals were based on being within 5ppm of the suspected mass values. This left a final peak list localized only on the tissue area that are considered as potentially identifiable metabolites. This mass list was searched against the Metlin metabolite database for possible matches within 5ppm.

Results and Discussion

Comparison of Different Classes of Compounds Between Matrixes

A corn seed was germinated for 29 hours and then serial sections of the seed were subjected to MALDI-MSI analysis with a different matrix applied to each section. To demonstrate selectivity of each matrix in characterizing specific compounds of interest, representative compounds detected in three matrixes in negative ion mode are compared in Figure 2. Detailed anatomy of a maize seed is shown in Supplemental Figure 1.

There is no perfect way to compare images between different matrixes due to multiple variables. The best method we found is to individually optimize each matrix, especially matrix application and laser energy, and compare ion signal levels obtained at each optimum condition. We minimized other variables as much as possible, such as use of serial sections to minimize biological variations and acquiring data next to each other on the same day to minimize analytical variations. We have found the results were fairly reproducible in this approach. The images shown in Figure 2 were produced in absolute ion intensity scale, without any normalization. The intensity scale is adjusted for each analyte that can best visualize the image contrast. Normalization to the total ion count (TIC) is often adopted to minimize spot-to-spot variation; however, it may not the best method for the current purpose because it can exaggerate relative ion signals for a matrix that gives low ion signals. We also compared the images with TIC normalization as shown in Supplementary Figure 2, which gave virtually the same results.

The selected compounds show a better desorption ionization efficiency for one matrix compared to the others, as can be observed through the quality of images generated. DAN, for example, shows high quality images for the acidic phospholipids and small

molecule classes (shown in the top set of images). Previously, Chaurand has shown DAN to be effective for lipids in negative mode[22] and here we observe similar results. Phosphatidylethanolamine (PE) 34:2 and phosphatidic acid (PA) 36:4 are both localized in the developing embryo and radicle tissue (initial root of the seed) in the images obtained with DAN, while 9AA and Ag NPs have almost no signal present. Other lipids, such as phosphatidylinositol (PI) and phosphatidylglycerol (PG) (not shown), also show much better signal with DAN compared to the other matrixes. DAN is also effective for small molecules such as glutamine, which appears in the radicle of the seed, and malic acid, which is present in the radicle and endosperm (starchy storage reserves of the seed). This is consistent with our recent report that DAN is a good matrix for small molecules in negative ion mode.[23]

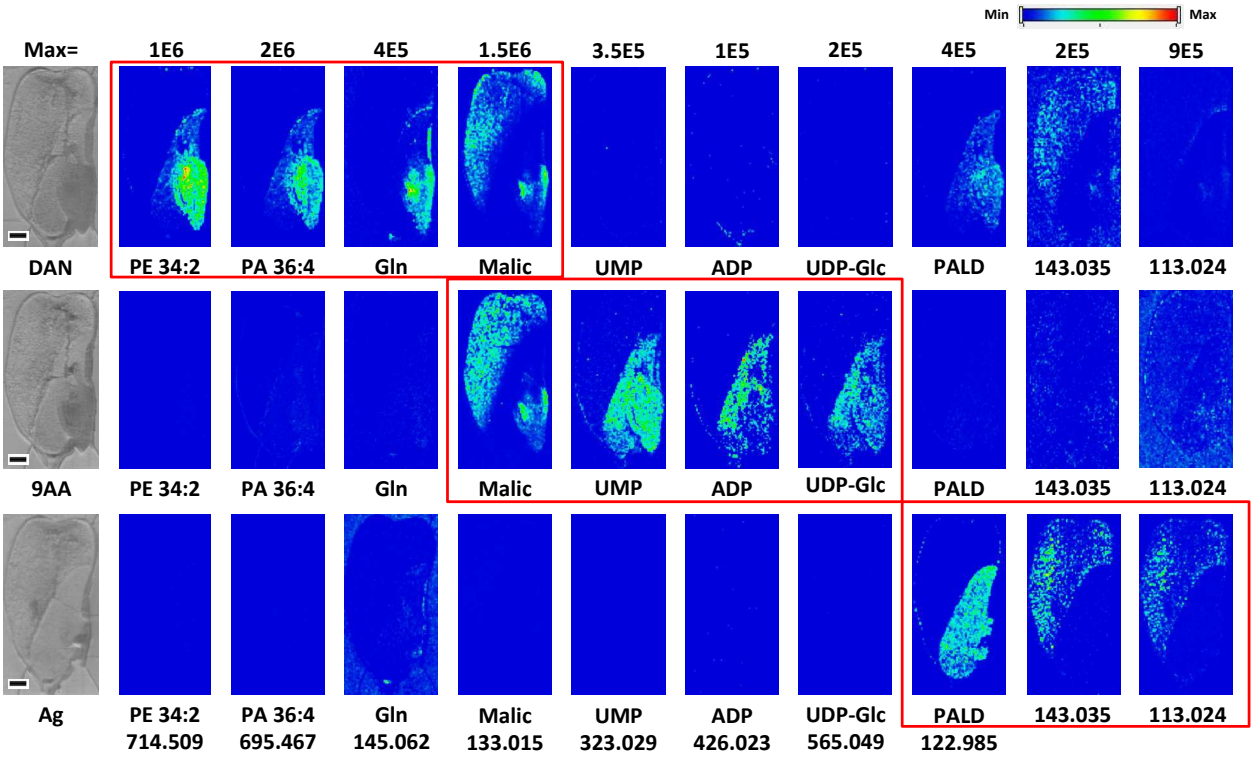


Figure 2. Comparison of ten representative compounds in a germinated maize seed showing different selectivity between three matrixes in negative ion mode. MS

images are shown for phosphatidylethanolamine 34:2 (PE 34:2), phosphatidic acid 36:4 (PA 36:4), glutamine (Gln), malic acid (Malic), uridine monophosphate (UMP), adenosine diphosphate (ADP), uridine diphosphate glucose (UDP-Glc), phosphonoacetaldehyde (PALD), and two unknown compounds at *m/z* 143.035 and 113.024. M/z values used in generating images are given at the bottom of the figure. Red boxes highlight ion images more selectively detected with a given matrix than in other matrixes. All ions are observed as deprotonated species. Scale bars represent 1mm.

9AA is well known for its effectiveness for small molecules in negative mode.[24, 25] As shown in Figure 2, it is comparable with DAN for some organic acids, such as malic acid, but inefficient for some other molecules, such as glutamine.[23] However, 9AA shows much better ionization efficiency for nucleotide type compounds, as has been previously demonstrated.[26] Examples shown here include uridine monophosphate (UMP), adenosine diphosphate (ADP), and uridine diphosphate glucose (UDP-Glc), all distributed throughout the radicle and embryo of the seed. DAN and Ag NPs show no signal for this class of compounds however.

Silver nanoparticles were selected as one of the matrixes in this study because they are known to be effective for fatty acids and phospholipids.[27, 28] However, at least for the current tissue sample of a germinated maize seed, they were not as efficient for phospholipids as DAN. Fatty acids could be successfully visualized with silver nanoparticles, but DAN produced much better images (not shown). Ag NPs showed better selectivity for some small molecules, such as phosphonoacetaldehyde (PALD), and for two unidentified peaks (likely small, organic acid type compounds) at *m/z* 143.035 and 113.024. PALD and *m/z* 143.035 were also detected with DAN, but with much lower

signals. PALD is evenly distributed throughout the embryo of the seed while m/z 143.035 and 113.024 species are distributed throughout the endosperm area of the seed.

Similar results were obtained in positive mode as demonstrated with nine representative compounds in Figure 3. DHB, tungsten oxide NPs, and iron oxide NPs show different desorption ionization efficiency behavior for various classes of compounds, although there was some similarity between the two metal oxide nanoparticles. DHB is well known to have good efficiency for a relatively wide variety of compounds, especially for large molecules including lipids, proteins, and oligosaccharides.[22, 29-31] Neutral or cationic lipids, such as ceramide 42:1 and phosphatidylcholine (PC) 34:2, are efficiently detected with DHB, along with large oligosaccharides and an unidentified compound at *m/z* 972.521. The sphingolipid ceramide is very uniquely localized on the boundary area between the embryo and endosperm part of the seed. PC 34:2 and the compound at *m/z* 972.521 are seen primarily in the radicle of the seed with some distribution throughout the rest of the embryo. Large oligosaccharides, such as heptahexose, are localized in the endosperm area. Neither of the NPs shows any signal for these classes of compounds.

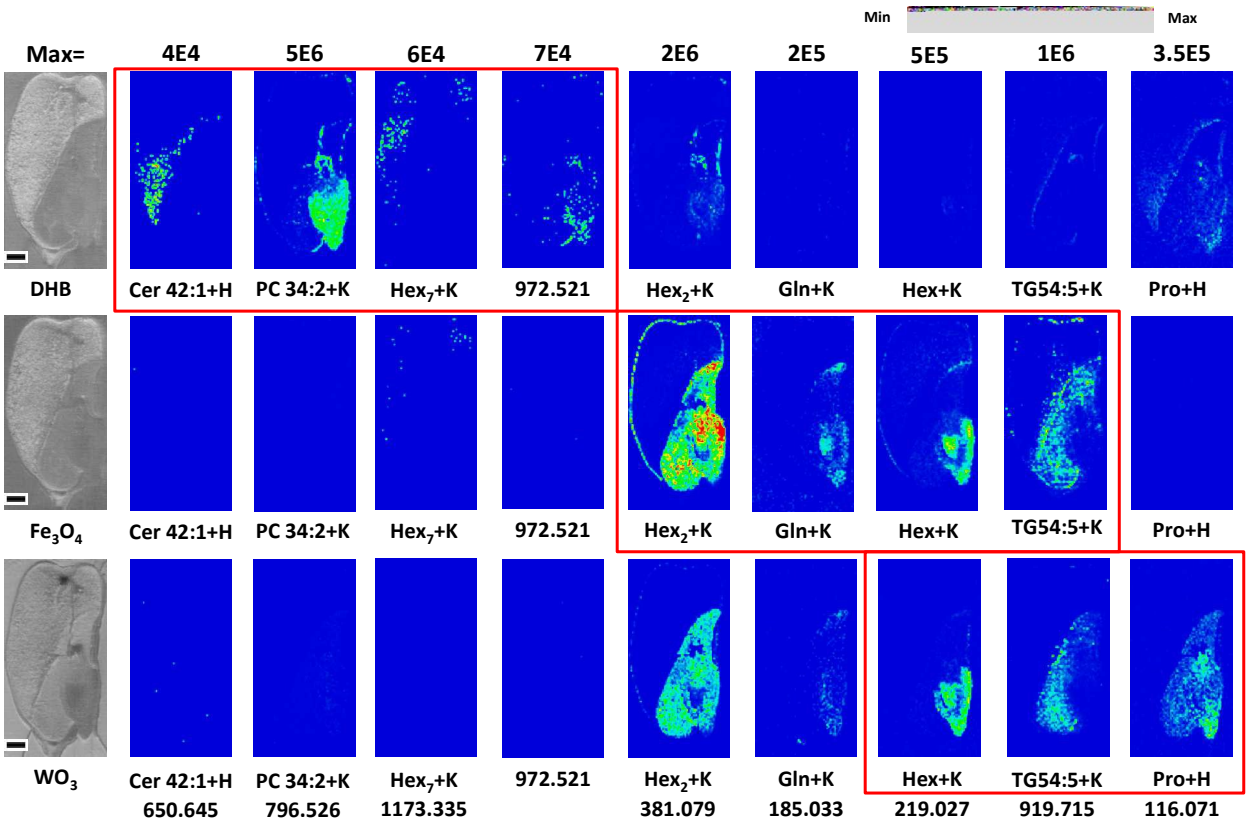


Figure 3. Comparison of nine representative compounds in a germinated maize seed between three matrixes in positive ion mode. MS images are shown for ceramide 42:1 (Cer 42:1), phosphatidylcholine 34:2 (PC 34:2), heptahexose (Hex_7), an unknown compound at *m/z* 972.521, dihexose (Hex_2), glutamine (Gln), monohexose (Hex), triacylglycerol 54:5 (TG 54:5), and proline (Pro). Red boxes highlight ion images more selectively detected with a given matrix than in other matrixes. All known compounds are shown as a potassium adduct, except for ceramide and proline, which are shown as a proton adduct. Scale bars represent 1mm.

Iron oxide and tungsten oxide NPs show significant differences with DHB, but share some similarity between themselves. Both are able to effectively ionize small sugars (e.g., monohexose and dihexose, most likely glucose and sucrose) and triacylglycerols (TG) much better than DHB. For instance, in the case of dihexose, iron oxide and tungsten oxide NPs have about fifteen times and five times higher ion signals than DHB, respectively. Overall,

iron oxide NPs are slightly better than tungsten oxide NPs, except for certain compounds such as proline.

It is interesting to note that large oligosaccharides (e.g., Hex_7; the same for Hex_3 to Hex_9, images not shown) are localized in the endosperm area whereas small oligosaccharides (i.e., Hex, Hex_2) are localized in the embryo area. This suggests that starch is broken down in the endosperm and then small sugars are transported to the embryo. Most importantly, the use of multiple matrixes could effectively visualize the localization of both small and large sugars.

All together, the results above suggest that a wide range of compounds can be characterized by combining several different matrixes in positive and negative mode. In negative mode, we are able to detect acidic phospholipids such as PE and PA, amino acids, nucleotides, nucleotide sugars, organic acids and other small molecules. Positive mode presents neutral or cationic phospholipids, amino acids, sugars, and non-acidic small molecules.

MS/MS Based Metabolite Identification

The previous section demonstrated that the combined use of various matrixes in dual polarities is a good strategy to increase the chemical coverage in MALDI-MSI. The other major obstacle to overcome for MSI in metabolomics scale is confident identification of compounds. In-parallel metabolomics analysis by traditional GC- or LC-MS methods is expected to assist metabolite identifications in MSI data; however, it is still necessary to identify metabolites on tissue because there may be several structural isomers with

different tissue localizations or possible alterations of metabolites during sample preparation.

As detailed in the experimental section, the six data sets shown in Figures 2 and 3 were acquired using a multiplex MSI strategy. Figure 1 summarized the overall data acquisition scheme. Two precursor MS scans in low and high mass range (*m/z* 50-800 and *m/z* 600-1600, respectively) were acquired in Orbitrap, followed by data-dependent MS/MS scans in ion trap. We envision high-throughput MS/MS database searches would be possible for thousands of MS/MS spectra acquired through this multiplex MSI. Currently, however, we encounter several obstacles, most notably the lack of an MS/MS database. Our MS/MS data was acquired in an ion trap mass spectrometer, and publicly available ion trap MS/MS databases are very limited. MassBank has some ion trap MS/MS spectra, but the total number is not any more than a few hundred compounds. Hence, in this proof-of-concept study, we performed manual evaluation of each MS/MS spectra for possible identification.

The total number of MS/MS spectra from the six multiplex MSI data sets, three positive and three negative, is 92,248. Many of them are from the same precursor masses, in spite of data-dependent acquisitions, but there are still a large amount of unique MS/MS spectra from different compounds. To minimize the workload in manually interpreting all the MS/MS spectra, we first filtered out low quality MS/MS spectra that were expected to be structurally uninformative. Then, to further minimize the manual labor, we ignored structural isomers and combined MS/MS spectra from the same precursor mass by averaging these spectra together. Additionally, low abundance ions in the precursor MS spectra and 13C isotope ions were removed as well as ions present outside the tissue area.

After this filtering, detailed in the experimental section, we obtained a total of 104 unique ions with high quality MS/MS spectra in positive mode and 62 in negative mode, as listed in Supplementary Tables 1 and 2 and summarized in Figure 4A. A few ions were obtained in all three matrixes, 13 and 4, for positive and negative mode, respectively; however, the majority were obtained only in one matrix and some in two matrixes, as expected from Figures 2 and 3.

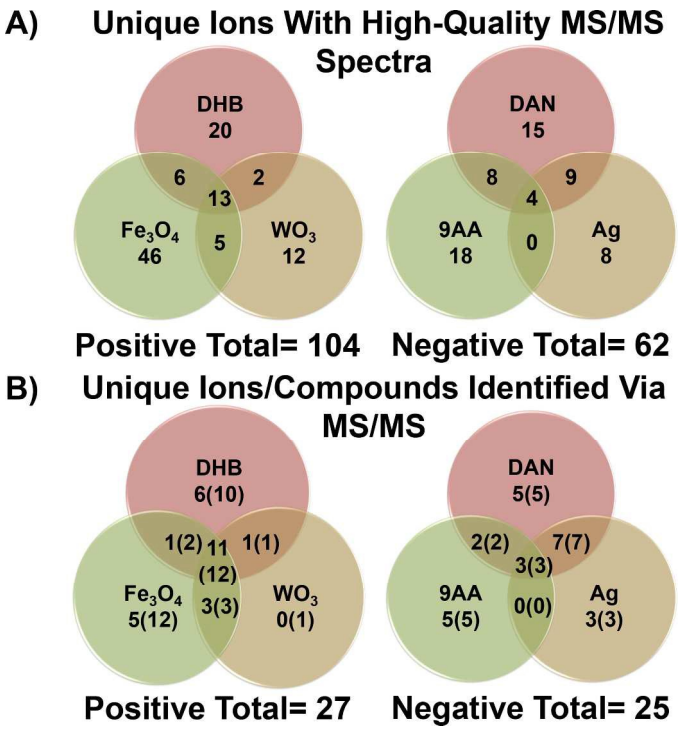


Figure 4. Venn diagrams displaying A) the number of unique ions with high-quality MS/MS spectra and B) the number of unique compounds (outside parentheses) and total number of ions (in parentheses) identified via manual interpretation of MS/MS spectra. The number of unique compounds identified is after the removal of multiple adducts or fragments of the same species.

Manual interpretation was performed for the 166 MS/MS datasets, as described in the Experimental Section, based on accurate precursor mass and MS/MS spectra. When

MS/MS spectra were available on Metlin or MassBank, careful comparison was made while considering the differences in MS/MS spectral patterns between QTOF and ion trap. Namely, our MS/MS spectra were acquired in an ion trap, which is known to have a low mass cutoff, losing the fragment ion signals below 28% of the precursor *m/z* values.[32] Meanwhile, QTOF fragmentation often produces significant low mass fragments due to multiple fragmentations. Accordingly, we mostly compared fragments in higher mass ranges and ignored the difference in fragment ion intensities. If MS/MS spectra were not available or did not match with the queried experimental spectra, metabolites matching the precursor mass either in Metlin, LipidMaps or ChemSpider were subjected to manual interpretations to attempt to match experimental fragment spectra with expected fragment structures.

After manual interpretation of the 166 MS/MS spectra, a total of 41 and 25 spectra were tentatively identified in positive and negative mode, respectively, as summarized in Supplementary Tables 3 and 4. Out of those, 27 were determined to be unique compounds in positive mode (multiple cation adducts and identifiable fragments such as PC trimethylamine losses were counted as the same species), and all 25 were unique in negative mode. These results are summarized in Figure 4B as Venn diagrams.

We were tentatively able to identify only ~40% of the ions in each polarity among the potentially identifiable MS/MS spectra, largely due to the lack of an MS/MS database. MS/MS spectra common in all three matrixes were mostly identified, with 12 out of 13 in positive mode and 3 out of 4 in negative mode. However, MS/MS spectra detected in only one matrix had a much lower success rate; e.g., only 1 out 12 and 12 out of 46 identified in WO_3 and Fe_3O_4, respectively. This is likely due to the fact that the metabolites found in all

three matrixes are present in high abundance, thus giving multiple sets of high-quality MS/MS spectra. In contrast, those found in only one matrix might be present in low abundance and/or with weak MS/MS signals. In other words, some high-abundance compounds could be detected regardless of matrix, although ion signals could be low in some matrixes, and thus have a good chance to be identified. Hence, it is especially for low abundance metabolites that we need to carefully optimize matrix selectivity and MS/MS conditions for effective detection and characterization.

Figure 5 shows two examples of metabolites identified through this analysis. MS/MS spectra for *m/z* 833.521 were obtained using both silver NPs and DAN. The two spectra are very similar to each other, but DAN gives better signals and is shown in Figure 5A. Searching the precursor mass against the Metlin database gives 19 metabolites within 5ppm mass tolerance, all various structural isomers of PI (34:2), but MS/MS spectra are present for none of them. Phospholipids can be manually characterized in MS/MS, most notably from the easily identifiable loss of fatty acid side chains.[33, 34] The most intense ion in Figure 5A, *m/z* 553.3, corresponds to C18:2 fatty acid loss (280.2 Da). The C16:0 fatty acid loss (256.2 Da) is also found at *m/z* 577.3. Further fragmentation results in the two lysophosphatidic acid (LPA) peaks seen at *m/z* 391.3 and 415.3, respectively. These two peaks represent species where the inositol headgroup and one of the fatty acid side chains have fragmented off. Additionally, deprotonated C16:0 and C18:2 fatty acid fragments are found at *m/z* 255.3 and 279.3 respectively. Thus, we can confidently assign this *m/z* 833.517 as PI (18:2/16:0).

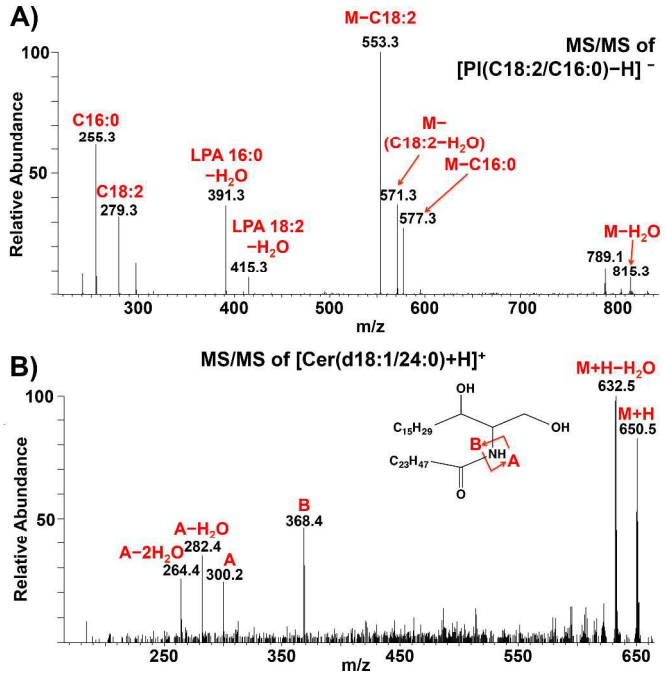


Figure 5. MS/MS spectra for (A) *m/z* 833.521 with DAN as a matrix in negative ion mode, and (B) *m/z* 650.644 with DHB as a matrix in positive ion mode. They are assigned as deprotonated PI (18:2/16:0) and protonated Cer (d18:1/24:0), respectively, according to accurate precursor mass and MS/MS analysis.

It should be noted that we might lose identification of possible structural isomers due to averaging multiple spectra. In the case of PI (34:2) (Figure 5A), we only see two fatty acid fragments in the averaged spectrum and therefore most likely there is only one structural isomer present in the spectra we averaged. However, some other lipids displayed patterns of fatty acids that could result in multiple structural isomers. For example, MS/MS of PA(36:3) shown in Supplementary Figure 3 is an averaged spectrum of 91 MS/MS spectra with the same precursor mass of *m/z* 697.483, and has fragments corresponding to 16:0, 18:0, 18:1, and 18:2 fatty acid chains, suggesting it is most likely the mixture of PA (16:0/20:3), PA (18:2/18:1) and PA (18:3/18:0), although 20:3 and 18:3 fatty acids and their losses are not detected. However, for the current purpose, we did not

verify each MS/MS spectrum, but simply assigned it as PA (36:3). Separate identification of structural isomers could be done efficiently by developing an MS/MS database for automatic searching.

In some cases, interpretation of MS/MS data was not straightforward without reference MS/MS spectra. Figure 5B shows MS/MS of *m/z* 650.644, which matches with ceramide 42:1 according to the Metlin database. MS/MS spectrum for Cer (d18:1/24:0) was available at Metlin with predicted fragment structures, based on which our MS/MS spectra could be interpreted and was found to match. Although the presence of ceramides is not reported in maize seeds to our knowledge, it has been studied in roots of maize[35] and seeds of other plants.[36] This fact, combined with our MS/MS spectrum matching with that on Metlin, allows us to be confident in the identification of this compound as Cer (d18:1/24:0).

To confirm the robustness of this approach, a replicate experiment was performed on another seed, which showed similar results, as summarized in Supplementary Figure 4. The number of metabolites was lower in negative mode, especially for small molecules, and it was attributed to have come from biological variations. Specifically, acquiring MS/MS of low abundance metabolites is significantly dependent on the native abundance.

Potential Metabolites Based on Accurate Mass

As summarized in Figure 4A, this work suggests that multiplex MSI can potentially identify and visualize over one hundred metabolite ions. Some of them are from the same compound with different alkali metal ion adducts, but there also are many compounds we ignored in this study, including structural isomers and those with low quality MS/MS spectra. As a result, we expect the number of metabolites we can identify to be much higher

if we have an available, comprehensive, automatically searchable MS/MS database. In addition to database development, another important step would be to optimize experimental conditions to increase ion signals for low abundance compounds and also their MS/MS spectral quality. While we are working on an MS/MS database for automatic searching and the optimization of experimental conditions, in this section we present the analysis of precursor mass scans to extract how many peaks could be assigned as potential metabolites based only on accurate mass. This allowed us to estimate how many metabolites could be potentially identified once we have an available database and optimized experimental conditions.

The details of how we extracted the potential metabolite mass list from the precursor mass spectra are described in the experimental section. In short, a peak list present exclusively on the tissue was generated using the peak finder tool in MSiReader with at least 0.1% ion abundance of the base peak; then, 13C isotopes, apparent multiple cation adducts, and possible in-source fragmentations of water loss were manually removed. Figure 6A shows the summary of this analysis. A total of 341 and 258 peaks were identified as potential unique metabolites present in the precursor mass spectra of positive and negative mode, respectively. The mass values were searched against the Metlin database and Figure 6B summarizes the result for those peaks that matched with at least one metabolite in the Metlin database within 5ppm. It suggests a total of 128 and 89 metabolites could be potentially characterized in positive and negative mode, respectively, totaling over two hundred compounds combining both ion modes even in the current 'limited' metabolome. Accordingly, this analysis suggests that over two hundred compounds are potentially identifiable, once an MS/MS database becomes available and

MS/MS experimental conditions are optimized. Ultimately, five hundred or more compounds could be identified if a more complete understanding of the corn seed metabolome becomes available.

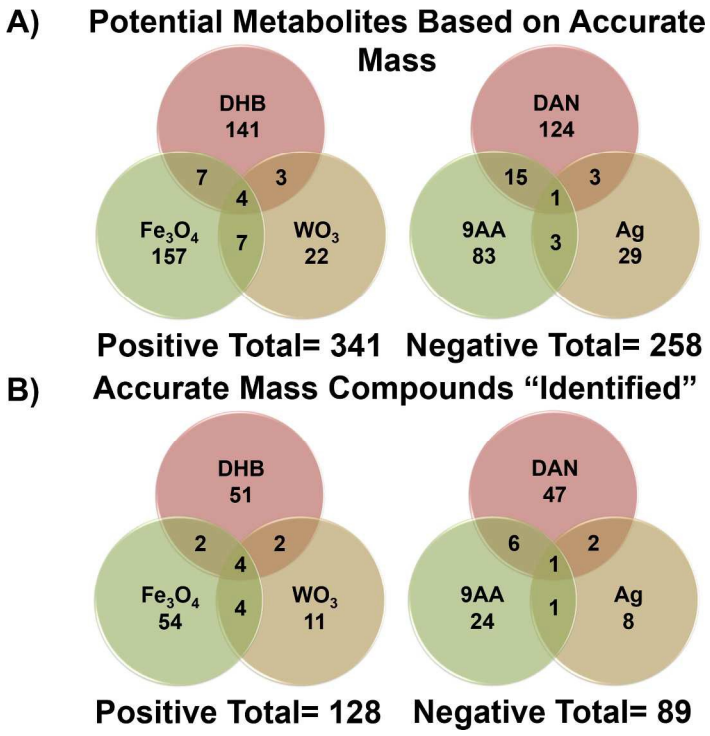


Figure 6. Venn diagrams displaying A) number of potential compounds based only on accurate mass, and B) number of potentially identified compounds from accurate mass search against the Metlin database.

Localization of Metabolites in Energy-related Metabolic Pathways

Once we identify hundreds of compounds, the next goal would be understanding the data in the context of metabolic biology. Quantitative analysis is often essential for this purpose in typical metabolomics analysis. Quantifying hundreds of compounds directly on the tissue in MSI, however, is a daunting task; yet, relative quantification could be done comparing different biological states. In the present work, we demonstrate how MSI can be utilized to visualize localization information of metabolic pathways. Two metabolic

pathways closely related with energy production, glycolysis and the tricarboxylic acid (TCA) cycle, are demonstrated here as examples. These two pathways are biologically relevant as both of them produce energy needed during seed development, while glycolysis also provides the base units for macromolecule generation and the starting materials for the TCA cycle.[37]

We could visualize the distributions of nine of the ten metabolites involved in glycolysis as shown in Figure 7. It should be noted that there are some ambiguities in these images. Glucose 6-phosphate (G6P), fructose 6-phosphate (F6P), and fructose 1,6-bisphosphate (F 1,6BP) are confidently identified through MS/MS, but others are solely based on accurate mass in the precursor spectra. Additionally, G6P/F6P and 2&3PG/PEP are structural isomers and cannot be easily distinguished, even in MS/MS. Thus, the images of G6P/F6P and 2&3PG/PEP shown in Figure 7 are the same. For the same reason, 2- and 3-phosphoglycerate (2PG, 3PG) are essentially identical.

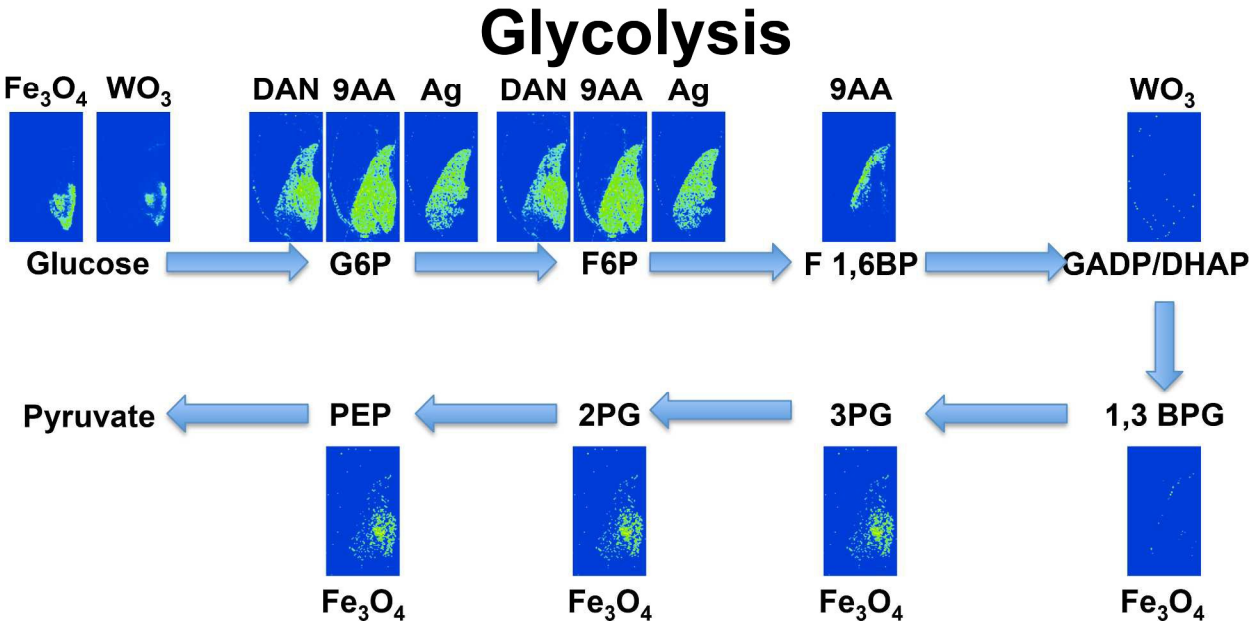


Figure 7. The images of metabolites involved in the glycolysis pathway. Compounds in negative mode are shown as deprotonated species. In positive mode, glucose,

glyceraldehyde 3-phosphate (GADP)/dihydroxyacetonephosphate (DHAP), 1,3 bisphosphoglycerate (1,3 BPG), and phosphoenolpyruvate (PEP) are seen as potassium adducts and 2- and 3-phosphoglycerate (2PG, 3PG) are seen as potassiated water loss species.

It is clear from the figure that the use of various matrixes and both polarities is essential to visualize as many compounds as possible in a metabolic pathway. Overall, metabolites involved in the glycolysis metabolic pathway appear to be mostly enriched in the embryo, which is not surprising as metabolic biology is most active when new tissues are being generated such as the radicle (root) and coleoptile (shoot). However, we could see some localization differences between the metabolites. Glucose looks to be localized predominantly to the radicle while fructose 1,6-bisphosphate seems to be more localized to the scutellum of the seed. Meanwhile, compounds like the phosphate sugars and phosphoglycerate compounds seem to be homogeneously distributed throughout the entire embryo region of the seed.

There are several possible explanations why some metabolites show different localization in the same metabolic pathway. For example, glucose is involved not only in glycolysis but also as a building block of many molecular components, which would be especially needed in newly developing tissues such as radicles. Pyruvate was not detected in this study and the images for GADP/DHAP and 1,3BPG were also not clear, potentially due to their low abundance and/or chemical instability.

Figure 8 shows eight out of the nine metabolites in the TCA Cycle, with all of them observed only in negative mode. This is not surprising as all the compounds in the pathway are organic acids, other than succinyl-CoA which is the only non-observed species in the

pathway. Malate, citrate, cis-aconitate, and isocitrate are confirmed with MS/MS, but others are assigned based only on accurate masses.

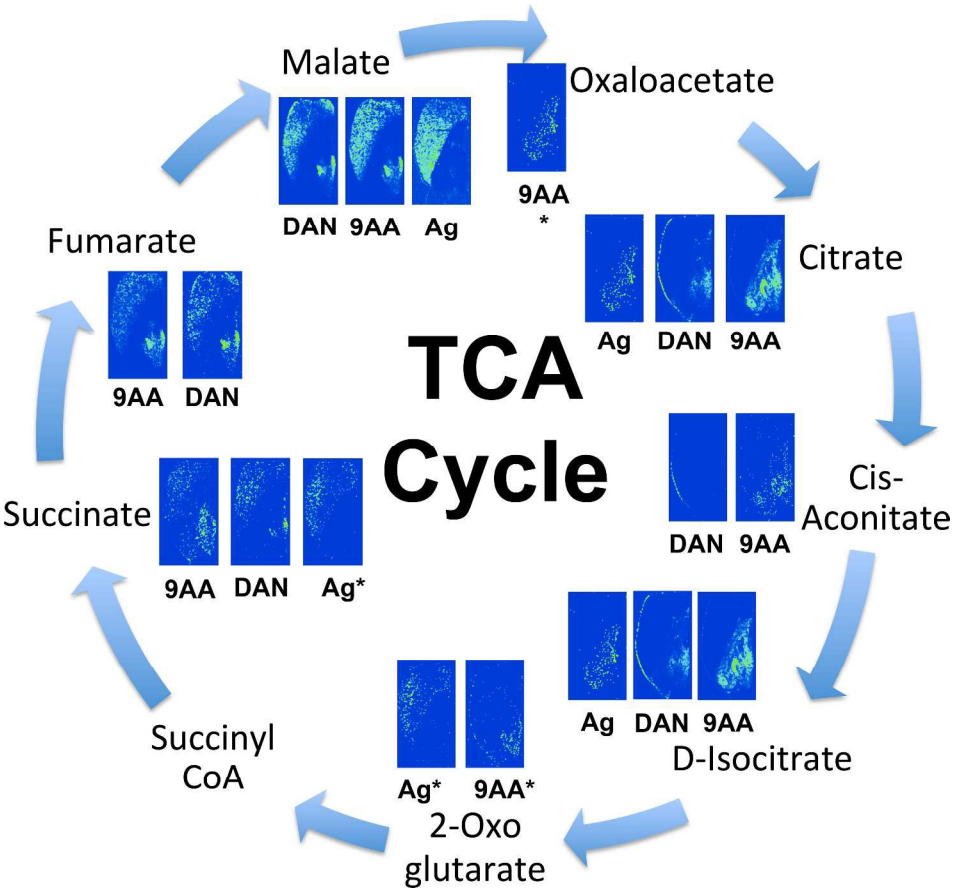


Figure 8. The images of metabolites involved in the TCA Cycle. Ions are all observed in negative mode as deprotonated species, except for those with asterisks, which are shown as deprotonated water loss species.

Similar to Figure 7, most of the metabolites appear to be localized in the embryo region of the seed. However, some small organic acid compounds (fumarate, succinate, malate) are also observed in the endosperm of the seed. This fact is surprising as typically the endosperm of the seed serves only as an energy source where starch and proteins are broken down for use by the developing seed.[38] According to an in-parallel GC-MS analysis (data not shown; a separate manuscript is in preparation to investigate metabolomic

changes at various germination time points), malate is in fact present both in the embryo and endosperm in significant levels. This may indicate that parts of the TCA cycle do occur in the endosperm of the seed, or they could be localized in those areas as parts of other metabolic processes.

Conclusions

A multi-matrix, dual polarity, tandem mass spectrometry imaging strategy is proposed to perform MSI in a large, metabolomics scale. This approach is applied to a germinated corn seed, visualizing a total of 599 potential metabolite ions based solely on accurate mass. Over 90,000 MS/MS spectra were acquired in six multiplex MS imaging data sets, among which at least 166 are unique ions with high quality MS/MS spectra that are potentially identifiable if we have an appropriate MS/MS library. Among those, 66 ions were confidently identified through manual MS/MS interpretation, with 52 of them being unique compounds. We were able to further demonstrate the utility of this approach by visualizing metabolites involved in the glycolysis and TCA pathways.

Limited by matrix selectivity, MALDI is often considered as a targeted analysis tool. As a result, MALDI-MSI has been limited to visualizing only certain classes of compounds. The current work intends to shift the paradigm and demonstrate MALDI-MSI can be used to visualize metabolites in a large scale. The proposed approach has several limitations compared to conventional untargeted metabolomics analysis. Most of all, due to the lack of chromatographic separation, metabolomic coverage and confidence in identification cannot reach those of LC-MS/MS. These limitations were partly overcome in this approach using natural localization of metabolites and matrix-specific analyte selectivity as ways of

separation. Future study will include 1) developing many more matrixes to widen the classes of metabolites that can be efficiently ionized, especially for low abundance compounds, 2) optimizing MS/MS conditions for small molecules, and 3) establishing an ion trap MS/MS database for confident identification and high-throughput analysis. Once successfully accomplished, we envision the proposed approach could be a powerful tool to fully understand metabolic biology with high-spatial localization information.

Acknowledgements

This work was supported by the US Department of Energy (DOE), Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences and Biosciences. The Ames Laboratory is operated by Iowa State University under DOE Contract DE-AC02-07CH11358. A special thanks to Drs. Javier Vela and Marna Yandau-Nelson from Iowa State University for providing materials.

References

1. B. Spengler, M. Hubert and R. Kaufmann, 1994.
2. R. M. Caprioli, T. B. Farmer and J. Gile, *Analytical Chemistry*, 1997, 69, 4751-4760.
3. A. Svatoš, *Trends in biotechnology*, 2010, 28, 425-434.
4. E. M. Weaver and A. B. Hummon, *Advanced drug delivery reviews*, 2013, 65, 1039-1055.
5. L. A. McDonnell and R. Heeren, *Mass spectrometry reviews*, 2007, 26, 606-643.
6. E. R. A. van Hove, D. F. Smith and R. M. Heeren, *Journal of chromatography A*, 2010, 1217, 3946-3954.
7. G. Marko-Varga, T. E. Fehniger, M. Rezeli, B. Döme, T. Laurell and Á. Végvári, *Journal of proteomics*, 2011, 74, 982-992.
8. H. Ye, E. Gemperline, M. Venkateshwaran, R. Chen, P. M. Delaux, M. Howes‐Podoll, J. M. Ané and L. Li, *The Plant Journal*, 2013, 75, 130-145.
9. R. A. Dixon and D. Strack, *Phytochemistry*, 2003, 62, 815-816.
10. R. Goodacre, S. Vaidyanathan, W. B. Dunn, G. G. Harrigan and D. B. Kell, *Trends in biotechnology*, 2004, 22, 245-252.
11. X.-L. Shu, T. Frank, Q.-Y. Shu and K.-H. Engel, *Journal of agricultural and food chemistry*, 2008, 56, 11612-11620.
12. U. Roessner, C. Wagner, J. Kopka, R. N. Trethewey and L. Willmitzer, *The Plant Journal*, 2000, 23, 131-142.
13. Y. Tikunov, A. Lommen, C. R. de Vos, H. A. Verhoeven, R. J. Bino, R. D. Hall and A. G. Bovy, *Plant Physiology*, 2005, 139, 1125-1137.
14. D. Perdian and Y. J. Lee, *Analytical chemistry*, 2010, 82, 9393-9400.
15. A. R. Korte and Y. J. Lee, *Journal of the American Society for Mass Spectrometry*, 2013, 24, 949-955.
16. J. D. Bewley, *The plant cell*, 1997, 9, 1055.
17. A. T. Klein, G. B. Yagnik, J. D. Hohenstein, Z. Ji, J. Zi, M. D. Reichert, G. C. MacIntosh, B. Yang, R. J. Peters and J. Vela, *Analytical chemistry*, 2015.
18. A. R. Korte, G. B. Yagnik, A. D. Feenstra and Y. J. Lee, in *Mass Spectrometry Imaging of Small Molecules*, Springer, 2015, pp. 49-62.
19. J. A. Hankin, R. M. Barkley and R. C. Murphy, *Journal of the American Society for Mass Spectrometry*, 2007, 18, 1646-1652.
20. J. H. Jun, Z. Song, Z. Liu, B. J. Nikolau, E. S. Yeung and Y. J. Lee, *Analytical chemistry*, 2010, 82, 3255-3265.
21. G. Robichaud, K. P. Garrard, J. A. Barry and D. C. Muddiman, *Journal of the American Society for Mass Spectrometry*, 2013, 24, 718-721.
22. A. l. Thomas, J. L. Charbonneau, E. Fournaise and P. Chaurand, *Analytical chemistry*, 2012, 84, 2048-2054.
23. A. R. Korte and Y. J. Lee, *Journal of Mass Spectrometry*, 2014, 49, 737-741.
24. R. L. Vermillion‐Salsbury and D. M. Hercules, *Rapid communications in mass spectrometry*, 2002, 16, 1575-1581.
25. M. Burrell, C. Earnshaw and M. Clench, *Journal of experimental botany*, 2007, 58, 757-763.

26. F. Benabdellah, D. Touboul, A. Brunelle and O. Laprévote, *Analytical chemistry*, 2009, 81, 5557-5560.
27. S. N. Jackson, K. Baldwin, L. Muller, V. M. Womack, J. A. Schultz, C. Balaban and A. S. Woods, *Analytical and bioanalytical chemistry*, 2014, 406, 1377-1386.
28. T. Hayasaka, N. Goto-Inoue, N. Zaima, K. Shrivas, Y. Kashiwagi, M. Yamamoto, M. Nakamoto and M. Setou, *Journal of the American Society for Mass Spectrometry*, 2010, 21, 1446-1454.
29. J. Schiller, J. Arnhold, S. Benard, M. Müller, S. Reichl and K. Arnold, *Analytical biochemistry*, 1999, 267, 46-56.
30. K. Strupat, M. Karas and F. Hillenkamp, *International Journal of Mass Spectrometry and Ion Processes*, 1991, 111, 89-102.
31. D. Harvey, *Rapid Communications in Mass Spectrometry*, 1993, 7, 614-619.
32. K. R. Jonscher and J. R. Yates, *Analytical biochemistry*, 1997, 244, 1-15.
33. S. N. Jackson, H.-Y. J. Wang and A. S. Woods, *Journal of the American Society for Mass Spectrometry*, 2005, 16, 2052-2056.
34. N. E. Manicke, J. M. Wiseman, D. R. Ifa and R. G. Cooks, *Journal of the American Society for Mass Spectrometry*, 2008, 19, 531-543.
35. L. D. Williams, A. E. Glenn, A. M. Zimeri, C. W. Bacon, M. A. Smith and R. T. Riley, *Journal of agricultural and food chemistry*, 2007, 55, 2937-2946.
36. J. J. Salas, J. E. Markham, E. Martínez-Force and R. Garcés, *Journal of agricultural and food chemistry*, 2011, 59, 12486-12492.
37. W. C. Plaxton, *Annual review of plant biology*, 1996, 47, 185-214.
38. D. Yan, L. Duermeyer, C. Leoveanu and E. Nambara, *Plant and Cell Physiology*, 2014, pcu089.