

Article

Metabolite Profiling of Soybean Seed Extracts From Near Isogenic Low and Normal Phytate Lines Using Orthogonal Separation Strategies

Judith Jervis, Christin Kastl, Sherry Hildreth, Ruslan Biyashev,
Elizabeth A. Grabau, Mohammad A. Saghai-Maroofo, and Richard Helm

J. Agric. Food Chem., **Just Accepted Manuscript** • DOI: 10.1021/acs.jafc.5b04002 • Publication Date (Web): 21 Oct 2015

Downloaded from <http://pubs.acs.org> on October 25, 2015

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.



ACS Publications

Journal of Agricultural and Food Chemistry is published by the American Chemical Society, 1155 Sixteenth Street N.W., Washington, DC 20036
Published by American Chemical Society. Copyright © American Chemical Society. However, no copyright claim is made to original U.S. Government works, or works produced by employees of any Commonwealth realm Crown government in the course of their duties.

**Metabolite Profiling of Soybean Seed Extracts From Near Isogenic Low and Normal
Phytate Lines Using Orthogonal Separation Strategies**

Judith Jarvis¹, Christin Kastl², Sherry B. Hildreth^{1,3}, Ruslan Biyashev², Elizabeth A. Grabau⁴,
Mohammad A. Saghai Maroof², and Richard F. Helm^{*1}

Departments of Biochemistry¹, Crop and Soil Environmental Sciences², Biological Sciences³,
and Plant Pathology, Physiology and Weed Science⁴, Virginia Tech, Blacksburg, VA, 24061

***Corresponding Author**

Contact Information:

Life Sciences 1
970 Washington Street, SW
Blacksburg, VA 24061-0910

E-mail: helmrf@vt.edu
Phone: 540-231-4088
FAX: 540-231-4043

ABSTRACT

Untargeted metabolomic profiling using liquid chromatography-mass spectrometry (LC-MS) was applied to lipid-depleted methanolic extracts of soybean seeds utilizing orthogonal chromatographic separations (reversed-phase and hydrophilic interaction) in both positive and negative ionization modes. Four near isogenic lines (NILs) differing in mutations for two genes encoding highly homologous multi-drug resistant proteins (MRPs) were evaluated. The double mutant exhibited a low phytate phenotype whereas the other three NILs, namely the two single mutants and the wild type, did not. Principal component analysis (PCA) of the four LC-MS datasets fully separated the low phytate line from the other three. While the levels of neutral oligosaccharides were the same for all lines, there were significant metabolite differences residing in the levels of malonyl isoflavones, soyasaponins, and arginine. Two methanol-soluble polypeptides were also found as differing in abundance levels, one of which was identified as the allergen Gly m 1.

KEYWORDS

Glycine max, hydrophobic seed protein, metabolomics, oligosaccharides, phytate, seed, soyasaponin

Introduction

Domesticated soybeans, *Glycine max* (L.) Merr., are one of the most valuable crops in the world, with a worldwide economic contribution of approximately \$48.6 billion.¹ The seeds are a source of oil and protein for both human and animal consumption, as well as industrially for a wide variety of end uses including adhesives, biodiesel, printing inks and structural materials. Over 85 million acres of soybeans were planted in the US in 2015, making it second only to corn in total acreage sown.²

The seeds of soybeans, cereal grains and other legumes contain high concentrations of phytic acid (phytate, *myo*-inositol 1,2,3,4,5,6-hexakisphosphate, InsP₆). Phytate is naturally found in its salt form and serves to store phosphate, *myo*-inositol and cations in seeds, substances that are mobilized upon germination.³ As phytate can chelate cations such as Ca⁺², Mg⁺², Zn⁺², and Fe⁺²/Fe⁺³ as well as protein, its presence in food and feedstuffs can limit the bioavailability of cations and/or protein when consumed by either humans or livestock.⁴ This anti-nutritional effect decreases feeding efficiencies for livestock and hence excretion of nutrients into the environment. While phosphate and/or phytase supplementation are options to meet nutritional demands, especially in non-ruminants such as pigs, poultry and fish, their use results in increased costs that must be balanced against the anti-nutritional losses.⁵

The negative effects of high phytate levels in soybean seed end use products has led to the development of low phytate lines.^{6, 7} ‘LR33’ and ‘V99-5089’ are low phytate soybean lines containing a mutation in the gene encoding *myo*-inositol 3-phosphate synthase (*MIPS1*).^{7, 8} This mutation restricts flow through the canonical phytic acid biosynthetic pathway, leading to a reduction in seed phytate and an increase in inorganic phosphate.⁷ A third low phytate line, ‘CX1834’,^{6, 9, 10} is the result of single base mutations in two multidrug resistance-associated

proteins (*MRPs*) located on chromosomes 19 (linkage group L, Glyma19g35230) and 3 (linkage group N, Glyma03g32500).¹⁰ The two genes encode proteins that are annotated as ATP-binding cassette, sub-family C, member 5 proteins (UniProt IDs: K7MYS3, I1JP84) and are 95% identical to one another (Supplemental Fig. S1). This class of conserved transmembrane proteins couple ATP hydrolysis to metabolite movement across a membrane, with the closest ortholog in *Arabidopsis* (*AtABCC5*) associated with multiple processes including phytate transport,¹¹ transition from primary to lateral root formation and drought resistance.¹² While it is established that the mutation of the two ATP-coupled transport proteins leads to both low phytate and low emergence,¹³⁻¹⁶ the biochemical mechanism behind the linked phenotypes and the two *MRPs* has yet to be established in soybean.⁴

Systems level insights into plant growth and development have been gained through the use of coupled chromatographic and mass spectrometric technologies (GC-MS, LC-MS) combined with chemical informatics tools.¹⁷ With respect to seed metabolomics, GC-MS based metabolite analysis was used to compare a low phytate mutant line and wild-type rice lines,¹⁸ as well as individual low and normal phytate maize kernels grown on the same ear.¹⁹ The analysis of two related salt-sensitive and salt-tolerant soybean lines by LC-MS permitted discrimination between the two genotypes, with genistin and Group B saponins being correlated with salt tolerance.²⁰ Sawada and Harai recently described the collection of over 40,000 MS/MS spectra collected from soybean seed (93 recombinant inbred lines), which were linked with data from quantitative trait loci (QTL).²¹ These individual MS² datafiles were uploaded (without identifications) as independent Accession Numbers to the publically available MS2T database.²² Additional analyses of soybean seeds in relation to cultivar differences have also been published.^{23, 24} Finally, in an untargeted GC-MS analysis of low phytate soybean seed extracts

from MIPS pathway mutants, it was shown that levels of non-reducing oligosaccharides and cyclitols, compounds downstream of the MIPS pathway, were decreased relative to wild type controls.²⁵ There are currently no studies available that have evaluated soybean low phytate lines generated by the two MRP mutations.

Our analyses centered on four near isogenic lines (NILs) differing in the two *MRP* genes associated with the low phytate and emergence phenotype. Of these four lines, only the double mutant would be low in phytate. As untargeted metabolomics is sensitive to both environmental conditions and genetic differences, the near isogenic lines provided an opportunity to minimize the noise associated with genetic background. We first removed the lipophilic fraction by extraction using dry ethyl acetate (analysis of this fraction will be the subject of another report). The lipid-depleted seed powder was extracted with MeOH:0.1% aqueous HOAc (9:1, v/v) with the resulting extract submitted to both reversed-phase and HILIC-based chromatographic separations combined with both positive and negative ion mode detection, thereby generating four LC-MS datasets per sample. The usage of orthogonal separation methods with positive and negative ion modes provided datasets to assess the roles *MRP* gene mutations have in generating the low phytate phenotype. These datasets have been made publically available (MTBLS120).²⁶

MATERIALS AND METHODS

Plant Material and Harvest. The NILs investigated in this study were developed from a cross of the low-phytate soybean lines CX1834-1-6 (hereafter CX1834, MRP mutations) and V99-5089 (MIPS mutation). A single plant that was heterozygous for the mutations in the two *MRP* genes and homozygous for the wild type allele of the *MIPS1* gene was selected from an advanced generation recombinant inbred line (RIL) population of this cross. The plant was

selfed to generate the four near isogenic lines (I) *mrp(3)/mrp(3)/mrp(19)/mrp(19)* (homozygous for mutations in both *MRP* genes, referred to here as the double mutant; DM), (II) *mrp(3)/mrp(3)/MRP(19)/MRP(19)* (homozygous for the *MRP* gene mutation on chromosome 3, and wildtype for *MRP* gene on chromosome 19, referred to here as single mutant SM3), (III) *MRP(3)/MRP(3)/mrp(19)/mrp(19)* (homozygous for the *MRP* mutation on chromosome 19, referred to here as single mutant SM19) and (IV) *MRP(3)/MRP(3)/MRP(19)/MRP(19)* (wildtype for both *MRP* genes, referred to here as wildtype, WT).

Plants were grown in Blacksburg, Virginia in 2012. For each class, ten seeds from the 2011 harvest were hand planted in one row. The lines were developed from a cross of CX1834 (Maturity Group III) and V99-5089 (Maturity Group V). All four lines had similar maturity dates, which were closer to that of V99-5089. Fully matured seeds were harvested from all plants of each row and bulked together. Seeds were stored at 4 °C and then used for metabolomic analysis and determination of phytate levels.

Nucleic Acid Methodologies. Confirmation of the *MRP* genotypes was performed on DNA extracted from trifoliolate leaf tissue using the CTAB method and submitted to the Competitive Allele-Specific PCR genotyping system (KASPar, Kbioscience) according to the methods of the manufacturer. All 3 sets of SNP primers (MIPS1, *MRP*-L, *MRP*-N) were designed, validated and synthesized by KBioscience based on DNA sequences 100 nucleotides in length with designated SNPs directly after the first 50 nucleotides. The template sequences containing the signature SNPs were obtained from respective genomic DNA sequences of MIPS1⁸, *MRP*-L and *MRP*-N^{9, 10} genes. Extracted DNA (4 µL, 15ng) was mixed with Reaction Mix (4µL, 2x) and an Assay solution (0.11 µL) that contained two allele-specific primers (one for each SNP allele)

with an unlabeled tail sequence, one common reverse primer, two fluorescently-labeled oligos and two complimentary quencher-labeled oligos. The reaction mix contained Taq polymerase enzyme and 5-carboxy-X-rhodamine (passive reference dye), succinimidyl ester, MgCl_2 and DMSO. A touch-down PCR profile specific for amplification of *MRP* and *MIPSI* SNPs was used: 1 cycle 94°C (15min), 61°C (60s); 1 cycle 94°C (20s), 60.4°C (60s); 1 cycle 94°C (20s), 59.8°C (60s); 1 cycle 94°C (20s), 59.2°C (60s), 1 cycle 94°C (20s), 58.6°C (60s), 1 cycle 94°C (20s), 58°C (60s), 1 cycle 94°C (20s), 57.4°C (60s), 1 cycle 94°C (20s), 56.8°C (60s), 1 cycle 94°C (20s), 56.2°C (60s), 1 cycle 94°C (20s), 55.6°C (60s), 35 cycles 94°C (20s), 55°C (60s). Confirmation of the Sg-1 alleles for the four NILs and the parental lines (CX-1834 and V99-5089) was based upon amplification of DNA extracted from trifoliolate leaf tissue using the CTAB method. Sg-1:151 primer, designed by Sayama *et al.*,²⁷ was used for amplification and sequencing. Sequencing was performed at the Virginia Bioinformatics Institute, Blacksburg, VA.

Phytate Content. Approximately 75 seeds of each of the four classes (2012 year of harvest) were ground with a Cyclone Sample Mill with a 0.5 mm mesh screen (UDY Corporation, Fort Collins, CO). Phytate concentrations were determined in triplicate according to the protocol of Burleson *et al.*²⁸ and reported as mg/g dry weight of seed.

Seed Extract Preparation. Solvents used were LC-MS quality (Spectrum Chemicals) with the exception of ethyl acetate, which was HPLC-grade (Fisher Scientific) and dried with anhydrous MgSO_4 powder before use. LC-MS grade organic acids (formic and acetic) were from Sigma-Aldrich. The four NILs were analyzed in biological triplicates using 5 randomly selected seeds for each replicate. Seeds were flash-frozen in liquid nitrogen and finely ground with P14 mill (Pulverisette 14, Fritsch). The powder was then transferred to pre-weighed 15 mL tubes,

weighed and stored at -80 °C. A subset (400 mg) of this powder was dried overnight on a high-vacuum line and the non-polar components were extracted using dry ethyl acetate (7 mL). The extraction procedure was repeated twice with a sequence of vortexing, sonication for 20 min and centrifugation (1680 x g for 15 min). The supernatants were combined, concentrated to oils and stored at -80 °C. The remaining ethyl acetate was removed from the soybean powder on the high-vacuum line and stored at -80 °C.

The polar metabolites were obtained from two 30 mg subsets of the dried lipid-free powders, each extracted with MeOH:0.1% aqueous HOAc (0.5 mL; 9:1, v/v). One subset was used for reversed-phase chromatography-mass spectrometry (RP-MS) and the other for hydrophilic interaction liquid chromatography-mass spectrometry (HILIC-MS). Internal standards were added to both extraction solvents, with the HILIC extracts employing ¹³C-labeled L-arginine (¹³C₆, Thermo Scientific, 0.09 µg/µl final concentration before extraction) and the RP extracts employing deuterated L-tryptophan (indole-d₅, Cambridge Isotope Labs, 0.05 µg/µl final concentration before extraction). After vortexing and sonication for 20 min, the samples were centrifuged (1680 x g, 15 min). This extraction was repeated twice and the extracts were pooled, reduced in volume with the aid of a centrifugal concentrator under vacuum (35 °C) and subsequently taken to dryness on a high-vacuum line.

Liquid Chromatography-Mass Spectrometry. Analyses were performed with an Acquity I-class UPLC interfaced with Synapt G2-S HDMS (Waters) in both positive and negative ion modes. Master mixes were prepared for each class (DM, WT, SM3, SM19) by combining aliquots (10 µL each) of each of the three biological replicates, with a complete master mix of all classes created by combining aliquots (10 µL) of the four class master mixes. An analysis "set"

consisted of 3 blank injections followed by 3 complete master mix injections (for column conditioning). A random set of all samples (17) plus a blank injection was then repeated three times in the same randomized order, which was generated at <https://www.random.org>. This was followed by a set of 4 complete master mix injections in MS^E mode, each at different collision energies (10, 20, 30, 40 V).

Reversed-Phase Separations. Dried extracts were reconstituted with 0.1% aqueous formic acid: acetonitrile (MeCN, 9:1, v/v, 120 μ L). Samples were briefly vortexed and sonicated for 10 min, followed by centrifugation (13,000 x g, 10 min, RT). An aliquot (10 μ L) was transferred to an LC-MS-grade vial and diluted with 90 μ L of 0.1% aqueous formic acid:MeCN (9:1, v/v). Sample separation was achieved with a binary solvent system of 0.1% formic acid (A) and MeCN (B) utilizing an Acquity UPLC BEH C18 column (1.7 μ m, 2.1 mm x 50 mm, Waters Corp., Milford, MA) with a flow rate of 200 μ L/min and a 15 minute gradient. The following gradient conditions were used: isocratic at 5% B (0-1 min), followed by linear gradient to 15% B (1-2 min), to 95% B (2-11 min), isocratic at 95% B (11-12.5 min), followed by return to initial conditions (12.5-15 min). Injection volume into the column was 2 μ L. The separated samples were ionized by electrospray ionization and analyzed in both positive and negative modes. The scan time was set to 0.20 sec and a mass range of 50-1800 m/z was scanned. The source parameters for positive ion mode were source temperature 125 °C, capillary voltage 3.0, cone voltage 70, source offset 80, desolvation temperature 300 °C, cone gas 50 L/h, desolvation gas 500 L/h and nebulizer gas 6.0 bar. The source parameters for negative ion mode were source temperature 125 °C, capillary voltage 2.4, cone voltage 40, source offset 80, desolvation temperature 300 °C, cone gas 50 L/h, desolvation gas 500 L/h and nebulizer gas 6.0 bar. A

reference sprayer continuously infused leucine-enkephalin (200 ng/mL, Waters Corp.) at 5 μ L/min with a scan frequency of 20 seconds.

HILIC Separations. Dried extracts were reconstituted with 0.1% aqueous formic acid:acetonitrile (1:1, v/v, 100 μ L). Samples were briefly vortexed and sonicated for 10 minutes, followed by centrifugation (13,000 x g, 10 min, RT). An aliquot (10 μ L) was transferred to a vial and diluted with 90 μ L of 0.1% aqueous formic acid:MeCN (1:1, v/v). Chromatography was performed on an Acquity UPLC BEH Amide Column (1.7 μ m, 2.1 mm x 50 mm, Waters Corp.) with a flow rate 400 μ L/min and a 10 minute gradient prepared from mobile phase A (0.1% aq. formic acid) and mobile phase B (acetonitrile). Gradient conditions: isocratic at 99% B (0-0.5 min), followed by linear gradient to 30% B (0.5-7 min), to 99% B (7-7.10 min), isocratic at 99% B (7.10-10 min), followed by returning to the initial conditions. The injection volume was 1 μ L. Column eluent was ionized by electrospray ionization in both positive and negative modes. The scan time was set to 0.20 sec and a mass range of 50-1800 m/z was scanned. The source parameters for positive ion mode were source temperature 120 $^{\circ}$ C, capillary voltage 3.0, cone voltage 30, source offset 80, desolvation temperature 500 $^{\circ}$ C, cone gas 50 L/h, desolvation gas 600 L/h and nebulizer gas 6.0 bar. The source parameters for negative ion mode were source temperature 120 $^{\circ}$ C, capillary voltage 2.2, cone voltage 30, source offset 80, desolvation temperature 500 $^{\circ}$ C, cone gas 50 L/h, desolvation gas 600 L/h and nebulizer gas 6.0 bar. A reference sprayer continuously infused leucine-enkephalin (1 ng/ μ L, Waters Corp., Milford, MA) at 5 μ L/min with a scan frequency of 20 sec.

Data Processing and Analysis. The MarkerLynx software (version 4.1, Waters Corp.) was used to process the raw UPLC/Q-TOF-MS data. The MarkerLynx parameters for the analyses of the

RP runs were: retention time range 2.0 – 9.0 min, mass 50 – 2000 m/z, mass window 0.05, retention time window of 0.15, noise elimination of level 4, peak intensity threshold of 10000, marker intensity threshold 2400 (positive ion mode) and 4800 (negative ion mode). The MarkerLynx parameters for the analyses of the HILIC runs were: retention time range 2.0 – 13.0 min, mass 50 – 2500 m/z, mass window 0.02, retention time window of 0.15, noise elimination of level 10, peak intensity threshold of 1000, marker intensity threshold 2400 (positive ion mode) and 5000 (negative ion mode). MarkerLynx generated a data matrix consisting of all exact mass-retention time pairs (EMRTs) found in the datasets along with their peak areas. The EZinfo 2.0 software (Umetrics) was used to conduct principal component analyses (PCA) and orthogonal partial least squared discriminate analyses (OPLS-DA) of the EMRT datasets. The latter was visualized using a score plot. The datasets were converted to spreadsheets to permit determination of p-values (t-test) and factors of change (peak area ratios). Peaks were identified with the help of commercial standards, available on-line databases such as Metlin (metlin.scripps.edu) and PRIME/MS2T,²² from the literature, and/or manual interpretation of MS/MS fragmentation patterns. Gradient and source parameters for MS/MS runs were identical to their MS runs, except the collision energy was applied in trap and ramped from 5 to 40 eV.

220

221 RESULTS AND DISCUSSION

222 **Phytate Concentrations and Metabolite Profiling Overview.** The mean phytate content of the
223 seeds harvested in 2012 are shown in Table 1, and were within the normal ranges reported for
224 low and high phytate lines.²⁹ The phytate values for the parents of these lines were 7.62 mg/g
225 (V99-5089) and 4.78 mg/g (CX1834); thereby both parents are considered to be low phytate.

The line with both mutations (DM, Table 1) exhibited the lowest phytate concentration whereas the other three showed a normal phytate phenotype (p -value < 0.02).

UPLC-MS methods were developed for the detection of methanol-soluble metabolites obtained from lipid-depleted, freeze-dried soybean powder. These extracts were analyzed by two orthogonal chromatographic methods using short gradient separations to minimize LC-MS time. Triplicate random injections for all formats in both positive and negative ion modes of biological triplicates resulted in a total of 144 (36 per column, per ion mode) individual LC-MS runs with the data collected in MS¹ mode. Along with these runs were triplicate injections of “master mixes” of each line; samples generated by combining equal volumes of each sample for each line into one vial. The final set of analyses in each mode utilized a master mix of all classes. This sample was submitted to fragmentation experiments conducted in data-independent acquisition mode (DIA, MS^E). The master mix data collected in MS¹ mode was employed during the statistical analysis phase of the work (data validation) and the fragmentation experiments aided in compound identification. Example UPLC-MS¹ chromatograms can be found in the Supplemental Materials (Figs. S2 and S3).

The MS¹ datasets were submitted to analysis with MarkerLynx (Waters) in order to convert the raw data into exact mass–retention time pairs (EMRTs). The number of EMRTs detected depended on the peak detection thresholds (see Materials and Methods), which were set in a stringent manner to limit noise and minimize the number of ions for downstream analysis. A total of 154 EMRTs were detected in reversed-phase positive mode and 355 EMRTs for negative mode. The HILIC-based separations resulted in 1066 EMRTs for positive-ion mode and 1221 EMRTs in negative mode (Table 2). To further limit the list of ions significantly different between the NILs; factor of change (peak area ratios) and p -values were determined for the

EMRTs in each dataset. EMRTs with p -values < 0.05 and a factor of change less than 0.5 or greater than 2 were selected for further analysis. The overlap between NIL, LC-MS format and identified EMRTs is shown as a Venn Diagram in Fig. 1. Considering only the EMRTs that contribute to the phytate phenotype, there were 24 in reversed phase positive ion mode (center of Venn diagram) and 141 in the negative ion mode. The HILIC interrogation led to the identification of 146 and 199 EMRTs in positive and negative ion modes, respectively. Many of these ions were derived from the same compound due to the presence of ion adducts and/or in-source decay. The full listing of ions associated with each region of Fig. 1 are provided in the Supplemental Materials (Figs. S4–S7), and the complete datasets for each mode are available at the MetaboLights website (Accession MTBLS120).²⁶

Principal Component Analyses and Ion Identification. Principal component analysis of each LC method-ionization mode pairing resulted in a clear separation of the three normal phytate lines from the low phytate line (Fig. 2), with the reversed-phase analyses providing over 70% of the variation in PC1 compared to 35% in the HILIC separations. Similar analyses of the three normal phytate lines displayed poorer overall class separations with the exception of the HILIC negative ion mode data (See Supplemental Fig. S8). As the ions identified as being significantly different between all three normal lines vs. the low phytate line may lend insight into the role of the MRPs in the low phytate and low emergence phenotype, we next utilized a combination of EMRT data, fragmentation patterns (MS/MS data), literature and database searches to provide the assignments shown in Table 3 for the reversed-phase datasets.

Malonyl daidzin and malonyl genistin were both decreased in the double mutant line. These metabolites are stored in vacuoles within seeds, and changes in their levels are known to

occur during the germination and emergence processes.^{30, 31} Both malonyl daidzin and malonyl genistin are selectively released almost immediately upon imbibition; a saturable event that occurs concomitantly with release of glucosidases that hydrolyze the glycosidic bond.³¹ Lower levels of these compounds may be related to low emergence rates but will require additional studies to evaluate more fully.

The most significant differences between the low and normal phytate lines were within the soyasaponin region (Fig. 3). The normalized base peak chromatograms from the reversed-phase separations in negative ion mode show that the soyasaponins are more abundant in the single MRP mutants, with overall profiles similar to that of the wild type line (for structural information, see Supplemental Table S1 and Fig. S9). The double mutant low phytate line (DM) exhibited a dramatically different profile, with the presence of at least three soyasaponins that were not present in the other three lines (Fig. 3). Structural evaluation determined that the low phytate line was enriched with C22-peracetylxylose terminated Group A soyasaponins, with severely reduced levels of C22-peracetylglucose terminated structures.

Metabolomics Informs Genomics. The soyasaponin composition of seeds is regulated by proteins encoded by five genes: *Sg-1*, *Sg-3*, *Sg-4*, *Sg-5* and *Sg-6* (on chromosomes 7, 10, 1, 15, 1, respectively), which control soyasapogenol A or B utilization and the sequence and types of sugar chains attached to the triterpene skeleton (Fig. S9). Such processes are differentially regulated depending on plant organ and variety.³²⁻³⁶ *Sg-1* controls terminal sugar deployment at the C-22 position of Group A soyasaponins.²⁷ Interestingly, this terminal glycosylation "choice" (acetylxylose, acetylglucose, or none) is controlled by multiple alleles of the *Sg-1* gene, which encodes a UDP-sugar-dependent glycosyltransferase (Glyma07g38460). Expression of one

allele (*Sg-1^a*) leads to UDP-acetylxylosyltransferase activity producing soyasaponin A4 whereas expression of *Sg-1^b* leads to UDP-acetylglucosyltransferase (UGT73F2) activity producing soyasaponin A1.²⁷ The two proteins have 98.3% identity with a single amino acid substitution (Ser/Gly-138) thought to be responsible for substrate specificity. A third allele, *Sg-1⁰*, is a loss of function allele and leads to soyasaponins devoid of the terminal acetylated sugar.²⁷

The dramatic change in soyasaponin profiles suggested that these near isogenic lines contained genetic differences in the *Sg-1* gene-coding region. In order to explore this possibility further, the *Sg-1* region of the parental lines (CX-1834 and V99-5089) as well as the four classes investigated here were amplified and sequenced using established *Sg-1* primers.²⁷ The sequencing results confirmed the metabolomic observations in that the low phytate line contained the sequence associated with the single amino acid substitution (*Sg-1a* allele, Ser/Gly-138) whereas the other three did not. The parent V99-5089 also contained the *Sg-1a* allele whereas the parent CX1834 encoded for the *Sg-1b* allele and hence glucose-terminated soyasaponins. Thus, the low phytate double mutant contains the *Sg-1a* allele, which was contributed by the V99-5089 parent, whereas the three normal phytate lines contained the *Sg-1b* allele, which was contributed by CX-1834. These differences were subsequently confirmed at the metabolite level by LC-MS analyses of both parent lines (Supplemental Fig. S10). While this result is an example of untargeted metabolomics work informing subsequent genomics efforts to confirm the alleles, it minimizes a direct link between phytate and emergence with Group A soyasaponins, as the CX-1834 line exhibits low phytate and emergence and encodes for the *Sg-1b* allele (glucosyl-terminated).

Hydrophobic seed protein differences. Two polypeptides extracted from the lipid-depleted powder were identified as significantly lower in the low phytate line in the positive-ion mode, reversed-phase separations. One of these polypeptides could not be assigned to a sequence, with the $[M+3H]^{+3}$ and $[M+4H]^{+4}$ monoisotopic ions calculated to be 1297.227 and 973.172, respectively (retention time 5.20 min, Table 3). The second polypeptide matched to the full-length form of hydrophobic seed protein (HPS, Glyma15g13750, UniProt ID P24337), more commonly referred to as the soybean allergen Gly m 1.³⁷⁻³⁹ Further analysis of the LC-MS data also led to the identification of the lower mass isoform in which the polypeptide is shortened by the loss of two amino acids at the *N*-terminus, with the $[M+6H]^{+6}$ form being the most abundant ion for each of these polypeptides. A subsequent analysis of the extracts from the low phytate parent lines also showed trace levels of HPS and the 3.9 kDa peptide relative to the normal phytate NILs (Supplemental Fig. S11), supporting the claim that reduced levels of these polypeptides are associated with the low phytate phenotype.

Hydrophobic seed protein is synthesized in the pod endocarp and deposited on the seed surface during maturation and has been associated with water uptake rates.⁴⁰ Both forms of HPS contain 8 cysteine residues, which form 4 disulfide bonds, generating the characteristic structure of a non-specific lipid transfer protein (nsLTP, Supplemental Fig. S12).⁴¹ Although the HPS structure is highly similar to that of plant nsLTPs, HPS is generally not included in discussions of this class of proteins due to the lack of several conserved moieties that are known to interact with lipids,⁴² even though the disulfide bonding pattern is a match to that of Type II nsLTPs.⁴³

Since HPS is hydrophobic and on the surface of the seed, it has been suggested to be involved in water absorption processes and/or seed-pathogen interactions.⁴⁰ A recent study of a similar apoplastic/extracellular sunflower protein HaAP10 (UniProt P82007) demonstrated that

this nsLTP is internalized upon water uptake, with evidence provided for its subsequent association with oil bodies and glyoxysomes.⁴⁴ The HaAP10 protein may be acting at the oil body-glyoxysome interface during seed germination and emergence, providing a means to transfer lipids to the glyoxosome during germination and emergence. Low levels of HPS may be due to lower biosynthetic capacity, the inability to transport the protein to the exterior of the developing seed, or changes in the seed coat that limit HPS binding.

HILIC-based Separations Enhance the Metabolome Coverage. Reversed-phase chromatography is generally not useful for hydrophilic compounds, which tend to elute at or near the solvent front, even under initial conditions of high aqueous mobile phase. As soybean seeds contain considerable amounts of hydrophilic compounds, separations that favor analysis of these compounds, in concert with the compounds identifiable by reversed-phase separations provide a fuller coverage of the seed metabolome relative to choosing only one separation platform and/or ionization mode. The orthogonal nature and ionization efficiencies of the two separation modes are exemplified in Fig. 4 for the γ -glutamyl dipeptides of phenylalanine and tyrosine, two abundant dipeptides in soybean seed.⁴⁵ The dipeptide γ -Glu-Tyr elutes before γ -Glu-Phe using reversed-phase chromatography but after under conditions of hydrophilic interaction chromatography. Relative ion intensities depend on both the ionization mode as well as the method of separation. The ion intensities of the two dipeptides in reversed-phase chromatography are two orders of magnitude less in positive ion mode than in negative ion mode, while the HILIC separations provide relatively similar intensities in each ionization mode.

As was seen in Table 2 and Fig. 1, HILIC-based separations provided more significantly changing EMRTs than reversed-phase. A high percentage of these ions do not represent

independent compounds, but are adducts or fragments of the same metabolite. For example, $[M+\text{formate}]^-$ and $[M+\text{Na}]^+$ are common ions found in negative and positive ion modes, respectively, along with their corresponding $[M-H]^-$ and $[M+H]^+$ ions. The combination of adducts and in-source decay led to several cases where there were more than 6 EMRTs per metabolite identified as being different across isogenic lines (See Supplemental, Figs. S3-S7).

Soybean seeds from low phytate lines associated with the MIPS mutations have reduced levels of raffinose-based oligosaccharides and galactosyl cyclitols.²⁵ While the HILIC separations provided a means to cleanly separate the predominant soybean seed neutral oligosaccharides (Fig. 5), their levels were not significantly different when comparing isogenic lines. Eight classes of oligosaccharides were identified, ranging from disaccharides to pentasaccharides, of which several were methylated. While the HILIC-MS system does not permit the detection of phytic acid, the approach can be used for studies evaluating oligosaccharide profiles, including those from low phytate MIPS mutant lines. That the MRP mutations did not result in changes in the oligosaccharide profile suggests that the mutations modulate transport processes unrelated to this compound class.

The compounds identified in both reversed-phase and HILIC trended in the same direction with the same relative ion abundance differences (Table 4), indicating the robust nature of the combined analysis approach. While data is provided for the soyasaponins, this class of compounds does not resolve as well in HILIC-based separations relative to reversed-phase. Interestingly, one hydrophilic compound that was identified as being in higher abundance in the double mutant line was arginine. Higher levels of the free amino acid arginine may be a result of changes in the partitioning of nitrogen during seed maturation, as a study aimed at reducing levels of the soybean seed storage protein β -conglycinin demonstrated increased levels of free

arginine in developing seeds.⁴⁶ Interestingly, an analysis of the parent lines showed that MRP low phytate CX1834 was extremely low in free arginine whereas the MIPS-mutant parent V99-5059 had levels in accordance to the WT, SM3 and SM19 (Supplemental Fig. S13).

In summary, orthogonal separations combined with positive and negative ionization modes resulted in the identification of several classes of compounds in four near isogenic soybean seeds, including soyasaponins, isoflavones, oligosaccharides, and amino acid-based compounds. When applied to low phytate mutant lines, there were no differences in oligosaccharide profiles indicating that the low phytate phenotype obtained by the MRP mutations does not affect pathways downstream of *myo*-inositol as has been observed for MIPS mutants.²⁵ Differences observed in soysaponins was linked to the parent lines and since both are low phytate, these compounds are probably not directly related to low phytate. Malonyl daidzin and malonyl genistin were both decreased in the low phytate line, which may be related to low emergence rates as these compounds are known to be excreted as part of the germination process. The relationship between HPS, phytate levels and emergence are worthy of further investigation.

Abbreviations Used: EMRT, exact mass/retention time pair; HILIC, hydrophilic interaction liquid chromatography; HPS, Hydrophobic Seed Protein; LC-MS, liquid chromatography-mass spectrometry; MRP, multi-drug resistance protein; NIL, near-isogenic line; PCA, principal components analysis;

Acknowledgements. The authors thank Neelam Redekar for DNA sequence analysis. Funding for this work was through the United Soybean Board as well as the John Lee Pratt Fellowship Program and the Bio-design and Bioprocessing Research Center (BBRC), both at Virginia Tech.

Additional funding was provided by the Fralin Life Science Institute, as well as the Virginia Tech Agricultural Experiment Station Hatch and McIntire-Stennis Programs.

Supporting Information Description

Soybean soyasaponin structural and monoisotopic mass information, alignment and structural overview of the two transport proteins, example chromatograms using reversed-phase and HILIC, EMRTs for all chromatography and ionization modes, PCA plots of the three normal phytate lines, additional soyasaponin analyses, structural details of hydrophobic seed protein (Gly m 1), additional LC-MS analyses including parental lines, and a link to the LC-MS data available at the MetaboLights Website. This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES

1. Wilson, R. F., Soybean: market driven research needs. In *Genetics and genomics of soybean*, Springer: 2008; pp 3-15.
2. U.S. Department of Agriculture, National Agricultural Statistics Service (NASS), *Acerage*; ISSN: 1949-1522; 2015.
3. Raboy, V., The ABCs of low-phytate crops. *Nat Biotechnol* **2007**, *25*, 874-875.
4. Raboy, V., Approaches and challenges to engineering seed phytate and total phosphorus. *Plant Sci* **2009**, *177*, 281-296.
5. Dersjant-Li, Y.; Awati, A.; Schulze, H.; Partridge, G., Phytase in non-ruminant animal nutrition: a critical review on phytase activities in the gastrointestinal tract and influencing factors. *J Sci Food Agr* **2015**, *95*, 878-896.
6. Wilcox, J. R.; Premachandra, G. S.; Young, K. A.; Raboy, V., Isolation of High Seed Inorganic P, Low-Phytate Soybean Mutants. *Crop Sci* **2000**, *40*, 1601.
7. Hitz, W. D.; Carlson, T. J.; Kerr, P. S.; Sebastian, S. A., Biochemical and molecular characterization of a mutation that confers a decreased raffinose and phytic acid phenotype on soybean seeds. *Plant Physiol* **2002**, *128*, 650-60.
8. Saghai Maroof, M. A.; Buss, G. R., *Low phytic acid, low stachyose, high sucrose soybean lines*. US Patent US20080199591 A1: 2008; p 14.
9. Gillman, J. D.; Pantalone, V. R.; Bilyeu, K., The Low Phytic Acid Phenotype in Soybean Line CX1834 Is Due to Mutations in Two Homologs of the Maize Low Phytic Acid Gene. *Plant Genome J* **2009**, *2*, 179.
10. Saghai Maroof, M. A.; Glover, N. M.; Biyashev, R. M.; Buss, G. R.; Grabau, E. A., Genetic basis of the low-phytate trait in the soybean line CX1834. *Crop Sci* **2009**, *49*, 69-76.
11. Nagy, R.; Grob, H.; Weder, B.; Green, P.; Klein, M.; Frelet-Barrand, A.; Schjoerring, J. K.; Brearley, C.; Martinoia, E., The Arabidopsis ATP-binding cassette protein AtMRP5/AtABCC5 is a high affinity inositol hexakisphosphate transporter involved in guard cell signaling and phytate storage. *J Biol Chem* **2009**, *284*, 33614-22.
12. Suh, S. J.; Wang, Y. F.; Frelet, A.; Leonhardt, N.; Klein, M.; Forestier, C.; Mueller-Roeber, B.; Cho, M. H.; Martinoia, E.; Schroeder, J. I., The ATP binding cassette transporter AtMRP5 modulates anion and calcium channel activities in Arabidopsis guard cells. *J Biol Chem* **2007**, *282*, 1916-24.
13. Anderson, B. P.; Fehr, W. R., Seed Source Affects Field Emergence of Low-Phytate Soybean Lines. *Crop Sci* **2008**, *48*, 929.

- 457 14. Gao, Y.; Biyashev, R. M.; Maroof, M. A. S.; Glover, N. M.; Tucker, D. M.; Buss, G. R.,
458 Validation of Low-Phytate QTLs and Evaluation of Seedling Emergence of Low-Phytate
459 Soybeans. *Crop Sci* **2008**, *48*, 1355.
- 460 15. Oltmans, S. E.; Fehr, W. R.; Welke, G. A.; Cianzio, S. R., Inheritance of Low-Phytate
461 Phosphorus in Soybean. *Crop Sci* **2004**, *44*, 433.
- 462 16. Oltmans, S. E.; Fehr, W. R.; Welke, G. A.; Raboy, V.; Peterson, K. L., Agronomic and
463 Seed Traits of Soybean Lines with Low-Phytate Phosphorus. *Crop Sci* **2005**, *45*, 593-598.
- 464 17. Sakurai, T.; Yamada, Y.; Sawada, Y.; Matsuda, F.; Akiyama, K.; Shinozaki, K.; Hirai,
465 M. Y.; Saito, K., PRIME Update: innovative content for plant metabolomics and integration of
466 gene expression and metabolite accumulation. *Plant Cell Physiol* **2013**, *54*, e5.
- 467 18. Frank, T.; Meuleye, B. S.; Miller, A.; Shu, Q. Y.; Engel, K. H., Metabolite profiling of
468 two low phytic acid (lpa) rice mutants. *J Agric Food Chem* **2007**, *55*, 11011-9.
- 469 19. Hazebroek, J.; Harp, T.; Shi, J.; Wang, H., Metabolomic analysis of low phytic acid
470 maize kernels. In *Concepts in plant metabolomics*, Springer: 2007; pp 221-238.
- 471 20. Wu, W.; Zhang, Q.; Zhu, Y.; Lam, H. M.; Cai, Z.; Guo, D., Comparative metabolic
472 profiling reveals secondary metabolites correlated with soybean salt tolerance. *J Agric Food*
473 *Chem* **2008**, *56*, 11132-8.
- 474 21. Sawada, Y.; Hirai, M. Y., Integrated LC-MS/MS system for plant metabolomics. *Comput*
475 *Struct Biotechnol J* **2013**, *4*, e201301011.
- 476 22. Matsuda, F.; Yonekura-Sakakibara, K.; Niida, R.; Kuromori, T.; Shinozaki, K.; Saito, K.,
477 MS/MS spectral tag-based annotation of non-targeted profile of plant secondary metabolites.
478 *Plant J* **2009**, *57*, 555-77.
- 479 23. Clarke, J. D.; Alexander, D. C.; Ward, D. P.; Ryals, J. A.; Mitchell, M. W.; Wulff, J. E.;
480 Guo, L., Assessment of genetically modified soybean in relation to natural variation in the
481 soybean seed metabolome. *Sci Rep* **2013**, *3*, 3082.
- 482 24. Lin, H.; Rao, J.; Shi, J.; Hu, C.; Cheng, F.; Wilson, Z. A.; Zhang, D.; Quan, S., Seed
483 metabolomic study reveals significant metabolite variations and correlations among different
484 soybean cultivars. *J Integr Plant Biol* **2014**, *56*, 826-36.
- 485 25. Frank, T.; Norenberg, S.; Engel, K. H., Metabolite profiling of two novel low phytic acid
486 (lpa) soybean mutants. *J Agric Food Chem* **2009**, *57*, 6408-16.
- 487 26. Haug, K.; Salek, R. M.; Conesa, P.; Hastings, J.; de Matos, P.; Rijnbeek, M.;
488 Mahendrakar, T.; Williams, M.; Neumann, S.; Rocca-Serra, P.; Maguire, E.; Gonzalez-Beltran,
489 A.; Sansone, S. A.; Griffin, J. L.; Steinbeck, C., MetaboLights--an open-access general-purpose
490 repository for metabolomics studies and associated meta-data. *Nucleic Acids Res* **2013**, *41*,
491 D781-6.

- 492 27. Sayama, T.; Ono, E.; Takagi, K.; Takada, Y.; Horikawa, M.; Nakamoto, Y.; Hirose, A.;
493 Sasama, H.; Ohashi, M.; Hasegawa, H.; Terakawa, T.; Kikuchi, A.; Kato, S.; Tatsuzaki, N.;
494 Tsukamoto, C.; Ishimoto, M., The Sg-1 glycosyltransferase locus regulates structural diversity of
495 triterpenoid saponins of soybean. *Plant Cell* **2012**, *24*, 2123-38.
- 496 28. Burleson, S. A.; Shang, C.; Rosso, M. L.; Maupin, L. M.; Rainey, K. M., A Modified
497 Colorimetric Method for Selection of Soybean Phytate Concentration. *Crop Sci* **2012**, *52*, 122.
- 498 29. Gao, Y.; Shang, C.; Maroof, M. A. S.; Biyashev, R. M.; Grabau, E. A.; Kwanyuen, P.;
499 Burton, J. W.; Buss, G. R., A Modified Colorimetric Method for Phytic Acid Analysis in
500 Soybean. *Crop Sci* **2007**, *47*, 1797.
- 501 30. Junior, A. Q.; Ida, E. I., Isoflavones of the soybean components and the effect of
502 germination time in the cotyledons and embryonic axis. *J Agric Food Chem* **2014**, *62*, 8452-9.
- 503 31. Graham, T. L., Flavonoid and isoflavonoid distribution in developing soybean seedling
504 tissues and in seed and root exudates. *Plant Physiol* **1991**, *95*, 594-603.
- 505 32. Hiroko, S.; Yoshitake, T.; Masao, I.; Keisuke, K.; Chigen, T., Estimation of the Mutation
506 Site of a Soyasapogenol A-Deficient Soybean [*Glycine max* (L.) Merr.] by LC-MS/MS Profile
507 Analysis. In *Chemistry, Texture, and Flavor of Soy*, American Chemical Society: 2010; Vol.
508 1059, pp 91-102.
- 509 33. Tsukamoto, C.; Kikuchi, A.; Harada, K.; Kitamura, K.; Okubo, K., Genetic and chemical
510 polymorphisms of saponins in soybean seed. *Phytochemistry* **1993**, *34*, 1351-1356.
- 511 34. Takada, Y.; Tayama, I.; Sayama, T.; Sasama, H.; Saruta, M.; Kikuchi, A.; Ishimoto, M.;
512 Tsukamoto, C., Genetic analysis of variations in the sugar chain composition at the C-3 position
513 of soybean seed saponins. *Breed Sci* **2012**, *61*, 639-45.
- 514 35. Takada, Y.; Sasama, H.; Sayama, T.; Kikuchi, A.; Kato, S.; Ishimoto, M.; Tsukamoto, C.,
515 Genetic and chemical analysis of a key biosynthetic step for soyasapogenol A, an aglycone of
516 group A saponins that influence soymilk flavor. *Theor Appl Genet* **2013**, *126*, 721-31.
- 517 36. Krishnamurthy, P.; Lee, J. M.; Tsukamoto, C.; Takahashi, Y.; Singh, R. J.; Lee, J. D.;
518 Chung, G., Evaluation of genetic structure of Korean wild soybean (*Glycine soja*) based on
519 saponin allele polymorphism. *Gen Res Crop Evol* **2014**, *61*, 1121-1130.
- 520 37. Gijzen, M.; Kuflu, K.; Moy, P., Gene amplification of the Hps locus in *Glycine max*.
521 *BMC Plant Biol* **2006**, *6*, 6.
- 522 38. Kuppanan, K.; Julka, S.; Karnoup, A.; Dielman, D.; Schafer, B., 2DLC-UV/MS assay
523 for the simultaneous quantification of intact soybean allergens Gly m 4 and hydrophobic protein
524 from soybean (HPS). *J Agric Food Chem* **2014**, *62*, 4884-92.
- 525 39. Odani, S.; Koide, T.; Ono, T.; Seto, Y.; Tanaka, T., Soybean hydrophobic protein.
526 Isolation, partial characterization and the complete primary structure. *Eur J Biochem* **1987**, *162*,
527 485-91.

40. Gijzen, M.; Miller, S. S.; Kuflu, K.; Buzzell, R. I.; Miki, B. L., Hydrophobic protein synthesized in the pod endocarp adheres to the seed surface. *Plant Physiol* **1999**, *120*, 951-9.
41. Baud, F.; Pebay-Peyroula, E.; Cohen-Addad, C.; Odani, S.; Lehmann, M. S., Crystal structure of hydrophobic protein from soybean; a member of a new cysteine-rich family. *J Mol Biol* **1993**, *231*, 877-87.
42. Yeats, T. H.; Rose, J. K., The biochemistry and biology of extracellular plant lipid-transfer proteins (LTPs). *Protein Sci* **2008**, *17*, 191-8.
43. Liu, F.; Zhang, X.; Lu, C.; Zeng, X.; Li, Y.; Fu, D.; Wu, G., Non-specific lipid transfer proteins in plants: presenting new advances and an integrated functional analysis. *J Expt Bot* **2015**.
44. Pagnussat, L.; Burbach, C.; Baluska, F.; de la Canal, L., An extracellular lipid transfer protein is relocalized intracellularly during seed germination. *J Exp Bot* **2012**, *63*, 6555-63.
45. Morris, C. J.; Thompson, J. F., The isolation and characterization of gamma-L-glutamyl-L-tyrosine and gamma-L-glutamyl-L-phenylalanine from soybeans. *Biochemistry* **1962**, *1*, 706-9.
46. Yamada, T.; Mori, Y.; Yasue, K.; Maruyama, N.; Kitamura, K.; Abe, J., Knockdown of the 7S globulin subunits shifts distribution of nitrogen sources to the residual protein fraction in transgenic soybean seeds. *Plant cell reports* **2014**, *33*, 1963-76.

Figure Captions

Figure 1. Venn diagrams displaying the EMRTs with p-values < 0.05 and Factor of Changes less than 0.5 or greater than 2 in all comparisons between the low phytate and normal phytate lines. The numbers in parentheses at the top of each diagram is the total number of EMRTs identified as significantly different. The number in parentheses below each pairwise comparison is the total number of EMRTs identified as significantly different for that pairwise comparison. The numbers within the diagram are the EMRTs unique to that particular grouping.

Figure 2. Principal component analyses of the four lines investigated. Each mode of separation and method of detection separated the low phytate line from the normal phytate lines.

Figure 3. Base peak ion chromatograms of each line (reversed-phase separation, negative ion mode detection). The low phytate near isogenic line was enriched in xylose-terminated soyasaponins.

Figure 4. Orthogonal nature of the reversed-phase and HILIC separation modes. The soybean seed dipeptides elute in accord with the separation medium. Ion intensities are condition-dependent as well.

Figure 5. Relative base peak ion chromatograms for the oligosaccharide region of the WT and DM/Low phytate lines (HILIC-based separations).

Table 1. Genetic Characteristics and Phytic Acid Concentrations of the Near Isogenic Lines (NILs) Investigated.

	Near Isogenic Lines (NILs)			
Mutations and Phytate Level	<i>mrp(3)/mrp(19)</i> (DM)	<i>mrp(3)/MRP(19)</i> (SM3)	<i>MRP(3)/mrp(19)</i> (SM19)	<i>MRP(3)/MRP(19)</i> (WT)
MRP on Chr 3	yes	yes	no	no
MRP on Chr 19	yes	no	yes	no
Phytate (mg/gm)	3.6 +/-1.2	14.2 +/- 0.8	16.5 +/- 0.5	14.5 +/- 0.6

Table 2. Comparison of EMRTs Across Separation Strategies and Ionization Modes (RP, Reversed-phase; HILIC, hydrophilic interaction).

Feature Category	RP-(+)	RP-(-)	HILIC-(+)	HILIC-(-)
Total EMRTs identified	154	355	1066	1221
Significantly changing EMRTs ¹	63	273	322	467
EMRTs changing across all lines ²	24	141	146	199
Total EMRTs from pairwise comparisons ³	29	87	104	128

¹p-value <0.05, Factor of Change < 0.5 or > 2.0.

²Center of Venn diagrams in Fig. 1.

³Sum of EMRTs in pairwise comparisons shown in Fig. 1.

Table 3. Selected Exact Mass/Retention Time Pairs With Significant Differences Between the Low and Normal Phytate Lines by Reversed-Phase LC-MS Analysis.¹

Compound	RT (min)	Mass obsd.	Ion	Peak Areas				
				DM ²	SM3	SM19	WT	LP:NP ³
Malonyldaidzin	3.93	1003.2140	[2M-H] ⁻	1785	6699	6266	5591	0.29
Malonylgenistin	4.35	1035.2040	[2M-H] ⁻	2891	11799	11456	9725	0.26
Gly m 1	4.42	1392.5460	[M+6H] ⁺⁶	trace	632	232	517	NP
3.9 kDa peptide	5.20	973.6720	[M+4H] ⁺⁴	trace	410	1076	331	NP
Soyasaponin A4/Aa	5.75	1363.6160	[M-H] ⁻	6783	11	17	13	492
	5.75	1365.6300	[M+H] ⁺	368	trace	trace	trace	LP
Soyasaponin Au	5.82	1347.6200	[M-H] ⁻	537	trace	trace	trace	LP
Soyasaponin A1/Ab	5.88	1435.6380	[M-H] ⁻	90	11040	14845	11689	0.01
	5.88	1437.6510	[M+H] ⁺	trace	597	852	637	NP
Soyasaponin Ac	5.95	1419.6420	[M-H] ⁻	trace	1388	1698	1365	NP
Soyasaponin Ad	5.98	1405.6270	[M-H] ⁻	trace	656	804	581	NP
Soyasaponin A5/Ae	6.00	1201.5630	[M-H] ⁻	2923	trace	trace	trace	LP
Soyasaponin A6/Ag	6.06	1171.5520	[M-H] ⁻	423	trace	trace	trace	LP
Soyasaponin A2/Af	6.12	1273.5840	[M-H] ⁻	trace	4752	6580	4999	NP
Soyasaponin Ba	6.47	957.5046	[M-H] ⁻	934	3346	4441	2861	0.26
	6.46	959.5198	[M+H] ⁺	118	435	616	360	0.25
Soyasaponin Bb	6.59	941.5105	[M-H] ⁻	9823	25765	30957	21604	0.38
	6.58	943.5246	[M+H] ⁺	1315	4470	4823	3578	0.31
Soyasaponin Bc	6.74	911.5000	[M-H] ⁻	2772	15991	14899	11995	0.19
	6.74	913.5144	[M+H] ⁺	49	297	282	216	0.18
Soyasaponin βg	7.09	1067.5420	[M-H] ⁻	5542	17891	30749	14696	0.26
	7.08	1069.5560	[M+H] ⁺	470	3027	5489	1999	0.13

¹Full listing of EMRTs can be found in the Supplemental (Figs. S4-S7).²Low phytate line (double mutant).³DM peak area/Average of other three lines, peak area ratio.

Table 4. Selected HILIC-Based Ions of Significance.

Compound	RT (min)	Mass obsd.	Ion	Peak Areas				
				DM ²	SM3	SM19	WT	LP/NP ³
Soyasaponin A2/Af	2.93	1273.5850	[M-H] ⁻	62	5339	6476	6024	0.010
	2.92	1275.5980	[M+H] ⁺	trace	734	768	808	NP
Soyasaponin A5/Ae	2.94	1201.5630	[M-H] ⁻	5630	338	433	160	18
	2.94	1203.5770	[M+H] ⁺	1030	trace	trace	trace	LP
Soyasaponin Ac	3.24	1419.6410	[M-H] ⁻	trace	551	701	539	NP
	3.23	1421.6560	[M+H] ⁺	trace	132	156	147	NP
Soyasaponin Au	3.25	1347.6200	[M-H] ⁻	435	trace	trace	trace	LP
	3.25	1349.6350	[M+H] ⁺	128	trace	trace	trace	LP
Soyasaponin A1/Ab	3.29	1435.6380	[M-H] ⁻	trace	4543	4741	4393	NP
	3.29	1437.6510	[M+H] ⁺	trace	830	965	888	NP
Soyasaponin A4/Aa	3.31	1363.6160	[M-H] ⁻	2518	trace	trace	trace	LP
	3.31	1365.6300	[M+H] ⁺	798	trace	trace	trace	LP
Arginine	4.73	173.1034	[M-H] ⁻	5553	2134	2279	2165	2.5
	4.75	175.1195	[M+H] ⁺	11547	4411	4401	4727	2.6

¹Full listing of EMRTs can be found in the Supplemental (Figs. S4-S7).

²Low phytate line (double mutant).

³DM peak area/Average of other three lines, peak area ratio.

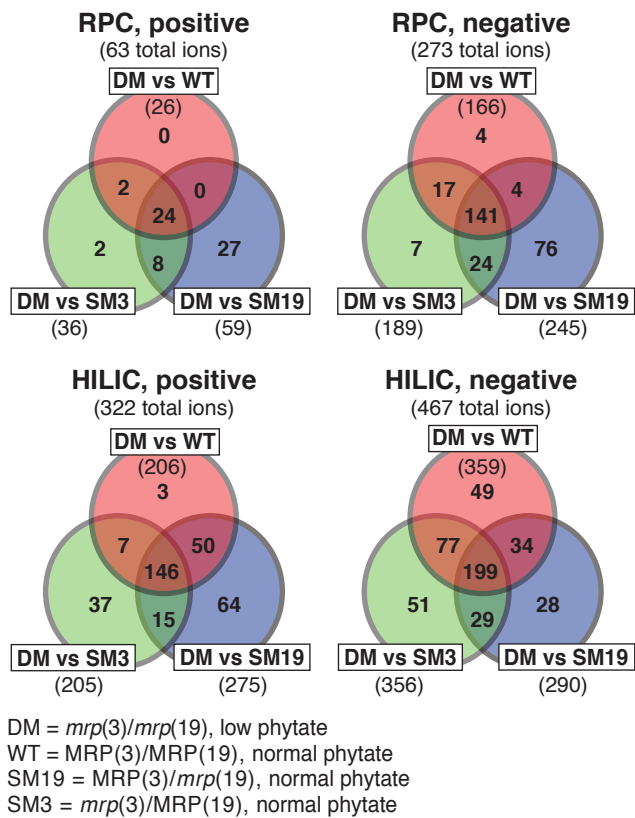


Figure 1

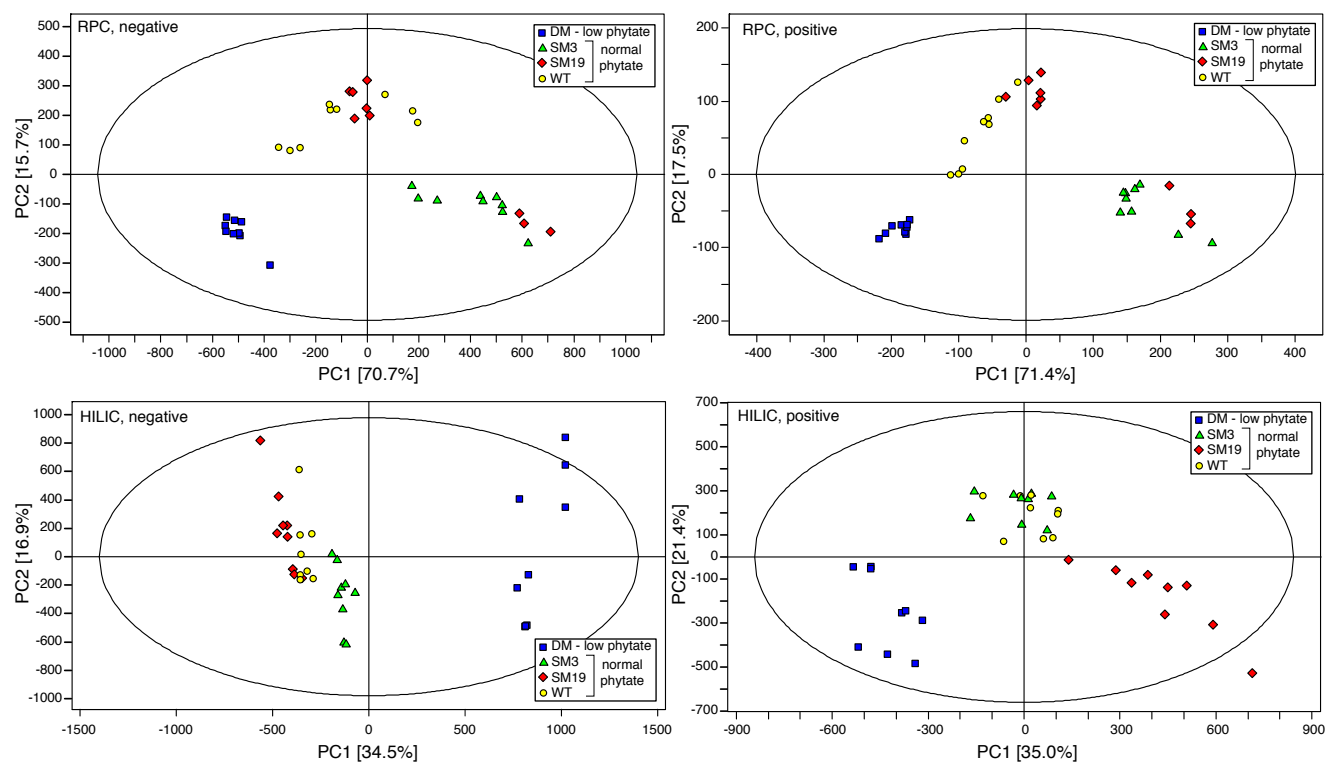


Figure 2

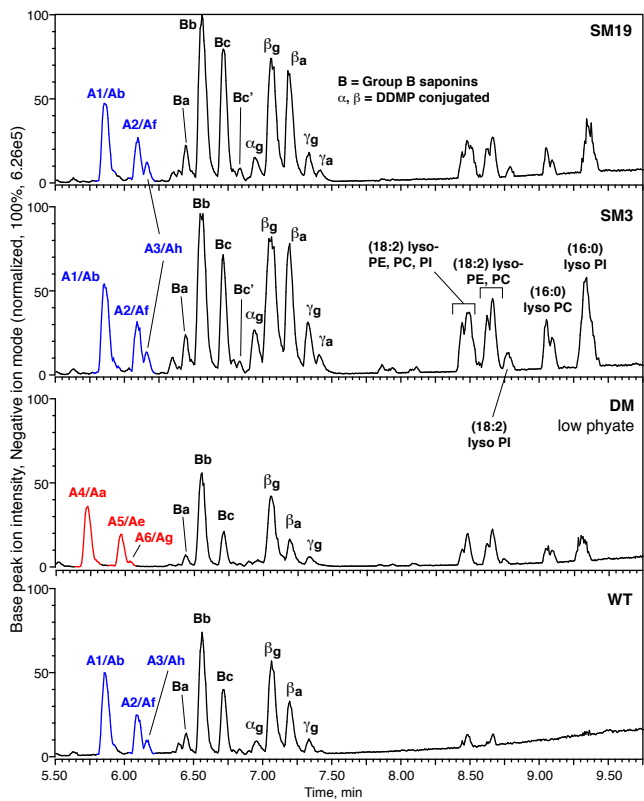
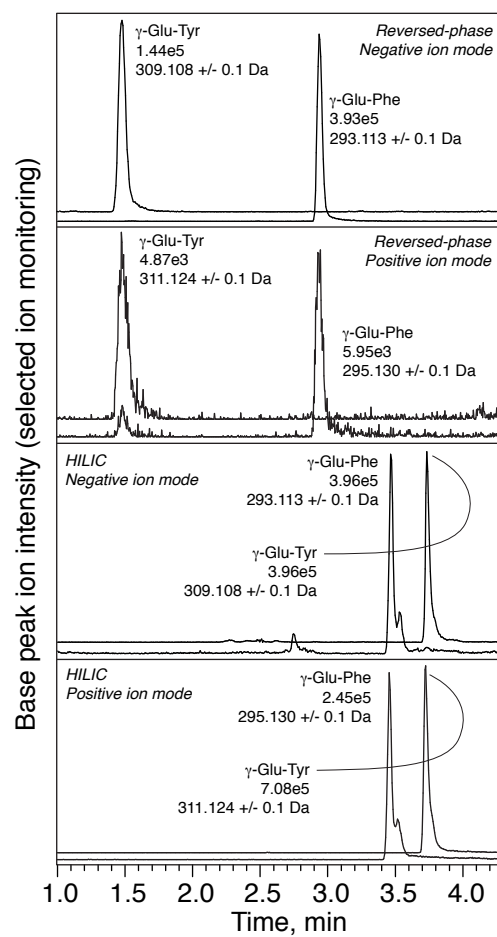


Figure 3

**Figure 4**

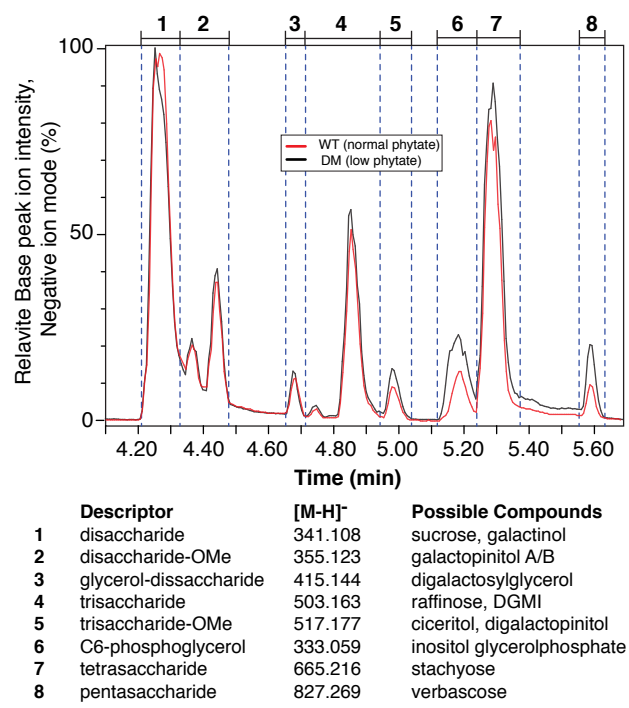


Figure 5

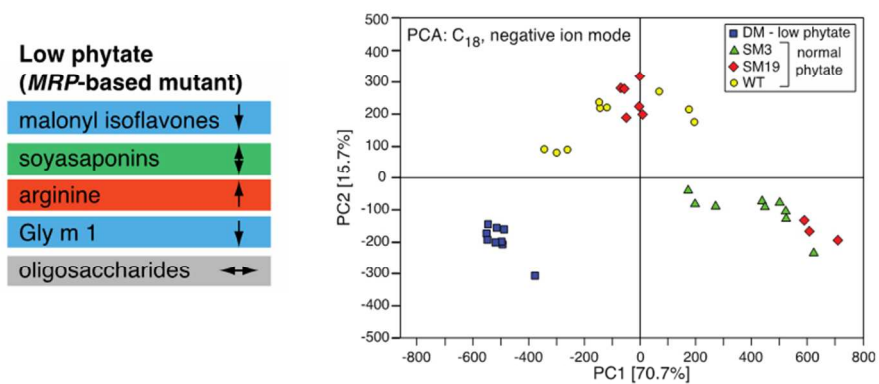


Table of Contents Graphic
245x118mm (96 x 96 DPI)