Published December 12, 2012

Ionomic Screening of Field-Grown Soybean Identifies Mutants with Altered Seed Elemental Composition

Greg Ziegler, Aimee Terauchi, Anthony Becker, Paul Armstrong, Karen Hudson, and Ivan Baxter*

Abstract

Soybean [*Glycine max* (L.) Merr.] seeds contain high levels of mineral nutrients essential for human and animal nutrition. High throughput elemental profiling (ionomics) has identified mutants in model plant species grown in controlled environments. Here, we describe a method for identifying potential soybean ionomics mutants grown in a field setting and apply it to 975 N-nitroso-N-methylurea (NMU) mutagenized lines. After performing a spatial correction, we identified mutants using either visual scoring of standard score (z-score) plots or computational ranking of putative mutants followed by visual confirmation. Although there was a large degree of overlap between the methods, each method identified unique lines. The visual scoring approach identified 22 out of 427 (5%) potential mutants, 70% (16 out of 22) of which were confirmed when seeds from the same parent plant were regrown in the field. We also performed simulations to determine an optimal strategy for screening large populations. All data from the screen is available at the Ionomics Hub File Transfer Page (http://www.ionomicshub.org/home/PiiMS/dataexchange).

Published in The Plant Genome 6.
doi: 10.3835/plantgenome2012.07.0012

5585 Guilford Rd., Madison, WI 53711 USA
An open-access publication

THE ELEMENTAL COMPOSITION of soybean is an important component of their overall nutritional value. Soybean is rich in Fe and other mineral nutrients required for human and animal nutrition (Messina, 1999). As most of the elemental composition of the seeds, with the exception of C and O_2, is obtained from the soil, seed composition is an indicator of the plant's interaction with the environment (Baxter and Dilkes, 2012). Accordingly, identifying the genetic factors controlling the elemental composition (the ionome) of soybean seeds has the potential to improve yield, quality, and our understanding of soybean–soil interactions.

Ionomics (high throughput elemental profiling) has been used to identify mutants and natural alleles in plants grown in potting mix or hydroponics in controlled environments (Lahner et al., 2003; Chen et al., 2009). These approaches identified mutants where single elements were altered as well as mutants with multi-element phenotypes. Cloning of the underlying loci driving two of the multi-element phenotypes identified by Lahner et al. (2003) led to the discovery of genes that had not been predicted to affect elemental composition. *ESB1*, a gene of unknown molecular function, and *TSC10a*, a gene in the sphingolipid pathway, have both been shown to affect suberin deposition in the root

G. Ziegler and A. Becker, Donald Danforth Plant Science Center, St. Louis, MO; A. Terauchi and I. Baxter, USDA-ARS, Plant Genetics Research Unit, St. Louis, MO; P. Armstrong, USDA-ARS, GGAHR, Manhattan, KS; K. Hudson, USDA-ARS, Agronomy, West Lafayette, IN. Greg Ziegler and Aimee Terauchi contributed equally to this work. Received 1 July 2012. *Corresponding author (Ivan.Baxter@ars.usda.gov).

Abbreviations: DRC, dynamic reaction cell; ICP-MS, ion coupled plasma mass spectrometry; IS, internal standard; NMU, N-nitroso-N-methylurea; q, studentized range statistic; UPW, ultrapure 18.2 MΩ water; z-score, standard score; Zmax, the selection of mutants based on the highest z-score of all of the elements measured; Zsum, the selection of mutants based on the sum of the z-scores for all of the elements measured.

endodermis and have a strong effect on the leaf ionome (Baxter et al., 2009; Chao et al., 2011). Both defects lead to increases in suberin levels, thereby affecting the movement of water and ions in the root, resulting in an altered multi-element profile.

Extending the ionomics approach beyond model systems to field-grown crop plants presents several challenges, principally the lack of control over environmental variation. Variation across the field in soil composition, hydrology, and topology can have large effects on the elements that are available for uptake. These differences can mask genetic variation, increasing both the false positive rate and false negative rates for mutant identification. Nevertheless, mutants with altered elemental profiles in a field setting have the potential to shed light on soil–plant interactions that are not visible in controlled growth environments and are necessary for genetic efforts to improve the nutritional content of crop seeds. Here, we demonstrate that the ionomics approach can be used to identify genetic mutants in soybean by using a high-throughput workflow, analyzing multiple soybeans per line, and performing a spatial correction to help control for field effects before identifying mutants. Furthermore, we explore several methods of identifying potential mutants and evaluate experimental designs to reduce the total number of samples needed to conduct such screens.

Materials and Methods

Population and Field Growth

The 'Williams-82' soybean population was mutagenized with N-nitroso-N-methylurea (NMU) as previously described in Cooper et al. (2008) (population D). This population is expected to have a mutation density of approximately one base pair change per 140 kb. Approximately 25 plants per line from the M_3 generation were grown in 1.8-m rows with 0.76-m row spacing and 1.2-m alleys between rows in the 2009 growing season in West Lafayette, IN. Rows were harvested in bulk. For the second round of screening, additional M_3 individuals were grown in the same field during the 2011 growing season in West Lafayette. In 2009, seed from nonmutagenized Williams-82 was collected from plants interspersed with the mutants. In 2011, Williams-82 was grown nearby but in nonadjacent plots of the same field as the regrown mutants. For a select number of M_3 individuals, pods were harvested individually so that pod-to-pod variation could be assessed.

Determination of Elemental Concentration by Ion Coupled Plasma Mass Spectrometry Analysis

Sample Preparation and Digestion

Nine hundred forty-seven lines from the M_3 generation of the mutagenized Williams-82 population were analyzed for the elemental concentrations of 20 elements. For each line, six or eight seeds were randomly pulled from the bulk seed packets for analysis. Seeds were sorted into 48-well tissue culture plates with one seed per well. A weight for each individual seed was determined using a custom built weighing robot. The weighing robot holds six 48-well plates and maneuvers each well of the plates over a hole that opens onto a three-place balance. After recording the weight, each seed was deposited using pressurized air into a 16 by 110 mm borosilicate glass test tube for digestion. The weighing robot can automatically weigh 288 seeds in approximately 1.5 h with little user intervention.

Seeds were digested in 2.5 mL concentrated HNO_3 (AR Select Grade, VWR International, LLC) with internal standard added (20 μg L^{-1} In) (Aristar Plus, BDH Chemicals). Seeds were soaked at room temperature overnight and then heated to 105°C for 2 h. After cooling, the samples were diluted to 10 mL using ultrapure 18.2 MΩ water (UPW) from a Milli-Q system (Millipore). Samples were stirred with a custom-built stirring rod assembly, which uses plastic stirring rods to stir 60 test tubes at a time. Between uses, the stirring rod assembly was soaked in a 10% HNO_3 solution. A second dilution of 0.9 mL of the first dilution and 4.1 mL UPW was prepared in a second set of test tubes. After stirring, 1.2 mL of the second dilution was loaded into 96-well autosampler trays.

Ion Coupled Plasma Mass Spectrometry Analysis

Elemental concentrations of B, Na, Mg, Al, P, S, K, Ca, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Rb, Mo, and Cd were measured using an Elan 6000 DRC (dynamic reaction cell)-e mass spectrometer (PerkinElmer SCIEX) connected to a Perfluoroalkoxy (PFA) microflow nebulizer (Elemental Scientific) and Apex HF desolvator (Elemental Scientific). Samples were introduced using a SC-FAST sample introduction system and SC4-DX autosampler (Elemental Scientific) that holds six 96-well trays (576 samples).

All elements were measured with DRC collision mode off. Before each run, the lens voltage and nebulizer gas flow rate of the ion coupled plasma mass spectrometry (ICP-MS) were optimized for maximum In signal intensity (>25,000 counts per second) while also maintaining low CeO+:Ce+ (<0.008) and Ba++:Ba+ (<0.1) ratios. This ensures a strong signal while also reducing the interferences caused by polyatomic and double-charged species. Before each run a calibration curve was obtained by analyzing six dilutions of a multi-element stock solution made from a mixture of single-element stock standards (Ultra Scientific). In addition, to correct for machine drift both during a single run and between runs, a control solution was run every 10th sample. The control solution is a bulk mixture of the remaining sample from the second dilution. Using bulked samples ensured that our controls were perfectly matrix matched and contained the same elemental concentrations as our samples, so that any drift due to the sample matrix would be reflected in drift in our controls. The same control mixture was used for every ICP-MS run in the project so that run-to-run variation could be corrected. A run of 576 samples took approximately 33 h with no user intervention. The time

required for cleaning of the instrument and sample tubes as well as the digestions and transfers necessary to set up the run limit the throughput to two 576 sample runs per week.

Drift Correction and Analytical Outlier Removal

Because our internal standard (IS) is added to our digesting acid, we are able to correct for losses due to differential sample evaporation, human error during the dilution process, and any sample introduction variability. So, if the final observed IS concentration is lower or higher than the starting IS concentration, all analyte concentrations are corrected equally for the percent difference between the observed IS concentration and the known starting IS concentration. Internal standard correction is handled automatically by the PerkinElmer Elan 6000 software. Additionally, drift was corrected using the values of the controls run every 10th sample using a method similar to Haugen (Haugen et al., 2000). In short, drift was corrected by calculating the percentage of concentration change between two controls. This percentage change was assumed to have occurred linearly during the sequence of 10 samples run between the two controls. So, for instance, the first sample run after the first control was corrected for 1/10th of the drift seen between the two controls. Finally, because responses from the machine may be different between runs, we also corrected for drift between runs. This was performed by calculating a correction factor from the control concentrations in this run and a reference control used for all the runs. After drift correction, samples were corrected for the dilution factor and normalized to the seed weights.

While biological outliers are of great interest to our analysis, analytical outliers (e.g., from contamination, spurious isobaric and polyatomic interferences, or poor sample uptake) introduce noise and could lead to a higher rate of incorrectly chosen potential mutants. Outlier removal was implemented using the algorithm described in Davies and Gather (1993). To remove outliers while ensuring that we were not removing biologically significant data points we removed data points on a per element basis from seeds whose reported elemental concentration was greater than a conservatively set 6.2 median absolute deviations from the median concentration of that element for that line.

Sample Size and Elements Omitted

We analyzed the first 427 lines at a depth of eight seeds per line. After visual inspection of the data for the line replicates, we determined that a depth of six seeds per line would provide enough information to select potential mutants, even for those lines that were potentially heterozygous.

Analysis of the concentrations of the elements measured across the ICP-MS runs for the experiment indicated that some of the elements measured were often below our limits of quantitation (As and Cd), prone to contamination during our sample preparation (Na and Al), or prone to interference from other components in the sample matrix (As and Al). To reduce the detection of false positives and incorrect conclusions based on noise from poorly quantified elements, we did not consider these elements in our screen for mutants.

Computational Analysis

All calculations were performed using a combination of in-house Perl and R scripts. The spatial correction plots used for visual inspection of field effects and evaluating the efficacy of our spatial correction technique were created using the lattice R package and the melt function of the reshape R package (Wickham, 2007). The Venn diagram was generated using the VennDiagram R package (Chen and Boutros, 2011). Standard Perl and R functions were used for drift correction, calculation of standard scores (z-scores), and creation of the z-score plots. Scripts used in this analysis are available at http://www.ionomicshub.org/home/PiiMS/data-exchange in the data exchange portal.

Results

Spatial Correction for Field Effects

We screened 947 field grown soybean lines using the ionomic pipeline detailed in the Materials and Methods. Differences in elemental profiles due to soil variation across the field site affected the distribution of ionomic phenotypes (Fig. 1A). We adjusted the elemental profiles for each genotype using to the elemental profiles from the adjacent plots. The correction for each genotype was performed by taking the median concentration of each element in a three by three block of plants centered around the genotype and subtracting these medians from the concentration of each element in the line. This has the effect of removing variance in the elemental profiles of plants when neighbors co-vary due to spatial effects while preserving the signal that is due to genetic variation (Fig. 1B). The spatially corrected values resulted in mutant identification throughout the field rather than the selection of one row of plants with a high elemental concentration as is evident without spatial correction (Fig. 1 and 2).

Putative Mutant Identification

To make a preliminary summarization of potential significant differences, each element measurement for every line was compared to all other samples of that element using a simple *t* test with an experiment-wide (14,205 tests) false discovery rate correction (Tables 1 and 2). Even with the significance cutoff level set quite stringently, studentized range statistic (q) < 0.001, 155 lines had at least one element significantly different from the population. The number of significant differences is likely due to both genetic factors (mutations) and environmental variation unaccounted for by the spatial correction combined with the precision of the ionomics techniques and the high statistical power of the $n = 6$ to 8 design. Significant differences were detected for all elements in at least three lines. The fewest mutants (three

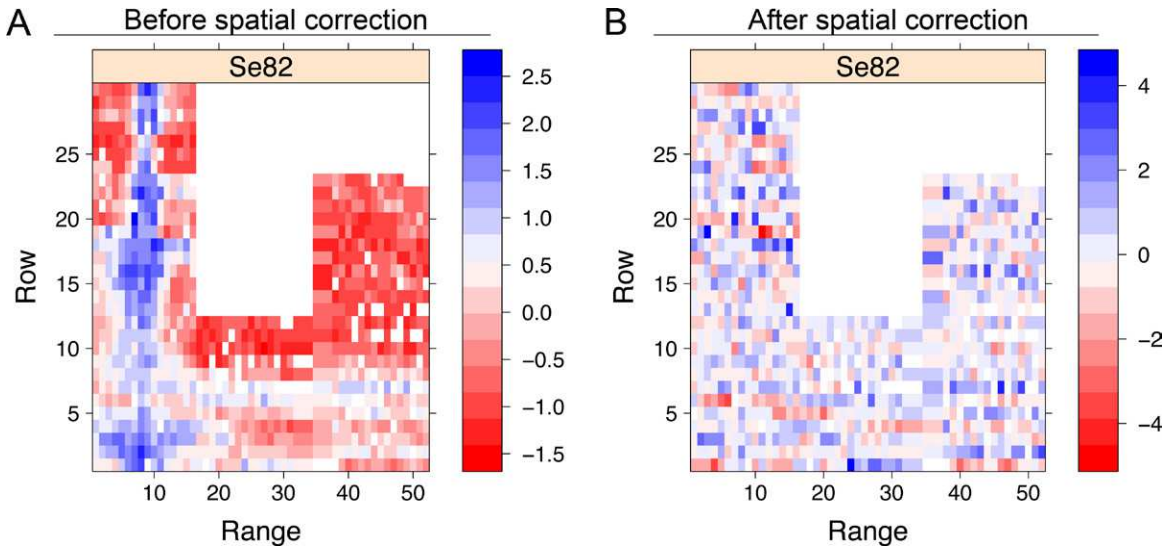

Figure 1. Spatial correction of field data. (A) Field effects before spatial correction. Selenium is shown as a representative example. High Se concentrations are represented in blue and low Se concentrations are shown in red. Scale represents number of standard deviations from the median. (B) Field effects were corrected for by comparing element content of lines to spatially adjacent lines. Selenium is shown as a representative example. Spatial correction for all elements is shown in Supplemental Fig. S2. Se82, ^{82}Se.

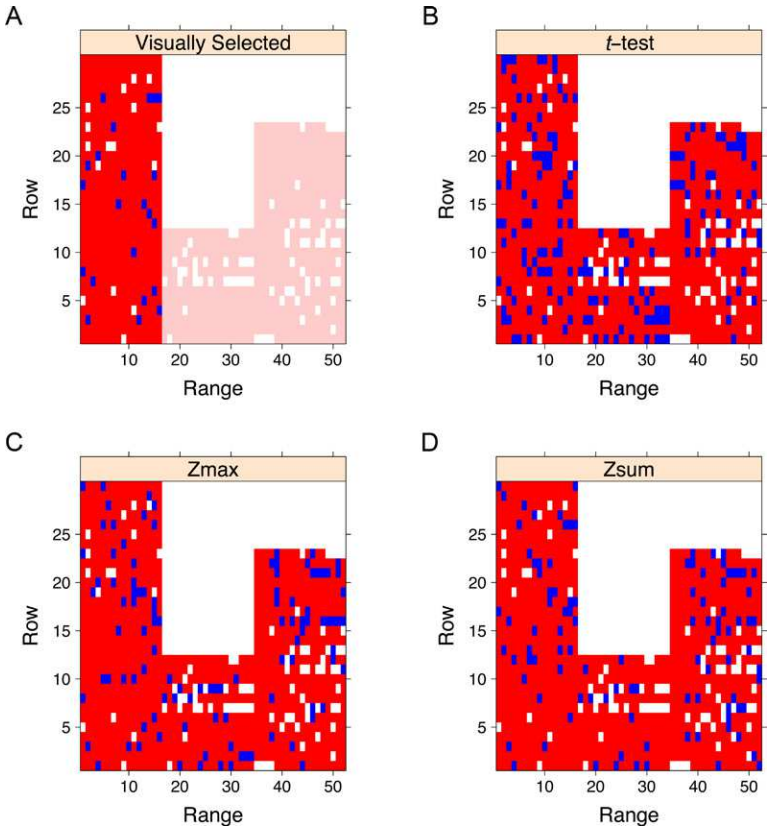

Figure 2. Field map of putative mutants. (A) Field location of putative mutants that were selected visually. The 22 putative mutant lines, shown in blue, were selected from the first 427 lines, represented in red. The rest of the lines were not considered in the visual screen (see text). (B) Field location of lines identified by *t* test (studentized range statistic [q] < 0.001). One hundred fifty-six lines identified highlighted in blue. (C) Field location of top 100 lines (blue) identified using the selection of mutants based on the highest z-score of all of the elements measured (Zmax) method. (D) Field location of top 100 lines (blue) identified using the selection of mutants based on the sum of the z-scores for all of the elements measured (Zsum) method.

each) were detected for K and Fe accumulation while 31 mutants were identified with a difference in Rb accumulation (Table 1). One hundred thirty-two of the lines (85%) had only a single significantly different element, with the rest having two to nine elements with altered accumulation.

Table 1. Statistically significant putative mutants as determined by *t* test (studentized range statistic [q] < 0.01).

Element	Mg25†	P31	S34	K39	Ca43	Mn55	Fe57	Co59	Ni60	Cu65	Zn66	Se82	Rb85	Sr88	Mo98
No. of putative mutant lines	5	17	12	3	10	8	3	25	28	11	14	10	31	13	10

†The numbers following the elements are the isotope measured by the ion coupled plasma mass spectrometry for that element. Elemental concentrations are then calculated based on the natural abundance of that isotope.

Table 2. Distribution of number of putative mutations per line determined by *t* test (studentized range statistic [q] < 0.01).

No. of elements	0	1	2	3	5	9
No. of putative mutant lines	791	132	14	7	2	1

We took three approaches to identifying mutants for confirmation: visual scoring of z-score plots, either ranking mutants by highest single absolute value of mean z-score (the selection of mutants based on the highest z-score of all of the elements measured [Zmax]) of any element, or ranking based on the sum of the absolute values of the mean elements across all 15 elements (the selection of mutants based on the sum of the z-scores for all of the elements measured [Zsum]) followed by visual confirmation (by A. Terauchi, G. Ziegler, and I. Baxter). Standard scores (z-scores) were selected as the statistical tests so that the computational method could be compared to the visual method. Visual scoring was only performed on the first 427 mutant lines while the ranked z-score tests were performed on all lines. The visual scoring approach identified 22 out of 427 (5%) potential mutants. Of the top 100 mutants identified by the Zmax method, 95 (10%, 37 in the first 427 lines) were confirmed as putative mutants by the visual scoring of z-score plots. Seventy-three of the top 100 lines selected using the Zsum were confirmed as putative mutants by visual inspection, 40 from the top 50, and 33 from 50 through 100. There is a large degree of overlap between the three methods as well as the *t* test method (Fig. 3), but each method identified unique lines. Plots of the z-score profile for each of the plants identified as putative mutants by the Zmax and Zsum method are contained in Supplemental File S1.

Mutant Confirmation

The 22 putative mutants identified from visual inspection of the first 427 lines were selected for further analysis. Individual seeds from the original seed stock were planted in 2011 at the Purdue Ag farm and the elemental profile was determined. Analysis of single seeds permits the detection of parent plants segregating for a phenotype within the row. For example, line 19147 (Fig. 4) appears to be homozygous for high S but segregating for high Co and Se. To determine if the phenotypes were indeed segregating by paternal plant, seeds from single plants were harvested and analyzed at $n = 4$ per plant. Standard scores (z-scores) for the accumulation of each element were calculated using wild-type Williams-82 grown in the same field. Of the 22 putative mutant lines replanted and analyzed, nine (41%) showed a similar z-score profile

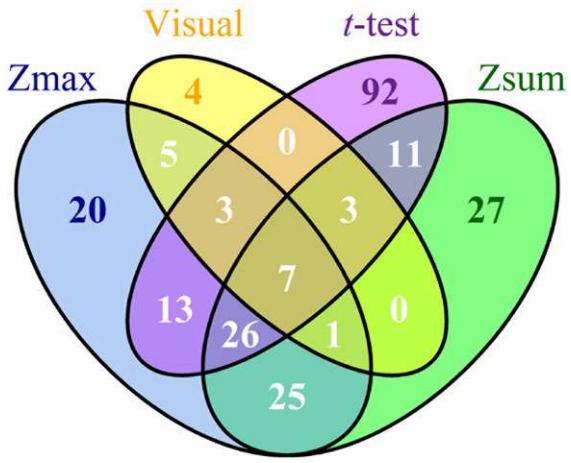

Figure 3. Number of putative mutants identified using each method (visual, *t* test, the selection of mutants based on the highest z-score of all of the elements measured [Zmax], and the selection of mutants based on the sum of the z-scores for all of the elements measured [Zsum]). A four-set Venn diagram shows the intersections of the top 100 lines identified using the Zmax method (blue), top 100 lines identified using the Zsum method (green), lines identified by *t* test (studentized range statistic [q] < 0.001) (purple), and lines chosen in the visual screen of 427 lines (yellow).

to the original data and were significantly affected for all elements identified in the primary screen (Table 3; Fig. 4). Seven additional lines displayed differences for at least one, but not all, element identified as significant in the original screen. Overall, 16 lines (72%) appeared to be mutant in the second round of analysis, higher than observed in controlled environment screens of *Arabidopsis thaliana* (L.) Heynh. where Lahner et al. had 14% reproducibility (Lahner et al., 2003) and in *Lotus corniculatus* L. var. *japonicus* Regel [syn. *Lotus japonicus* (Regel) K. Larsen] where Chen et al. observed 26% with an additional generation where 17% replicated again (Chen and Boutros, 2011). Of the 22 putative mutants identified, eight displayed patterns suggesting that a mutation was segregating within the row although few showed segregation for all of the significantly different elements. Of these, only one was confirmed in the single plant regrow analysis (Fig. 4A). An additional line was confirmed as having a segregating phenotype for one of two elements. The other six were unable to be confirmed, some perhaps due to low numbers of plants analyzed but also suggesting that some of the segregating elements were false positives. Collectively, these results demonstrate the potential for using an ionomics approach to identify genetic mutants in soybean and the accuracy of visual assessment of z-score profiles for selection of mutants in element accumulation. The 16 confirmed mutant lines displayed a wide variety of elemental

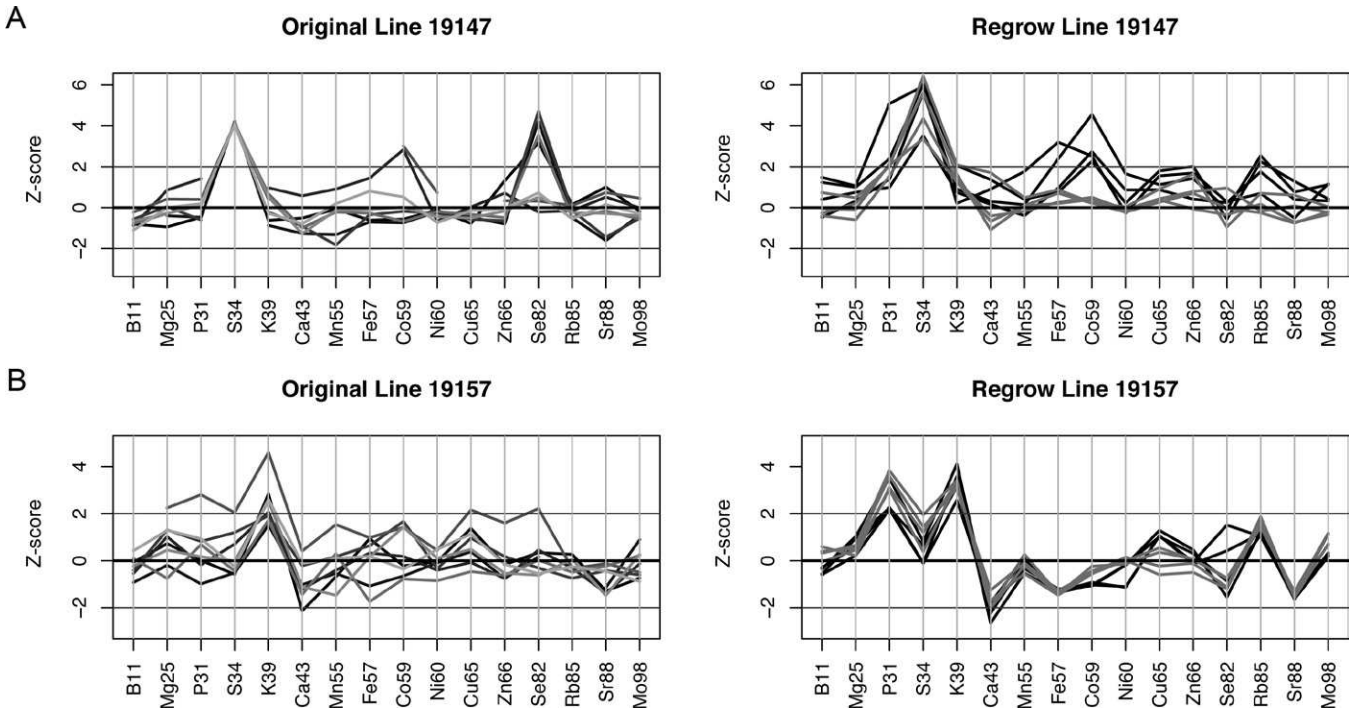

Figure 4. Representative standard score (z-score) plots. (A) Representative z-score plots showing visually selected putative mutants in the initial screen and corresponding mutant confirmation from regrowth. Colors in the regrow plots correspond to different plants of the same line to look for segregating phenotypes. Example of a line with a mutant phenotype for S and a segregating phenotype for Co is shown. Note that the putative segregating phenotype for Se was not confirmed in the regrow line. (B) Example of a line chosen for a mutant phenotype for K and Ca confirmed in the regrow line. Standard score (z-score) plots comparing all 22 visually selected to the regrown lines can be found in Supplemental File S3.

Table 3. Twenty-two putative mutants were selected visually based on standard score (z-score) profiles and regrown in the field to confirm the phenotype. Phenotypes that were confirmed in the regrow analysis are shown in italics. Lines that had some seeds exhibiting a mutant phenotype and some seeds exhibiting a wild-type phenotype were noted as segregating (seg).

Line	Original phenotype	Regrow phenotype	No. of plants analyzed in regrow
19147	*High S, Co (seg),* and Se (seg)	*High S* and *Co (seg)*	8
19157	*High K* and *low Ca*	*High K* and Rb and *low Ca*	6
19177	*High Co* and low Mn (seg)	*High Co*	4
19190	*Low Cu* and high Fe (seg)	*Low Cu*	6
19208	High Mg, S, and Sr and low Rb	High Co and low Ca	2
19226	*High Mg*	*High Mg*	7
19268	High Zn (seg)	Not significant	5
19320	High Ca, *Ni,* Se, and Sr	High P (seg) and *Ni* and low Cu	7
19350	*High Ni*	*High Ni* and Zn	2
19370	*Low P*	*Low P (seg)*	6
19394	High Mo (seg) and low Co	Not significant	9
19400	High B, *P, S,* Ni, and Se and *low K (seg)*	*High P* and *S* and *low K (seg)*	6
19421	High Mg, Ca, and Sr	*High Mg, Ca,* and *Sr*	9
19472	Low Ca	*Low Ca*	4
19518	High Co and Ni and low S and Mn	*High Co* and *Ni* and low Fe	6
19527	High Ni and P	*High P* and K and low Ca and Sr	9
19528	High Rb and Sr	Not significant	9
19529	*High P, K, Ni,* and *Zn* and *low Ca*	*High P, K, Ni,* and *Zn* and *low Ca*	2
20318	*Low Ca*	*Low Ca*	3
20322	High S and low K (seg) and B	Not significant	1
20324	*High P* and S and low Se	*High P*	1
20325	High Mg (seg) and Ca (seg)	Not significant	1

profiles (Table 3) with differences in both macro- and micronutrients and both single and multi-element profiles. Visual inspection of the lines that replicated at the V6 to V7 stage revealed several lines with morphological abnormalities and chlorosis but also lines that looked phenotypically wild-type (Supplemental Fig. S1). For several plants, the seeds from pods were kept together, allowing us to determine the variability within and between pods for a single plant (Supplemental File S2).

Screen Design Simulations

Profiling mutants at $n = 6$ or 8 is a labor- and resource-intensive sampling scheme with 1000 lines requiring approximately 5.5 wk on the ionomics pipeline as currently configured. One way to reduce the number of samples required would be to profile a smaller number seeds per line and select a subset to run at higher depth based on the limited sampling, for example, using an $n = 1$ to 3 screen to identify the top 200 candidates, which would then be further screened to get to $n = 6$ for visual scoring of z-score plots. To test this proposed strategy, we simulated 100 datasets for $n = 1$ to 7 by randomly selecting samples from each line. The 113 putative mutants selected using the Zmax and Zsum methods were assumed to be our "true" positives. For each simulated dataset, we identified the putative top 200 mutants by taking the top 150 ranked lines identified by the Zmax method and the top 50 lines from the Zsum method that were not included in the Zmax list. Since the Zmax method produced more visually confirmed (by z-score) putative mutants, we believe that Zmax method should be prioritized when combining rankings. All of the simulation sets with $n \geq 4$ included an average of at least 94% of the 113 "true" positives found in the full dataset (Fig. 5). The percentage dropped significantly from $n = 3$ to $n = 1$, but the top 200 selected from the $n = 2$ screen still contained 75% of the 113.

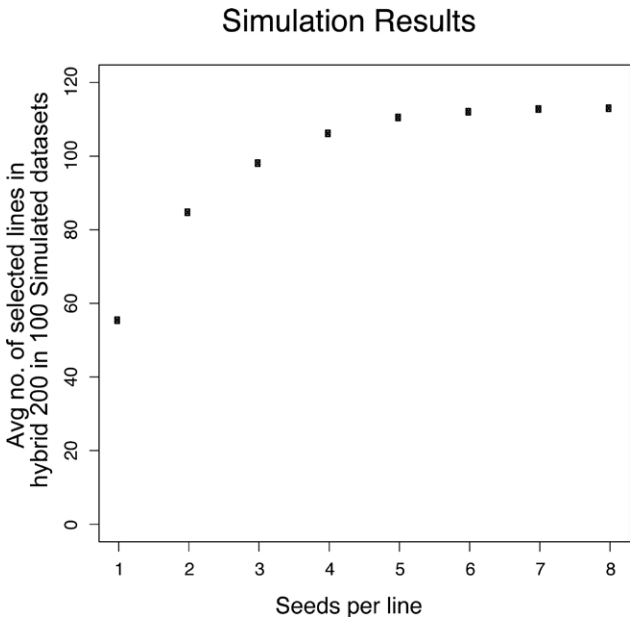

Figure 5. Results of experimental design simulation. The average number of the 113 "true" positives present in the composite top 200 selection of mutants based on the highest z-score of all of the elements measured (Zmax) and the selection of mutants based on the sum of the z-scores for all of the elements measured (Zsum) ranked lines in 100 simulated datasets with $n = 1$ to 7 seeds per line. The value for $n = 8$ is taken from the full dataset.

Discussion

We have used ionomics to identify mutants with altered seed elemental composition from a population of 947 NMU mutagenized soybean families. Throughput and accuracy was increased by analyzing single whole seeds, as homogenizing samples by grinding seeds is labor intensive and has the potential to introduce elemental contaminants. Single seed analysis enabled us to develop a straightforward automated processing workflow for soybean ionomic profiling. This approach should be generally applicable to all large seeded species and therefore many crops. Due to the high throughput and lower cost associated with automation, analyzing multiple seeds allows for the estimation of the seed to seed variances, the detection of genetic segregation, and greater power to detect differences between genotypes than would be possible otherwise. To screen the thousands of seeds required for this project (and others), we developed a high throughput elemental profiling facility that includes a custom built weighing robot and an ICP-MS. To simplify the logistics of planting and harvesting the 18,000 to 20,000 field grown plants in the M_3 plants, families were planted in rows and the rows were bulk harvested. In this design, 25% of the loci will be homozygous in all the maternal plants in a row and an additional 19% will be homozygous in approximately one-fourth of the plants in a row. The seed ionome is a summation of many plant processes, and most ionomics mutants identified to date are in genes that function in the root (Rus et al., 2006; Baxter et al., 2008, 2009; Chao et al., 2011). We therefore propose that the seed ionome is primarily indicating the phenotype of the maternal plant. The low variation within plant as compared to between plants (Table 3) is consistent with this hypothesis.

Given the number of segregating loci in each line and number of different independent mutagenized families that must be screened to discover mutants, it is inherently difficult to replicate plots to minimize spatial variation within the field. As most of the elemental content of the seed comes from the soil, ionomic phenotypes are likely to be particularly susceptible to spatial variation. In addition, the majority of mutant lines had no discernable effect on the ionome, allowing sibling families to be used as controls for the effect of position in the field on the ionome. To determine the affects of spatial variation on the ionome, we plotted the normalized levels of each element by the field location of the row, which in this case also plotted them in run order. We removed elements below the level of detection, prone to contamination, and prone to analytical interference from the dataset. To reduce the

effect of spatial variation on the remaining elements, we normalized each plot to the plots surrounding it. This technique is not likely to remove all spatial variation nor would it affect lines where the variation affected a single plot, but it should account for gross variation between areas of the field (Fig. 1). One caveat to our analysis that complicates the interpretation of the effects is field-based environmental variation is confounded by analytical variation. We did not randomize the seeds from the original screen (current practice is to randomize every experiment) so we could not separate spatial variation from variation introduced by sample processing.

The variation due to spatial and other field effects has the potential to confound mutant identification, but unlike most ionomic leaf screens that profile a single plant, profiling multiple seeds per line increases the power to detect homozygous mutants. Previous ionomics screens have identified putative mutants from a single sample by either human inspection of z-score plots (Lahner et al., 2003) or by selecting all lines with at least one element statistically different from the population (Chen et al., 2009). The statistical approach is both unbiased and quick but lacks additional context (which elements are measured well and/or which elements may be more important to the researcher), may miss strong phenotypes that are segregating within a row, and may miss multi-element phenotypes that are below the significance threshold in all elements. Human interpretation of z-scores plots can does not have the aforementioned deficiencies but is time consuming and prone to bias (Nickerson, 1998).

Here we took two approaches, visual scoring only and using statistical methods to rank mutants followed by visual inspection of z-scores. The overlap between the two methods was quite high, with only three lines selected by human scoring not showing up on the statistical lists. While there were more lines identified by the computational methods and not by human selection, it should be noted that the human list was generated specifically to identify lines for the reproducibility testing experiment, which put a limit to the number of lines selected. There was also large overlap between the Zmax and Zsum methods, indicating that most mutants with multiple element phenotypes will have one element that is significantly different. Sixty-three of the lines in the combined Zmax and Zsum confirmed list were in the 155 lines identified by the *t* test with multiple testing correct ($q < 0.001$), suggesting that the *t* test method alone is not sufficient to select mutants for human inspection (Fig. 3). The R code for all methods is available at http://www.ionomicshub.org/home/PiiMS/dataexchange, so all methods can be used and compared by other researchers taking an ionomic approach to identifying mutants.

A large majority (77%) of the lines selected for regrowth had at least one element significantly different in the regrowth, 41% matching the phenotype from the preliminary screen well while 36% only partially matched the phenotype. The large number of lines that did not match the phenotype from the first round screen but were still significantly different in other elements is not surprising. Some ionomic phenotypes appear to be environmentally plastic, both *esb1* and *tsc10a* in *Arabidopsis thaliana* displayed varying phenotypes when grown in different soil batches (http://www.ionomicshub.org/arabidopsis). The main reason for the higher confirmation rate than in previous studies of ionomic mutants (Lahner et al., 2003; Chen and Boutros, 2011) is likely that the selection of mutants is based on multiple samples and is therefore less sensitive to false positive selection of mutants among single outliers generated by biological or analytical variation. The 4% of the lines from the initial screen that confirmed after rescreen is consistent with the estimate of 2 to 4% of the genome having a strong effect on the ionome (Lahner et al., 2003).

The confirmation rate of the putative mutants identified in this screen indicates that the visual rating of z-score plots of $n = 6$ or more seeds is highly effective for identifying potential mutants. However, the number of samples required for a screen of large populations, such as the fast neutron population from the soybean mutagenesis project (Bolon et al., 2011), would be prohibitive. We reasoned that potential lines could be identified with a much lower *n* screen and then a subset with promising characteristics could be further profiled to get the replication necessary for visual screening. This design requires the matrix-matched control solution to link the samples from the prescreen to the follow-up screen for the combined z-score analysis. To test this, we used our data to evaluate in silico several scenarios with a lower *n* prescreen. The results demonstrate that while there might be lines with real phenotypes that do not make it through the prescreen, an $n = 2$ prescreen followed by an $n = 4$ follow-up (for a total of $n = 6$) of the top 200 ranked mutants would still allow for identification of 75% of the putative mutants. In the case of the population studied in this paper, this design would cut the number of samples in the primary screen in half allowing for the entire first round screen to be completed in less than a month. Losing 25% of the putative mutants may have limited consequences for the final confirmed data sets as the missing 25% are likely to come from the lower ranked mutants. As an illustrative example, assuming a 15,000 sample experiment with a true mutation rate of 5%: 2500 lines could be screened at $n = 6$ to identify putative mutants, approximately 125 of which would be confirmed mutants, or 5000 lines could be screened at $n = 2$ to identify 1250 lines for an additional $n = 4$ so putative mutants could be selected with $n = 6$ data. In the second scenario, even if only 75% of true 250 mutants were among the 1250 rescreened, an additional approximately 62 mutants would be identified.

Conclusions

We have demonstrated that high throughput elemental profiling approaches can be used to identify elemental composition mutants in soybean. Using a multistage process where few seeds from all lines are sampled followed by profiling

of a relatively large subset at greater depth, a population of approximately 1000 lines can be screened in less than a month. The data and scripts from this manuscript are available on the Ionomics Hub File Transfer Page (http://www.ionomicshub.org/home/PiiMS/dataexchange).

Supplemental Information Available

This article includes supplemental information.

References

Baxter, I., and B. Dilkes. 2012. Elemental profiles reflect plant adaptations to the environment. Science 336:1661–1663. doi:10.1126/science.1219992

Baxter, I., P.S. Hosmani, A. Rus, B. Lahner, J.O. Borevitz, B. Muthukumar, M.V. Mickelbart, L. Schreiber, R.B. Franke, and D.E. Salt. 2009. Root suberin forms an extracellular barrier that affects water relations and mineral nutrition in *Arabidopsis*. PLoS Genet. 5(5):e1000492. doi:10.1371/journal.pgen.1000492

Baxter, I., B. Muthukumar, H.C. Park, P. Buchner, B. Lahner, J. Danku, K. Zhao, J. Lee, M.J. Hawkesford, M.L. Guerinot, and D.E. Salt. 2008. Variation in molybdenum content across broadly distributed populations of *Arabidopsis thaliana* is controlled by a mitochondrial molybdenum transporter (MOT1). PLoS Genet. 4(2):e1000004. doi:10.1371/journal.pgen.1000004

Bolon, Y.T., W.J. Haun, W.W. Xu, D. Grant, M.G. Stacey, R.T. Nelson, D.J. Gerhardt, J.A. Jeddeloh, G. Stacey, G.J. Muehlbauer, J.H. Orf, S.L. Naeve, R.M. Stupar, and C.P. Vance. 2011. Phenotypic and genomic analyses of a fast neutron mutant population resource in soybean. Plant Physiol. 156:240–253. doi:10.1104/pp.110.170811

Chao, D.Y., K. Gable, M. Chen, I. Baxter, C.R. Dietrich, E.B. Cahoon, M.L. Guerinot, B. Lahner, S. Lu, J.E. Markham, J. Morrissey, G. Han, S.D. Gupta, J.M. Harmon, J.G. Jaworski, T.M. Dunn, and D.E. Salt. 2011. Sphingolipids in the root play an important role in regulating the leaf ionome in *Arabidopsis thaliana*. Plant Cell 23:1061–1081.

Chen, H., and P.C. Boutros. 2011. VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics 12:35–35. doi:10.1186/1471-2105-12-35

Chen, Z., T. Shinano, T. Ezawa, J. Wasaki, K. Kimura, M. Osaki, and Y. Zhu. 2009. Elemental interconnections in *Lotus japonicus*: A systematic study of the affects of elements additions on different natural variants. Soil Sci. Plant Nutr. 55:91–101. doi:10.1111/j.1747-0765.2008.00311.x

Cooper, J.L., B.J. Till, R.G. Laport, M.C. Darlow, J.M. Kleffner, A. Jamai, T. El-Mellouki, S. Liu, R. Ritchie, N. Nielsen, K.D. Bilyeu, K. Meksem, L. Comai, and S. Henikoff. 2008. TILLING to detect induced mutations in soybean. BMC Plant Biol. 8:9. doi:10.1186/1471-2229-8-9

Davies, L., and U. Gather. 1993. The identification of multiple outliers. J. Am. Stat. Assoc. 88:782–792. doi:10.1080/01621459.1993.10476339

Haugen, J.-E., O. Tomic, and K. Kvaal. 2000. A calibration method for handling the temporal drift of solid state gas-sensors. Anal. Chim. Acta 407:23–39. doi:10.1016/S0003-2670(99)00784-9

Lahner, B., J. Gong, M. Mahmoudian, E.L. Smith, K.B. Abid, E.E. Rogers, M.L. Guerinot, J.F. Harper, J.M. Ward, L. McIntyre, J.I. Schroeder, and D.E. Salt. 2003. Genomic scale profiling of nutrient and trace elements in *Arabidopsis thaliana*. Nat. Biotechnol. 21:1215–1221.

Messina, M.J. 1999. Legumes and soybeans: Overview of their nutritional profiles and health effects. Am. J. Clin. Nutr. 70:439S–450S.

Nickerson, R.S. 1998. Confirmation bias: A ubiquitous phenomenon in many guises. Rev. Gen. Psychol. 2:175–220. doi:10.1037/1089-2680.2.2.175

Rus, A., I. Baxter, B. Muthukumar, J. Gustin, B. Lahner, E. Yakubova, and D.E. Salt. 2006. Natural variants of AtHKT1 enhance Na+ accumulation in two wild populations of *Arabidopsis*. PLoS Genet. 2(12):e210. doi:10.1371/journal.pgen.0020210

Wickham, H. 2007. Reshaping data with the reshape package. J. Stat. Software 21:1–20.