Canopy position has a profound effect on soybean seed composition

Steven C Huber [1, 2, 3], **Kunzhi Li** [1, 2, 4], **Randall Nelson** [1, 3], **Alexander Ulanov** [5], **Catherine M DeMuro** [1], **Ivan Baxter** Corresp. [6, 7]

[1] Global Change and Photosynthesis Research Unit, United States Department of Agriculture, Agricultural Research Service, Urbana, IL, United States
[2] Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL, United States
[3] Department of Crop Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States
[4] Lab of plant nutrition genetic engineering, Kunming University of Science and Technology, Kunming, Yunnan, China
[5] Metabolomics Facility, Carver Biotechnology Center, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States
[6] Plant Genetics Research Unit, United States Department of Agriculture Agricultural Research Service, St. Louis, Missouri, United States
[7] Donald Danforth Plant Science Center, Creve Coeur, Missouri, United States

Corresponding Author: Ivan Baxter
Email address: ivan.baxter@ars.usda.gov

Although soybean seeds appear homogeneous, their composition (protein, oil and mineral concentrations) can vary significantly with the canopy position where they were produced. In studies with 10 cultivars grown over a 3-yr period, we found that seeds produced at the top of the canopy have higher concentrations of protein but less oil and lower concentrations of minerals such as Mg, Fe, and Cu compared to seeds produced at the bottom of the canopy. Among cultivars, mean protein concentration (average of different positions) correlated positively with mean concentrations of S, Zn and Fe, but not other minerals. Therefore, on a whole plant basis, the uptake and allocation of S, Zn and Fe to seeds correlated with the production and allocation of reduced N to seed protein; however, the reduced N and correlated minerals (S, Zn and Fe) showed different patterns of allocation among node positions. For example, while mean concentrations of protein and Fe correlated positively, the two parameters correlated negatively in terms of variation with canopy position. Altering the microenvironment within the soybean canopy by removing neighboring plants at flowering increased protein concentration in particular at lower node positions and thus altered the node-position gradient in protein (and oil) without altering the distribution of Mg, Fe and Cu, suggesting different underlying control mechanisms. Metabolomic analysis of developing seeds at different positions in the canopy suggests that availability of free asparagine may be a positive determinant of storage protein accumulation in seeds and may explain the increased protein accumulation in seeds produced at the top of the canopy. Our results establish node-position variation in seed constituents and provide a new experimental system to identify genes controlling key aspects of seed composition. In addition, our results provide an

unexpected and simple approach to link agronomic practices to improve human nutrition and health in developing countries because food products produced from seeds at the bottom of the canopy contained higher Fe concentrations than products from the top of the canopy. Therefore, using seeds produced in the lower canopy for production of iron-rich soy foods for human consumption could be important when plants are the major source of protein and human diets can be chronically deficient in Fe and other minerals.

Canopy position has a profound effect on soybean seed composition

Steven C Huber[1,2,3] ^*, Kunzhi Li[1,2,4,^], Randall Nelson[1,3^], Alexander Ulanov[5], Catherine M. DeMuro[1], and Ivan Baxter[6,7]*

1 Global Change and Photosynthesis Research Unit, United States Department of Agriculture, Agricultural Research Service, Urbana, IL, United States

2 Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL, United States

3 Department of Crop Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States

4 Lab of plant nutrition genetic engineering, Kunming University of Science and Technology, Kunming, Yunnan, China

5 Metabolomics Facility, Carver Biotechnology Center, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States

6 Plant Genetics Research Unit, United States Department of Agriculture Agricultural Research Service, St. Louis, Missouri, United States

7 Donald Danforth Plant Science Center, Creve Coeur, Missouri, United States

^These authors contributed equally to this work

*Correspondence: ivan.baxter@ars.usda.gov or steven.huber@ars.usda.gov

Abstract

Although soybean seeds appear homogeneous, their composition (protein, oil and mineral concentrations) can vary significantly with the canopy position where they were produced. In studies with 10 cultivars grown over a 3-yr period, we found that seeds produced at the top of the canopy have higher concentrations of protein but less oil and lower concentrations of minerals such as Mg, Fe, and Cu compared to seeds produced at the bottom of the canopy. Among cultivars, mean protein concentration (average of different positions) correlated positively with mean concentrations of S, Zn and Fe, but not other minerals. Therefore, on a whole plant basis, the uptake and allocation of S, Zn and Fe to seeds correlated with the production and allocation of reduced N to seed protein; however, the reduced N and correlated minerals (S, Zn and Fe) showed different patterns of allocation among node positions. For example, while mean concentrations of protein and Fe correlated positively, the two parameters correlated negatively in terms of variation with canopy position. Altering the microenvironment within the soybean canopy by removing neighboring plants at flowering increased protein concentration in particular at lower node positions and thus altered the node-position gradient in protein (and oil) without altering the distribution of Mg, Fe and Cu, suggesting different underlying control mechanisms. Metabolomic analysis of developing seeds at different positions in the canopy suggests that availability of free asparagine may be a positive determinant of storage protein accumulation in seeds and may explain the increased protein accumulation in seeds produced at the top of the canopy. Our results establish node-position variation in seed constituents and provide a new experimental system to identify genes controlling key aspects of seed composition. In addition, our results provide an unexpected and simple approach to link agronomic practices to improve human nutrition and health in developing countries because food products produced from seeds at the bottom of the canopy contained higher Fe concentrations than products from the top of the canopy. Therefore, using seeds produced in the lower canopy for production of iron-rich soy foods for human consumption could be important when plants are the major source of protein and human diets can be chronically deficient in Fe and other minerals.

Introduction

Although soybean seeds from a given plant may appear physically homogeneous, it has long been known that seed produced at the top of the canopy can have higher protein and less oil compared to seeds from the bottom of the canopy (Collins & Cartter 1956). Subsequently it was demonstrated that positional effects are observed with determinate as well as indeterminate soybeans (Escalante & Wilcox 1993a) and in normal protein as well as high-protein breeding lines (Escalante & Wilcox 1993b). While these effects on protein and oil concentrations have been documented to occur, they are nonetheless not widely recognized today and there are no insights concerning possible physiological mechanisms that may underlie these positional effects. There are many other important seed constituents, in particular minerals, but the impact of canopy position on many of these seed constituents is unknown. Because legumes like soybean can contribute not only protein to the human diet but also minerals like iron (Fe) and zinc (Zn), canopy position effects on the concentrations of essential minerals could be important, especially for the health and nutrition of children and women. According to the World Health Organization, Fe deficiency is currently the most widespread mineral deficiency affecting more than 30% of the world's population (http://www.who.int/nutrition/topics/ida/en/). One approach to control this problem is to increase Fe intake via dietary diversification with Fe-rich foods and it is possible that variation with canopy position could be exploited.

Several factors could affect the development of seeds at the top of the plant differently than those at the bottom of the canopy and therefore could be responsible for differences in seed composition at maturity. First, flowering in the indeterminate soybean plants as used in the present study occurs first at lower nodes; thus, there is the potential for seeds lower in the canopy to develop over a longer period. However, while there is a lot of information about node position and flowering, there are few reports that have documented differences in duration of the seed fill period (SFP) as a function of node, as was demonstrated in cultivar 'Williams79' (Raboy & Dickinson 1987). A second factor is that seeds lower in the canopy also develop under altered environmental conditions in terms of temperature, irradiance, light quality and humidity, which are recognized to impact soybean seed composition (Carrera et al. 2009; Carrera et al. 2011; Wolf et al. 1982). A third factor is the contribution of remobilization of reserves, including minerals, from leaves that may vary among minerals and with node position.

Therefore, the role of canopy microenvironment and node position on seed composition warrants further consideration.

In the present study, we grew a core group of ten soybean lines in Urbana, IL, over a 3-yr period and monitored seed composition (protein, oil and mineral element concentration) at maturity as a function of node position. In general, there was a continuum in composition with seed that developed at the top of the canopy having more protein but less oil and reduced concentrations of minerals such as Mg, Fe, and Cu compared to seeds produced at the bottom of the canopy. Of particular note was the variation in Fe concentration, which was generally ~20% higher in seeds from the bottom of the canopy. The differences in mineral concentrations such as Fe could have direct impact on use of soybeans for human food in countries that primarily depend on plant protein sources for intake of minerals. We also tested several possible developmental and micro-environmental factors for their ability to influence the seed compositional gradients, and used metabolomic profiling of developing seeds to investigate biochemical determinants of the protein and oil gradients. Collectively, the results establish a new type of seed heteromorphism in soybean where seeds appear physically homogenous but differ in composition and provide new insights to some of the underlying factors that may be responsible for the gradients in composition from bottom to top of the canopy.

Materials and Methods

Plant growth and sampling

Soybean lines were grown at the University of Illinois South Farm, Urbana, IL, in a randomized complete block design with three replicates each year. Each plot consisted of three rows 2.5 m long, with 0.75 m between rows and a planting density of roughly 30 seeds m^{-1}. To produce the thinning treatment, all but three plants were removed from each row shortly after flowering. Delaying thinning until after the reproductive period had begun minimized branching on the remaining plants. Approximately 20 cm of plants were thinned from the ends of each row and the third plant was left in the middle in the row. The remaining plants were spaced approximately 1 m apart.

Plants were harvested at maturity. All plants were cut close to ground level and brought into the laboratory. Each stem was divided into four quadrants and the stem fractions in each quadrant were threshed together for each plot. Only normal-sized plants were included in the

analysis, and extremely small, wrinkled or off-color seeds were manually removed from all samples before analysis.

Soy products

To produce flour, soybeans were blanched (boiled for ~25 minutes) and then baked before grinding. To produce soymilk and okara (remaining solids), soybeans were blanched (boiled for ~5 min) twice and then ground in water and cooled slightly. The soymilk (liquid phase) and okara (solid phase) were separated using a cheesecloth and then dried separately and reground before analysis.

Seed storage product analysis

Protein and oil were measured with an Infratech 1241 Grain Analyzer (FOSS Analytical AB, Höganäs, Sweden), which is a true Near Infrared Transmission instrument that generates a spectrum from 850 to 1050 nm via the monochrome light source and mobile grating system. A 50-ml seed sample was used that allowed for 10 subsample readings reported on a 13 % moisture basis.

Ionomic analysis

Seed analysis was conducted as described in Ziegler et al. 2012. Briefly, single seeds from each quadrant were weighed using a custom-built seed weighing robot and then digested in concentrated nitric acid before loading onto an Elan ICP-MS. Internal standards were used to control for differences in dilution and sample injection. Leaf and soy products were analyzed in the same manner except that samples were added to digestion tubes by hand and weighed. Custom scripts were used to correct for internal standards and correct for sample weight.

Metabolomic analysis

Metabolome analysis was done through Metabolomics Center, Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign. Frozen seeds of the cultivar 'Williams 82' with attached seed coats were homogenized in liquid nitrogen and about 25 mg FW was extracted at room temperature with 1 mL of 50% methanol followed by addition of 800 µ1 of methanol:chloroform (1:2) as outlined in **Supplemental File 7**. Each extraction was followed by centrifugation (5 min at 15,000 g), and the supernatants were collected. With the

exception of samples for analysis of coenzymes, final extracts were evaporated under vacuum at -60 ˚C and subjected to GC/MS analysis.

Metabolic profiling: Dried extracts were derivatized with 100 µL methoxyamine hydrochloride (40 mg ml^{-1} in pyridine) for 90 min at 50 ˚C, then with 100 µL MSTFA at 50 ˚C for 120 min, and following 2-hour incubation at room temperature. 5 µL of the internal standard (hentriacontanoic acid, 10 mg ml^{-1}) was added to each sample prior to derivatization. Metabolites were analyzed using a GC-MS system (Agilent Inc, CA, USA) consisting of an Agilent 7890 gas chromatograph, an Agilent 5975 mass selective detector and a HP 7683B autosampler. Gas chromatography was performed on a ZB-5MS (60m×0.32mm I.D. and 0.25µm film thickness) capillary column (Phenomenex, CA, USA). The inlet and MS interface temperatures were 250 ˚C, and the ion source temperature was adjusted to 230 ˚C. An aliquot of 1 µL was injected with the split ratio of 10:1. The helium carrier gas was kept at a constant flow rate of 2 ml min^{-1}. The temperature program was: 5-min isothermal heating at 70 ˚C, followed by an oven temperature increase of 5 ˚C min^{-1} to 310 ˚C and a final 10 min at 310 ˚C. The mass spectrometer was operated in positive electron impact mode (EI) at 69.9 eV ionization energy at m/z 30-800 scan range.

Amino acid analysis: A 20 µl aliquot of the internal standard DL-chlorophenylalanine (1mg ml^{-1} in 0.1M HCI) was added to the extracts, dried under vacuum, derivatized with 50 µl of neat N-Methyl-N-tert-butyldimethylsilyltrifluoroacetamide (MTBSTFA), and 50 µL of acetonitrile at 80 ^{0}C for 4 h, cooled to room temperature and centrifuged briefly to remove condensate from the top of tube prior to injection of 1 µL at 5:1 split ratio into the GC/MS system, which consisted of an Agilent 6890N (Agilent Inc, Palo Alto, CA, USA) gas chromatograph, an Agilent 5973 mass selective detector and Agilent 7683B autosampler. Gas chromatography was performed on a 60 m ZB-5MS column with 0.32 mm inner diameter (I.D.) and 0.25 µm film thickness (Phenomenex, CA, USA) with injection temperature and MSD transfer line of 230 ˚C both, and the ion source adjusted to 230^{0}C. The helium carrier gas was set at a constant flow rate of 2 ml min^{-1}. The temperature program was 5 min at 150 ˚C, followed by an oven temperature ramp of 5 ˚C min^{-1} to 315 ˚C for a final 3 min. The mass spectrometer was operated in positive electron impact mode (EI) at 69.9 eV ionization energy in m/z 50-800 scan range. Acquired data were

normalized to the internal standard (DL-p-Chlorophenylalanine) and sample fresh weight. Amino acid concentrations were calculated based on 2to 75 µg ml^{-1} standard curves.

Free fatty acids, total fatty acids and coenzymes were also measured and values obtained used in the global analysis, but specific results are not presented. Detailed methods for the analysis are available on request.

The spectra of all chromatogram peaks were compared with electron impact mass spectrum libraries NIST08 (NIST, MD, USA), W8N08 (Palisade Corporation, NY, USA), and a custom-built database (460 unique metabolites). All known artificial peaks were identified and removed. To allow comparison between samples, all data were normalized to the corresponding internal standard and the sample fresh weight (FW). The spectra of all chromatogram peaks were evaluated using the AMDIS 2.71 (NIST, MD, USA) program. Metabolite concentrations were reported as concentrations relative to the internal standard (*i.e.*, target compound peak area divided by peak area of internal standard: $N_I = X_i \times X^{-1}_{IS}$) per gram sample weight. The instrument variability was within the standard acceptance limit (5%).

Metabolites with more than 50% of missing data were removed and for the rest of the metabolites, any missing data was imputed with one-half of the minimum positive value in the original data assuming their level was below the instrument detection limit. MVA and visualization was performed with SIMCA-P+ 12.0 software (Umetrics AB, Umeå, Sweden) and MetaboAnalyst (Xia & Wishart 2011) using log-transformed and autoscaled data and validated by sevenfold Cross-Validation and permutation with 500 random. To address the problem of multiple comparisons the False Discovery Rate (FDR) test was adopted. (Storey 2002).

Data analysis

Protein, oil, and elemental data were analyzed using R and the packages dplyr, ggplot2, grid, reshape2, qtlcharts and gplots. All data and analysis scripts used in the analysis are included as a supplemental file and are available on www.ionomicshub.org.

Results

Canopy position affects soybean seed protein, oil and mineral concentrations

We investigated positional effects with a core group of ten soybean lines (**Supplemental Table 1**) grown in Urbana, IL, over a 3-year period. Main stems were harvested at maturity and divided into four canopy position quadrants (**Fig. 1**) and the seeds collected from each quadrant were analyzed separately for major storage products (protein and oil) and various minerals. Representative results obtained for one cultivar ('Chamberlain') are presented in **Fig. 2A** with full plots provides as **Supplemental File 1**. As shown, protein concentration increased with node position at which seeds developed going from bottom to top of the mainstem while oil and iron (Fe) concentration decreased. For both protein and oil, which are the major seed constituents, there was variation in the absolute concentrations among the 3 years of study, but general trends were similar. Differences in absolute concentrations among years were most apparent for protein concentration with highest levels obtained in 2010 and lowest in 2011, presumably reflecting the impact of weather on seed development and composition. Another confounding source of variation for canopy position analysis is genotype, and **Fig. 2B** highlights the substantial variation in absolute concentrations of seed constituents due to both genotype and year. As expected, absolute concentrations of Mg, S, K, P and Ca were highest (> 1000 ppm); Mn, Fe, Rb, and Zn were intermediate (10 to 100 ppm), and Na, Co, Ni, Cu, Sr, Mo, and Cd were present at trace levels (< 10 ppm).

In order to compare positional effects for various parameters across genotypes and years without the confounding effects of differences in absolute values, we normalized each canopy gradient to a mean value of one and the values for each quadrant were then expressed relative to the normalized mean. However, because the weather in each year of the study differed (**Suppleemntal Table 2**), the normalized results for each parameter are presented separately for each year. Across the 10 soybean lines, oil concentration decreased progressively from bottom to top of the canopy and was associated with a reciprocal increase in protein concentration (**Fig. 3A**). Protein and oil concentrations in soybean seeds are usually inversely related (Wilcox 1998) and this was apparent with variation within the canopy as well. Single seed weight (designated as sample weight in Fig. 3A) varied with canopy position with seed produced in the middle portion tending to be slightly heavier than seeds produced at either the bottom or top of the canopy; however, the storage product gradients were independent of seed weight variation. Storage product gradients did not vary significantly across the three years of the study; however, absolute

protein and oil concentrations varied among the three years of the study (**Figure S1**), This is perhaps a result of weather that differed substantially in terms of temperature and precipitation among the three growing seasons (**Supplemental Table 2**).

We also found that canopy position significantly affected the seed ionome, which comprises all of the minerals and trace elements found in mature seeds (**Fig. 3B** and **Supplemental File 2)**. While there have been several studies of the soybean seed ionome (McGrath & Lobell 2013; Myers et al. 2014; Sha et al. 2012; Ziegler et al. 2013), to our knowledge this is the first report demonstrating variation with canopy position. **Figure 3B** shows normalized canopy gradient plots for elements where there was a statistically significant ($p < 0.01$) variation in concentration with position. Several groups of minerals exhibited common responses with canopy position. The elements Mg, Fe, Cu, and Cd were present at highest concentrations in seeds from the bottom of the canopy and decreased progressively to the top of the canopy. Within this group, the profiles for Mg and Fe were similar to one another in that variation was relatively low and the gradients were almost identical across the three years; however, the relative changes in Fe concentration were much greater in magnitude compared to changes in Mg concentration. Cu, Zn and Cd showed similar patterns, but were more variable among years. The second group that was apparent included Ca and Sr, where seeds from the middle of the canopy exhibited the lowest concentrations except in 2010, when concentrations of both Ca and Sr tended to increase going up the canopy. Finally, Mn was alone in the third category that increased in concentration towards the top of the canopy in all 3 years. Ca and Sr, and Cd and Zn, are chemically similar which may explain their parallel profiles. It is interesting to note that while Rb is a chemical analog of K and the two are often closely correlated (Baxter 2009), that was not the case for soybean seeds where significant position effects on Rb were observed (**Fig. 3B**) but not for K (see Supplemental file 2) It is also noteworthy that 2010 was the one year where mineral profiles were often distinct from those in 2011 and 2012. All three years were above normal in terms of temperature, but 2010 was the only year with above normal precipitation. Thus, water availability may be a major environmental factor impacting positional effects on the seed ionome, and interestingly some minerals were affected (Ca, Mn, Cu, Zn, Sr) while others (Mg, Fe, Co, Rb, Cd) were not. We also measured other minerals (B, Na, Al, P, S, K, Ni, As, Se and Mo) that did not show statistically significant variation with nodal position and are presented in Supplemental File 2.

Another way to compare canopy profiles for the minerals measured is to do an overall correlation matrix of quadrant variation normalized to plot averages. In this way, one can look across the entire data set for parameters that are correlated based on variation with nodal position. A strong positive correlation would indicate that both components changed not only in the same direction but also to the same relative extent. As shown in **Fig. 4A**, only a few strong correlations were apparent among the measured parameters. Variation in seed size (sample weight in Fig. 4A) did not significantly correlate with positional variation of any of the measured elements or storage products. Protein and oil concentrations were strongly negatively correlated, as expected. In terms of minerals and storage products, the quadrant variation in protein concentration correlated negatively with Fe and Cu, and positively with Mn, and the reciprocal pattern was apparent with oil concentration. Among the minerals, highly correlated element pairs included Fe-Cu, Ca-Sr, Ca-Mn, and Zn-Cu, and between P and S, Zn, and Co. As noted earlier, Ca and Sr are chemical analogs and frequently correlated (Baxter 2009), but surprisingly, other chemical analog pairs such as K-Rb were not observed. Fe and Cu were positively paired and have been reported to be positively correlated in soybean seeds (Vasconcelos et al. 2014) but the basis for the pairing is unknown. Correlations between P and minerals are often considered to reflect association of the mineral with seed phytate, the principal form of P in seeds (Vreugdenhil et al. 2004).

In addition to comparing parameters based on quadrant variation, it is also worthwhile to compare plot averages, which will reflect genetic and environmental effects on absolute values of the parameters. **Figure 4B** shows a matrix plot of correlations between plot means. Compared to the corresponding plot that focused on quadrant variation (**Fig. 4A**), many more strong correlations were apparent when comparing plot means. For example, protein concentration was positively correlated with S and Zn (and more weakly with Fe). The correlation with S is expected as the total seed S has been shown to track closely with high cysteine- and methionine-containing proteins in the soybean seed (Krishnan et al. 2012). The correlations between protein content, Zn and Fe could be due to their primary role as cofactors of metalloproteins or to variations senescence in leaves leading to nutrient remobilization (Uauy et al. 2006). Accordingly, there was a significant negative correlation of Fe, S, and Zn with oil concentration. Interestingly, there was also a strongly significant negative correlation of P with oil, whereas the positive correlation of P with protein concentration was relatively weak. The majority of mineral

correlations were positive in nature, with a maxi-cluster of Rb, Mn, Sr, Mg, Ni, and Na and a mini-cluster of Fe, Cu and Zn. The mini-cluster pairs of Fe-Cu and Cu-Zn were noted in the plot of **Fig. 4A**, but several members of the maxi-cluster correlation were not reported in the plot normalized correlation matrix. For example, Mn and Mg concentrations did not relate to each other in terms of quadrant variation but were strongly positively correlated based on plot means, indicating that mineral uptake may be similar but allocation among seeds in different quadrants is controlled separately. Finally, P concentration exhibited a positive correlation with Mn, Fe, Cu, Zn, S and Co. The link among P and Zn, S and Co concentrations with quadrant variation was observed (**Fig. 4A**), but when analyzed in terms of plot means in **Fig. 4B** the association of P with Mn, Fe, and Cu became apparent as well. It is worth noting that in terms of plot means, there was no association between Ca and Sr suggesting that these chemical analogs do not always behave similarly. There was one also a strong negative correlation between Mo and Sr, and Mo and S, perhaps suggesting a common component(s) of the uptake system. Readers can explore all of the correlations and the underlying data in using Supplemental Files 3,4 and 5.

Canopy microenvironment impacts seed composition

Our understanding of the environmental factors responsible for the positional effects on seed composition is limited; however, many microclimatic factors vary from the top to the bottom of the closed soybean canopy (Baldocchi et al. 1983). Environment is well known to impact soybean seed protein and oil composition (Rotundo & Westgate 2009). Therefore, we conducted experiments to broadly evaluate microclimatic differences within the canopy by thinning plants at flowering to remove the influence of neighboring plants. Removal of neighboring plants increased protein concentration at the expense of oil in seeds throughout the canopy of the spaced plants but the changes were greatest in pods lower on the main stem. As a result, the thinning treatment reduced the positional effect on protein and oil by 50-60% (**Fig. 5A**). Increased light energy to drive photosynthesis at most leaf positions and increased temperature at lower positions could both favor increased protein accumulation at lower nodes thereby reducing the difference between top and bottom seeds. However, while thinning significantly altered the main stem gradients in major storage products there was relatively little effect on minerals. As shown in **Fig. 5B**, the canopy positional effect on Mg, Fe and Cu was unaltered by the thinning treatment whereas Ca and Sr were similar to one another and showed a significant effect of thinning but only in one of the two test years (2010). The general conclusion is that thinning

affects the canopy positional effect on some but not all minerals. This suggests that at least for Mg, Fe and Cu, the transport and homeostasis mechanisms are generally independent of instantaneous environmental factors and the transport of sucrose and amino acids into the developing seeds is not the sole factor driving their movement into seeds.

Seed fill period and seed composition

Another factor that could contribute to canopy position effects on seed composition is the duration of the seed-fill period (SFP), which is affected by genetic and environmental factors and is one of the major determinants of yield potential in soybean (Evans et al. 1995). Soybeans flower in response to photoperiod and the first flowers form lower in the canopy followed by flowering at upper nodes. Pods then form in the same order and when fully elongated the process of seed development is considered to begin when seeds are approximately 0.34 centimeter long (by visual inspection). In general, seeds lower in the canopy fill over a longer period but at a lower rate compared to seeds at the top of the canopy (Raboy & Dickinson 1987) so that at maturity, final seed size tends to be rather constant through the canopy rather than increasing progressively from bottom to top of the canopy. We measured the SFPs with our core group of ten lines and found substantial differences in SFPs at the bottom and top of the canopy (**Supplemental Table 3**). Top SFP was generally correlated with bottom SFP, as would be expected, but the difference in SPF (bottom – top position) was not correlated with the canopy gradients of protein, oil, or Fe (**Fig. 6**). Therefore, factors other than the duration of the SFP are responsible for the documented variation in composition with nodal position.

Iron concentrations of soybean seed products

Our results raise the question of whether soy food products made from seed from different portions of the canopy would vary in terms of their mineral concentrations. Three of the most common and simplest products to make from soybean seeds are flour, milk and okara (the particulate material remaining after preparation of milk). Because Fe is one of the most critical minerals to human health and anemia is a global epidemic, we focused our initial analysis on the Fe content of these soy food products. We prepared flour from seven lines, and milk and okara from four lines and **Fig. 7** summarizes the results (All elements displayed in Supplemental file 6). With all three products, the concentration of Fe was highest in products made from seeds

produced at the bottom of the canopy and decreased progressively with canopy position of the seeds used. Thus, as would be expected the concentration of seed Fe affects the concentration of Fe in the flour, milk or okara produced from those seeds. Although many questions remain, the public health implications of our findings are apparent. Given that mineral content of seeds, especially Fe, is important our results uncover another source of variation that can be directly exploited.

The vegetative soybean ionome

The canopy effect on seed mineral concentration prompted us to look at the distribution of minerals in the shoots of vegetative plants. Minerals deposited in seeds are derived from continued uptake from the soil or remobilization of previously accumulated minerals (Hocking & Pate 1977; Waters & Grusak 2008), and therefore the leaf ionome of the vegetative plant is relevant to studies of the mature seed ionome. Consequently, we examined the leaf ionome from four genotypes as a function of canopy position. As shown in **Fig. 8,** the concentrations of Mg, Al, Ca, Mn, Fe, Co, As, and Sr were highest in leaves at the bottom of the canopy and decreased progressively to the top of the canopy. Concentrations of P, S, K, Cu, Zn, Rb, and Mo increased from bottom to top leaves. Na and Ni were present at low absolute concentrations and fluctuated but not in a progressive pattern as for the other minerals. Although leaves at different positions are often analyzed together (or as part of the 'shoot'), two previous studies with soybean also reported differences in mineral concentrations of lower, middle and upper leaves (Drossopoulos et al. 1994) or young and old leaves (corresponding to different node positions) (Duke et al. 2012) that are generally consistent with our results. The basis for differential accumulation of foliar minerals at different positions within the canopy is not clear and will be important to address in future studies. One possible explanation is that the greater phloem mobility of P and K facilitates their enhanced remobilization to upper nodes whereas other less mobile elements (e.g. Fe, Ca, and Mg) tend to remain at their point of initial deposition. This would not readily explain the observed profiles for Cu, Zn and Mo, however, highlighting the complexities involved in metal homeostasis and the significant variation with canopy position. Another working hypothesis could be that K, P, Cu, Zn and Mo are mineral markers of metabolic activity and accumulate in leaves at the top of the canopy that have highest rates of photosynthesis. Because minerals can be remobilized from leaves to developing seeds (Drossopoulos et al. 1994;

Jiménez et al. 1996; Sankaran & Grusak 2014), it is tempting to speculate that canopy seed gradient in Fe and Mg may be related to greater stores of both metals in leaves lower in the canopy. Opposite patterns were observed for other minerals (Ca, Mn, and Cu) suggesting that remobilization is either mineral specific or not quantitatively important in delivery of minerals to developing seeds.

A final point to note is that the potential exists for some soil particles to adhere to vegetative plant parts, especially lower in the canopy, while seeds are protected from soil contamination by the pods. Since some minerals exhibited opposite patterns, it seems that soil adhesion could not be completely responsible for the patterns observed.

Node position and the developing seed metabolome

Developing seeds were analyzed to determine whether canopy position affected seed metabolism sufficiently to explain the observed differences in protein and oil concentrations at maturity. To do this, we collected developing seeds (cultivar 'Williams 82') from the top and bottom of the canopy at several time points over a 24-h period as illustrated in **Figure S2**. Because seeds at the top and bottom of the canopy differed in size on the day of the experiment, seeds from the top of the canopy were also collected 6 days later when they had reached the same size (fresh weight seed^{-1}) as the bottom seeds on the first collection date. All seeds were at the stage of development where cell expansion and accumulation of storage compounds (protein and oil) were the dominant metabolic processes (Collakova et al. 2013). Untargeted metabolite profiling was conducted for analysis of polar compounds, free amino acids, free fatty acids, and total fatty acids (**Supplemental File 7**).

In general, most metabolites did not show diurnal changes in concentration, but there were differences in concentrations as a function of seed size and node position. The metabolite plots in **Fig. 9** illustrate some of the different patterns observed. The concentration of sucrose (**Fig. 9A**) in developing seeds did not vary diurnally and remained relatively constant but the concentration was slightly higher in the smallest seeds (day 1, top seed) compared to the larger seeds sampled at the bottom position on day 1 or top position on day 7. The decrease in sucrose concentration comparing top seed on day 1 and day 7 likely reflects in part the dilution effect caused by storage product accumulation as the seeds increased in size by roughly 2-fold. In contrast, the concentration of citrate in developing seeds was roughly equal among the three samples (**Fig.**

9B). These results suggest that seeds actually accumulate sucrose and to a larger extent citrate as they increase in dry matter during seed fill (thereby negating the dilution effect caused by seed growth). This also indicates that developing seeds have ample sugars and organic acids irrespective of size and node position and time of day. In marked contrast to sucrose and citrate were the dramatic differences observed in free asparagine (Asn) concentration (**Fig. 9C**), which was highest in top seed sampled on day 1 (Aug 20), and lowest in bottom seed sampled on the same day; the difference was roughly 8-fold. Sampling top seed on day 7 (Aug 26), when seed size was equivalent to that of bottom seed on day 1, still resulted in a ~4-fold elevation of free Asn concentration. The roughly 2-fold decrease in Asn concentration in seeds at the top of the canopy from day 1 to day 7 likely reflects the dilution effect of growth. The pattern for Asn concentration is potentially of interest because free Asn concentration during seed development correlates with protein concentration at maturity (Herman 2014b; Hernandez-Sebastia et al. 2005; Miller et al. 2008; Pandurangan et al. 2012). The results obtained in the present study suggest that greater supply of Asn to developing seeds at the top of the canopy may contribute to the observed greater accumulation of storage protein.

Importantly, Asn was also one of the important metabolites that distinguished the three sets of seeds collected based on a global metabolite analysis (**Supplemental File 8**). Mean values for Asn, and other protein amino acids are shown in **Fig. 10**. The concentrations of the free amino acids was highest in the small seed (top seed, day 1). Concentrations of Ala, Asn, Gly, and Thr were substantially higher in top seed at day 7 relative to bottom seed at day 1 (when seed sizes were similar). Of those amino acids, Asn was present at the highest absolute concentrations and may contribute to the storage protein biosynthesis either by acting as a signal metabolite or providing substrate for protein biosynthesis.

Discussion

The present study yields two major conclusions. First, the position along the main stem at which soybean seeds develop has a profound impact on seed composition, affecting the concentrations of protein, oil and certain minerals at maturity. Second, the canopy position effects on seed mineral concentrations (in particular Fe) are sufficiently large that there may be direct

implications for human nutrition in countries where plants are the main source of protein and soybeans are used for human food. These conclusions are discussed in more detail below.

Positional effects on seed protein and oil concentration are broadly observed

Results of the present study demonstrate that for 10 lines grown over a period of 3 years there were remarkably consistent gradients in protein and oil concentrations in mature seeds as a function of nodal position (**Figs. 2A and 3**). Increased concentration of oil in seeds from lower nodes could result from the increased duration of the SFP documented for lower pods (**Supplemental Table 3**) because the accumulation of oil in seeds often starts earlier than protein (Rotundo & Westgate 2009; Saldivar et al. 2011). However, oil accumulation tends to plateau before protein accumulation and therefore, percent oil will often decrease with increasing duration of the SFP rather than increase. In the present study, the protein and oil concentration gradients from bottom to top of the canopy were not correlated with the difference in SFP between the two positions (**Fig. 6**) and thus it appears that SFP does not determine the observed gradients in protein and oil concentration. Micro-environment appears to be one factor controlling protein and oil concentration gradients in the canopy because removal of neighboring plants at flowering increased protein concentration at all positions and decreased the difference between top and bottom nodes (**Fig. 5**). While it is not clear which micro-environmental factor(s) might actually be involved, we suggest that increased light energy reaching lower leaves may be a contributing factor. Metabolomic analysis of developing seeds that identified free Asn as one of the primary metabolites distinguishing seeds at the bottom and top of the canopy supports this conclusion. Asparagine is the major free amino acid in developing soybean seeds and differences in Asn concentration during development are positively correlated with protein concentration at seed maturity (Hernandez-Sebastia et al. 2005; Pandurangan et al. 2012). Furthermore, over-expression of asparaginase in soybean, driven by an embryo-specific promoter, resulted in a reduction in free Asn concentration during development and reduced protein concentration in mature seed, measured by nitrogen analysis (Pandurangan et al. 2015). Collectively, these results suggest that free Asn is a sensor or regulator of processes that determine protein accumulation in soybean seeds (Herman 2014a). Our results are consistent with this hypothesis and suggest that differences in free Asn concentration may explain the position effects on seed protein (and oil) concentration. Nitrogen and carbon flux into pods is

largely provided by nearest sources (Seddigh & Jolliff 1986; Streeter & Jeffers 1979) including the nearest trifoliolate leaves. We speculate that decreased light at lower positions in the closed canopy (i.e., with neighboring plants) would reduce leaf metabolism as well as the xylem flux of ureides and/or nitrate from roots to the lower leaves, thereby restricting the ability of those leaves to provide Asn (and Gln) to developing seeds. In contrast, removal of neighboring plants (in the 'thinned' plant treatment) would increase light at lower nodes thereby enhancing overall leaf metabolism and the flux of reduced nitrogen to subtending pods resulting in increased protein (and reduced oil) accumulation.

Positional effects on seed mineral concentration are documented

The concentration of minerals in seeds reflects the combined action of transport processes and regulation at multiple steps starting with mobilization from the soil, uptake into the root, and transport to the shoot for distribution among organs (Grusak et al. 1999; Waters & Grusak 2008). Deposition of some minerals in seeds can also involve remobilization from leaves during seed filling (Grusak et al. 1999; Hocking & Pate 1977), and it is interesting that different minerals show fundamentally different profiles of accumulation in seeds as a function of canopy position (**Fig. 3**). These differences could reflect alternate routes from the apoplast to the symplast or differences in mobility in the phloem (White 2012). Interestingly, minerals that tended to have highest concentrations in seeds at the bottom of the canopy (e.g., Mg, Fe, and Cu) are considered to have moderate to good phloem mobility compared to the minerals that tended to concentrate in the top of the canopy such as Mn (and in some cases Ca) that are considered to have poor phloem mobility. These results suggest that remobilization from leaves may be playing some role at least in the positional effects on the mature seed ionome. Another factor that may impact the distribution of minerals in seeds along the mainstem is precipitation. This speculation is based on the increased concentrations of Ca, Mn, and Sr found in seeds at the top of the canopy in 2010, which had above normal precipitation. It is possible that increased precipitation resulted in greater xylem transport of certain minerals (including Ca, Mn, and Sr) to developing seeds at the top of the canopy, or alternatively, that weather conditions in 2010 allowed greater remobilization of selected minerals from leaves via the xylem. It is recognized that while Ca and Mn are generally considered to have very low phloem mobility and are therefore not remobilized from senescing leaves, there is variation among species in the extent of remobilization (Maillard

et al. 2015). Conceivably, remobilization may also be triggered from leaves of all species under certain conditions.

While multiple seed constituents exhibited canopy concentration gradients, it seems unlikely that they are all caused by the same factors. Changing the microenvironment by thinning plants to allow increased light penetration into the canopy altered the protein and oil gradients but did not affect observed gradients for most of the minerals (**Fig. 5**). Furthermore, while the slope of many gradients changes across lines, treatment and year, the way that they change is not well correlated between the different constituents, as illustrated in the plot normalized correlation matrix (**Fig. 4A**), where relatively few strong correlations among the various parameters were apparent. However, numerous correlations were apparent when mean plot values were compared (**Fig. 4B**). Several minerals (e.g., P, Mn, Fe, Zn, S, and Co) had a negative relationship with oil concentration and increased with protein concentration. Thus, some coordination between seed storage product accumulation and mineral uptake into seeds is evident. However, the results suggest that total uptake of a mineral and the allocation among nodal positions are controlled by different mechanisms, and in general, canopy positional effects on minerals and protein/oil appear to be controlled by distinct mechanisms. It should be noted that altering the microenvironment by thinning plants did affect the observed gradients in seed concentrations of Ca, Mn, and Sr, which were also the minerals altered in distribution in 2010 (the year of this study with above normal precipitation). These results highlight the differences among minerals in terms of factors controlling their distribution among seed produced at different node positions. Clear, continued studies in the future will be required to sort out the different mechanisms involved.

Human nutrition implications for variation in seed composition

Soybeans are valued for their protein and oil content, but when used for human nutrition the content of minerals such as iron and zinc is also critically important. On a global scale, human iron deficiency is one of the most prevalent nutritional disorders (McLean et al. 2009) especially in countries where plant-based diets are prominent. As discussed above, nodal position affected the concentration of several minerals such as Mg, Fe, and Cu that were present at higher concentrations in seeds produced at the bottom of the canopy. Iron is of particular interest and was generally 20% higher in seeds produced lower in the canopy relative to the top and as

expected, differences in seed iron concentrations affected the concentration of iron in soy food products made from those seeds (**Fig. 7**). Soy flour preserved more Fe than did milk; perhaps mineral retention improvement through product preparation is possible. An immediate application of our results with respect to human nutrition would be to use seeds from the top and bottom halves of the canopy for different purposes, with seeds produced in the lower half reserved for production of iron-rich soy foods for human consumption. Thus, knowledge of these canopy position effects provides an unexpected approach to link agronomic practices to improve human nutrition and health.

New type of seed heteromorphism and implications for climate change impacts

Seed heteromorphism is well established (Matilla et al. 2005) but the seed heterogeneity documented here establishes a new category where an individual plant produces a continuum of seeds that differ in major aspects of their composition (protein, oil, and minerals) but are morphologically very similar. Overall, our results raise a number of questions and directions for future research. For example, it would be interesting to explore whether there are positional effects on soybean seed functional traits such as seed vigor or seedling stress tolerance. Because environment during reproductive development of plants is now recognized to broadly impact seed properties, such as growth performance and stress tolerance of the progeny (Biodner et al. 2007; Tricker et al. 2013), it will be interesting to further explore similar properties of soybean seed produced at the different parts of the canopy. Our results also raise the question of whether similar effects occur in other species including non-domesticated plants where there might be some ecological significance.

Another area that will be interesting to explore is the impact of elevated CO_2 on the canopy positional effects described in the present study. It was recently reported (Loladze 2014; Myers et al. 2014) that grain from many species, including soybean, have lower concentrations of Zn and Fe when plants are grown at elevated CO_2 thereby uncovering a new climate change challenge to global health. The meta-analysis established a ~5% reduction in soybean seed Fe and Zn concentrations at high CO_2. It is relevant to note that variation in seed Fe concentration with node position established in the present study is substantially larger (4-fold greater) compared to the impact of climate change on mean seed Fe concentration. Therefore, our results are likely to be meaningful from a quantitative standpoint and have important implications for

examining the impact of climate change on the seed ionome. For example, it will be interesting to determine how this overall reduction in mean seed Fe concentration at elevated CO_2 is related (if at all) to canopy position effects; is Fe reduced 5% in seeds at nodes throughout the canopy or are certain positions affected to a greater degree than others? Identifying the molecular mechanisms underlying canopy gradients in composition may provide new approaches to controlling soybean seed quality for various uses, including food for human consumption under conditions of global climate change.

Acknowledgements

The authors thank Kunming University of Science and Technology (KUST), Kunming City, P.R. China for supporting the visit of Prof. Kunzhi Li to UIUC; Karl E. Weingarter and Marilyn L. Nash for advice on soy food preparation and providing access to the Test Kitchen facility of the National Soybean Research Laboratory at the University of Illinois; Greg Ziegler for expert technical assistance with ionomic analysis and Sarah Schultz for help with field sampling. Support for this work was provided by the United Soybean Board and the USDA-ARS.

REFERENCES

Baldocchi DD, Verma SB, and Rosenberg NJ. 1983. Microclimate in the soybean canopy. *Agricultural Meteorology* 28:321-337. http://dx.doi.org/10.1016/0002-1571(83)90009-2

Baxter I. 2009. Ionomics: studying the social network of mineral nutrients. *Current Opinion in Plant Biology* 12:381-386. http://dx.doi.org/10.1016/j.pbi.2009.05.002

Biodner C, Goebel C, Feussner I, Gatz C, and Polle A. 2007. Warm and cold parental reproductive environments affect seed properties, fitness, and cold responsiveness in Arabidopsis thaliana progenies. *Plant, Cell & Environment* 30:165-175. 10.1111/j.1365-3040.2006.01615.x

Carrera C, Martínez MJ, Dardanelli J, and Balzarini M. 2009. Water Deficit Effect on the Relationship between Temperature during the Seed Fill Period and Soybean Seed Oil and Protein Concentrations. *Crop Science* 49:990-998. 10.2135/cropsci2008.06.0361

Carrera C, Martínez MJ, Dardanelli J, and Balzarini M. 2011. Environmental Variation and Correlation of Seed Components in Nontransgenic Soybeans: Protein, Oil, Unsaturated Fatty Acids, Tocopherols, and Isoflavones. *Crop Science* 51:800-809. 10.2135/cropsci2010.06.0314

Collakova E, Aghamirzaie D, Fang Y, Klumas C, Tabataba F, Kakumanu A, Myers E, Heath L, and Grene R. 2013. Metabolic and Transcriptional Reprogramming in Developing Soybean (Glycine max) Embryos. *Metabolites* 3:347-372.

Collins FI, and Cartter JL. 1956. Variability in Chemical Composition of Seed From Different Portions of the Soybean Plant1. *Agron J* 48:216-219. 10.2134/agronj1956.00021962004800050006x

Drossopoulos JB, Bouranis DL, and Bairaktari BD. 1994. Patterns of mineral nutrient fluctuations in soybean leaves in relation to their position. *Journal of Plant Nutrition* 17:1017-1035. 10.1080/01904169409364785

Duke SO, Reddy KN, Bu K, and Cizdziel JV. 2012. Effects of Glyphosate on the Mineral Content of Glyphosate-Resistant Soybeans (Glycine max). *Journal of Agricultural and Food Chemistry* 60:6764-6771. 10.1021/jf3014603

Escalante EE, and Wilcox JR. 1993a. Variation in Seed Protein among Nodes of Determinate and Indeterminate Soybean near-Isolines. *Crop Science* 33:1166-1168.

Escalante EE, and Wilcox JR. 1993b. Variation in Seed Protein among Nodes of Normal-Protein and High-Protein Soybean Genotypes. *Crop Science* 33:1164-1166.

Evans TA, Fuhrmann JJ, Nelson RL, and Vasilas BL. 1995. Relationship of nitrogen utilization patterns with soybean yield and seed-fill period. *Crop Science* 35:809+.

Grusak MA, DellaPenna D, and Welch RM. 1999. Physiologic processes affecting the content and distribution of phytonutrients in plants. *Nutrition reviews* 57:27-33.

Herman E. 2014a. Soybean Seed Proteome Rebalancing. *Frontiers in Plant Science* 5. 10.3389/fpls.2014.00437

Herman EM. 2014b. Soybean seed proteome rebalancing. *Front Plant Sci* 5:437. 10.3389/fpls.2014.00437

Hernandez-Sebastia C, Marsolais F, Saravitz C, Israel D, Dewey RE, and Huber SC. 2005. Free amino acid profiles suggest a possible role for asparagine in the control of storage-product accumulation in developing seeds of low- and high-protein soybean lines. *J Exp Bot* 56:1951-1963. 10.1093/jxb/eri191

Hocking PJ, and Pate JS. 1977. Mobilization of Minerals to Developing Seeds of Legumes. *Annals of Botany* 41:1259-1278.

Jiménez MP, Effrón D, de la Horra AM, and Defrieri R. 1996. Foliar potassium, calcium, magnesium, zinc, and manganese content in soybean cultivars at different stages of development. *Journal of Plant Nutrition* 19:807-816. 10.1080/01904169609365163

Krishnan HB, Jang S, Baxter I, and Wiebold WJ. 2012. Growing location has a pronounced effect on the accumulation of cancer chemopreventive agent Bowman-Birk inhibitor in soybean seeds. *Crop Science* 52. 10.2135/cropsci2011.11.0593

Loladze I. 2014. Hidden shift of the ionome of plants exposed to elevated CO2 depletes minerals at the base of human nutrition. *Elife* 3. 10.7554/eLife.02245

Maillard A, Diquélou S, Billard V, Laîné P, Garnica M, Prudent M, Garcia-Mina J-M, Yvin J-C, and Ourry A. 2015. Leaf mineral nutrient remobilization during leaf senescence and modulation by nutrient deficiency. *Frontiers in Plant Science* 6:317. 10.3389/fpls.2015.00317

Matilla A, Gallardo M, and Puga-Hermida MI. 2005. Structural, physiological and molecular aspects of heterogeneity in seeds: a review. *Seed Science Research* 15:63-76. doi:10.1079/SSR2005203

McGrath JM, and Lobell DB. 2013. Reduction of transpiration and altered nutrient allocation contribute to nutrient decline of crops grown in elevated CO(2) concentrations. *Plant Cell Environ* 36:697-705. 10.1111/pce.12007

McLean E, Cogswell M, Egli I, Wojdyla D, and de Benoist B. 2009. Worldwide prevalence of anaemia, WHO Vitamin and Mineral Nutrition Information System, 1993-2005. *Public Health Nutr* 12:444-454. 10.1017/S1368980008002401

Miller AJ, Fan X, Shen Q, and Smith SJ. 2008. Amino acids and nitrate as signals for the regulation of nitrogen acquisition. *J Exp Bot* 59:111-119. 10.1093/jxb/erm208

Myers SS, Zanobetti A, Kloog I, Huybers P, Leakey AD, Bloom AJ, Carlisle E, Dietterich LH, Fitzgerald G, Hasegawa T, Holbrook NM, Nelson RL, Ottman MJ, Raboy V, Sakai H, Sartor KA, Schwartz J, Seneweera S, Tausz M, and Usui Y. 2014. Increasing CO2 threatens human nutrition. *Nature* 510:139-142. 10.1038/nature13179

Pandurangan S, Pajak A, Molnar SJ, Cober ER, Dhaubhadel S, Hernandez-Sebastia C, Kaiser WM, Nelson RL, Huber SC, and Marsolais F. 2012. Relationship between asparagine metabolism and protein concentration in soybean seed. *J Exp Bot* 63:3173-3184. 10.1093/jxb/ers039

Pandurangan S, Pajak A, Rintoul T, Beyaert R, Hernandez-Sebastia C, Brown DC, and Marsolais F. 2015. Soybean seeds overexpressing asparaginase exhibit reduced nitrogen concentration. *Physiol Plant*. 10.1111/ppl.12341

Raboy V, and Dickinson DB. 1987. The timing and rate of phytic Acid accumulation in developing soybean seeds. *Plant Physiol* 85:841-844.

Rotundo JL, and Westgate ME. 2009. Meta-analysis of environmental effects on soybean seed composition. *Field Crops Research* 110:147-156. http://dx.doi.org/10.1016/j.fcr.2008.07.012

Saldivar X, Wang Y-J, Chen P, and Hou A. 2011. Changes in chemical composition during soybean seed development. *Food Chemistry* 124:1369-1375. http://dx.doi.org/10.1016/j.foodchem.2010.07.091

Sankaran RP, and Grusak MA. 2014. Whole shoot mineral partitioning and accumulation in pea (Pisum sativum). *Front Plant Sci* 5:149. 10.3389/fpls.2014.00149

Seddigh M, and Jolliff GD. 1986. Remobilization Patterns of C and N in Soybeans with Different Sink-Source Ratios Induced by Various Night Temperatures. *Plant Physiology* 81:136-141.

Sha Z, Oka N, Watanabe T, Tampubolon BD, Okazaki K, Osaki M, and Shinano T. 2012. Ionome of soybean seed affected by previous cropping with mycorrhizal plant and manure application. *J Agric Food Chem* 60:9543-9552. 10.1021/jf3024744

Storey JD. 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64:479-498. 10.1111/1467-9868.00346

Streeter JG, and Jeffers DL. 1979. Distribution of Total Non-Structural Carbohydrates in Soybean Plants Having Increased Reproductive Load. *Crop Science* 19:729-734. 10.2135/cropsci1979.0011183X001900050046x

Tricker P, López C, Gibbings G, Hadley P, and Wilkinson M. 2013. Transgenerational, Dynamic Methylation of Stomata Genes in Response to Low Relative Humidity. *International Journal of Molecular Sciences* 14:6674-6689.

Uauy C, Distelfeld A, Fahima T, Blechl A, and Dubcovsky J. 2006. A NAC Gene Regulating Senescence Improves Grain Protein, Zinc, and Iron Content in Wheat. *Science* 314:1298-1301. 10.1126/science.1133649

Vasconcelos MW, Clemente T, and Grusak MA. 2014. Evaluation of constitutive iron reductase (AtFRO2) expression on mineral accumulation and distribution in soybean (Glycine max. L). *Frontiers in Plant Science* 5. 10.3389/fpls.2014.00112

Vreugdenhil D, Aarts MGM, Koornneef M, Nelissen H, and Ernst WHO. 2004. Natural variation and QTL analysis for cationic mineral content in seeds of Arabidopsis thaliana. *Plant, Cell & Environment* 27:828-839. 10.1111/j.1365-3040.2004.01189.x

Waters BM, and Grusak MA. 2008. Whole-plant mineral partitioning throughout the life cycle in Arabidopsis thaliana ecotypes Columbia, Landsberg erecta, Cape Verde Islands, and the mutant line ysl1ysl3. *New Phytol* 177:389-405. 10.1111/j.1469-8137.2007.02288.x

White PJ. 2012. Chapter 3 - Long-distance Transport in the Xylem and Phloem A2 - Marschner, Petra. *Marschner's Mineral Nutrition of Higher Plants (Third Edition)*. San Diego: Academic Press, 49-70.

Wilcox JR. 1998. Increasing Seed Protein in Soybean with Eight Cycles of Recurrent Selection. *Crop Sci* 38:1536-1540. 10.2135/cropsci1998.0011183X003800060021x

Wolf RB, Cavins JF, Kleiman R, and Black LT. 1982. Effect of temperature on soybean seed constituents: Oil, protein, moisture, fatty acids, amino acids and sugars. *Journal of the American Oil Chemists' Society* 59:230-232. 10.1007/BF02582182

Xia J, and Wishart DS. 2011. Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. *Nat Protocols* 6:743-760.

Ziegler G, Terauchi A, Becker A, Armstrong P, Hudson K, and Baxter I. 2013. Ionomic Screening of Field-Grown Soybean Identifies Mutants with Altered Seed Elemental Composition. *The Plant Genome* 6. 10.3835/plantgenome2012.07.0012

Figure 1. Quadrants of a Soybean Plant. The mature plant is divided up into quadrants upon harvest and each quadrant is analyzed separately. Plat normalized data uses the average of all four quadrants to normalize year, plot and line affects.

Figure 2. Canopy Gradients of Seed Composition Traits Before Normalization and Line and Year Effects on total accumulation. A) Composition gradients from the bottom to the top of the canopy for cultivar 'Chamberlain'. The plots display the quadrant average as a line with the 95% confidence interval calculated using standard error as the ribbon. Units are mg (Single seed weight), PPM (Fe) and percentage for Protien and Oil) B) Year and line effects for each compositional trait, represented as boxplots. Units are PPM.

Figure 3. Canopy Gradients of Seed Composition Traits. For each trait, the data was normalized to the plot average to remove the effect of environment and genotype. The plots display the quadrant average as a line with the 95% confidence interval calculated using standard error as the ribbon. A) Percentage protein, percentage Oil and single seed weight. B) Elements with a significant ($p < 1$ e -10) effect of gradient in an ANOVA analysis that included Entry, Year and Position.

Figure 4. Correlation Plot among Composition Traits. Pearson correlation values between compositional traits. **A**) Correlation across 832 quadrants normalized to the plot average. **B**) Correlation across 208 plot means.

Figure 5. Effect of Thinning on Compositional Traits. For each trait, the data was normalized to the plot average to remove the effect of environment and genotype. The plots display the quadrant average as a line with the 95% confidence interval calculated using standard error as the ribbon. **A**) Percentage protein and percentage Oil in 2010. **B**) Elements (from 2010 and 2012) with a significant ($p < 1$ e -10) effect of gradient in an ANOVA analysis that included Entry, Year, Position and thinning.

Figure 6. Difference in Top/Bottom composition traits is not correlated with Seed Fill Period. The difference in plot normalized composition between the top quad and the bottom quad for protein, oil and iron plotted versus the difference in seed fill period for 51 plots in 2012.

Figure 7. Canopy differences in Iron are reflected in food products. Fe content of the products from 3 replicates of seven lines (flour) and four lines (Milk and Okara). Boxplots display the five number summary (median, 25, and 75% percentile define the box, with whiskers extending to 1.5 x interquartile range).

Figure 8. Canopy Gradients of Leaf Composition Traits. For each trait, the data was normalized to the plot average to remove the effect of environment and genotype. The plots display the quadrant average as a line with the 95% confidence interval calculated using standard error as the ribbon. Elements with a significant ($p < 1 e -10$) effect of gradient in an ANOVA analysis that included Entry, Collection Date and Position

Figure 9. Concentrations of selected primary metabolites in developing seeds of cultivar 'Williams 82'**. A**, Suc; **B**, citrate; and **C**, Asn. Boxplots display the five number summary (median, 25, and 75% percentile define the box, with whiskers extending to 1.5 x interquartile range) for three replicates at each sampling time: 7AM (7) , 12N (12) , 7PM (19) and the following morning at 7AM (31). The black vertical bars represent the intervening night period. Values are µg (g DW).

Figure 10. Concentrations of free amino acids in developing seeds**.** Boxplots display the five number summary (median, 25, and 75% percentile define the box, with whiskers extending to 1.5 x interquartile range) for values from each sampling interval (3 replicates and 4 time points are merged within each box) and nodal position. Ornithine levels reflect both ornithine and arginine as arginine is converted to ornithine during sample prep for GC-MS. D1.bot, D1.top and D7.top refer to the samples collected on day one top and bottom quadrants and the day seven top quadrant respectively.

Supplemental File 1. Raw Gradient plots for each line/year combination. Values are PPM (Elements), mg (SampleWeight), and Percentage (Protein/Oil).

Supplemental File 2. All Compositional Traits Normalized Gradients.

Supplemental File 3. Interactive Correlation plots of all plot normalized data.

Supplemental File 4 . Interactive Correlation plots of plot mean data.

Supplemental File 5. Interactive Correlation plots of all data, not normalized.

Supplemental File 6. Composition of Food products for all elements.

Supplemental File 7. Metabolomics analysis scheme.

Supplemental File 8. Global analysis of metabolome of developing soybean seeds. (A) PLS-DA scores plot (R^2 = 98.7%, Q2 = 81.1%, $P < 0.001$ by permutation test) of soybean seeds at different canopy position and time of day. (B) Variable Importance in the Projection (VIP) for the first component showing the fifteen most important compounds.

Supplemental Table 1. Indeterminate lines used in the present study and selected characteristics

Supplemental Table 2. Weather summary (June 1– August 31) during the 2010 to 2012 growing seasons

Supplemental Table 3. Genotype differences in Seed fill period (SFP) and the difference in SFP at two node positions (bottom minus top; delSFP)

Figure 1

Quadrants of a Soybean Plant.

The mature plant is divided up into quadrants upon harvest and each quadrant is analyzed separately. Plat normalized data uses the average of all four quadrants to normalize year, plot and line affects.

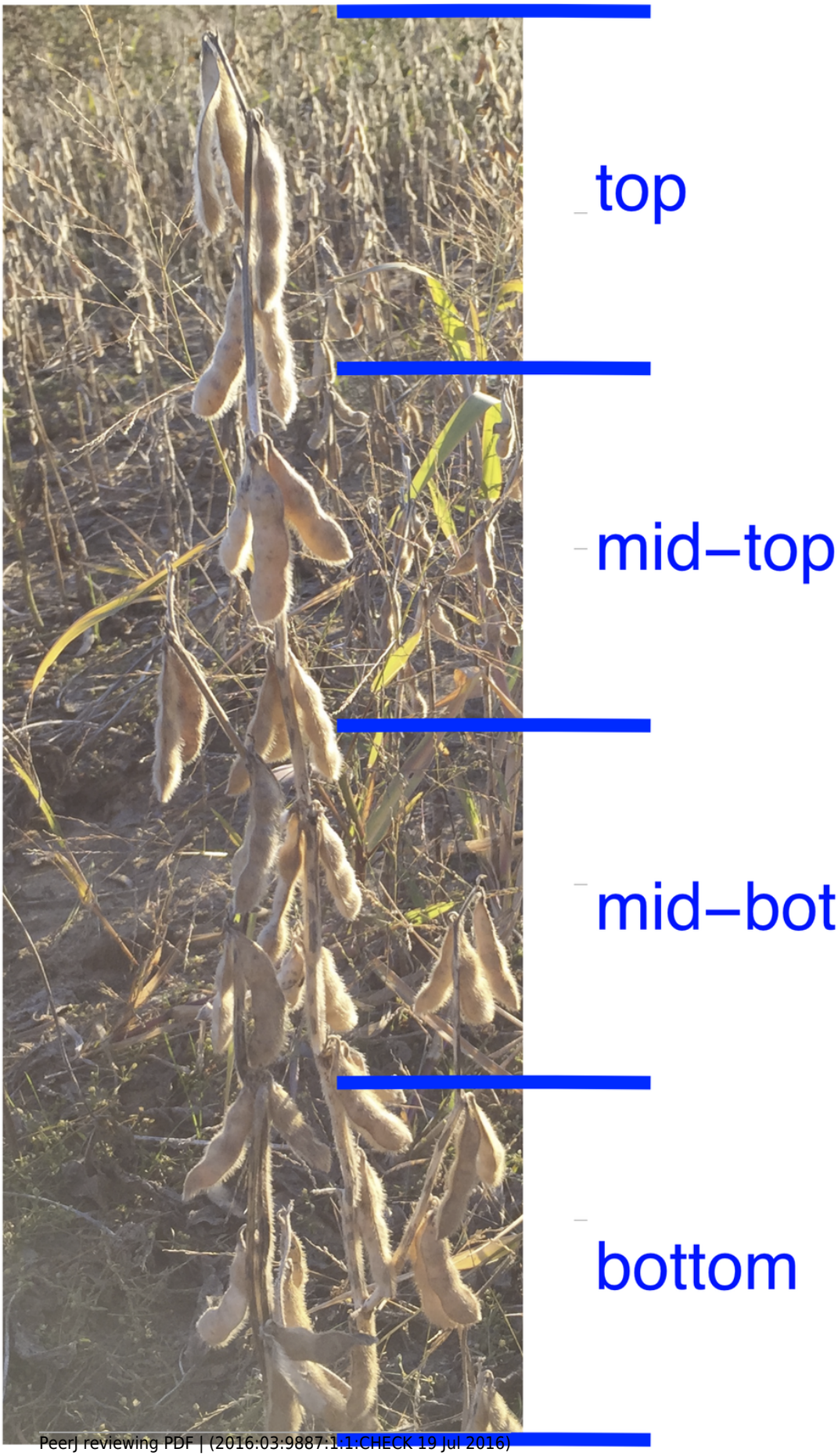

top
mid–top
mid–bot
bottom

Figure 2

Canopy Gradients of Seed Composition Traits Before Normalization and Line and Year Effects on total accumulation.

A) Composition gradients from the bottom to the top of the canopy for cultivar 'Chamberlain'. The plots display the quadrant average as a line with the 95% confidence interval calculated using standard error as the ribbon. Units are mg (Single seed weight), PPM (Fe) and percentage for Protien and Oil) B) Year and line effects for each compositional trait, represented as boxplots. Units are PPM.

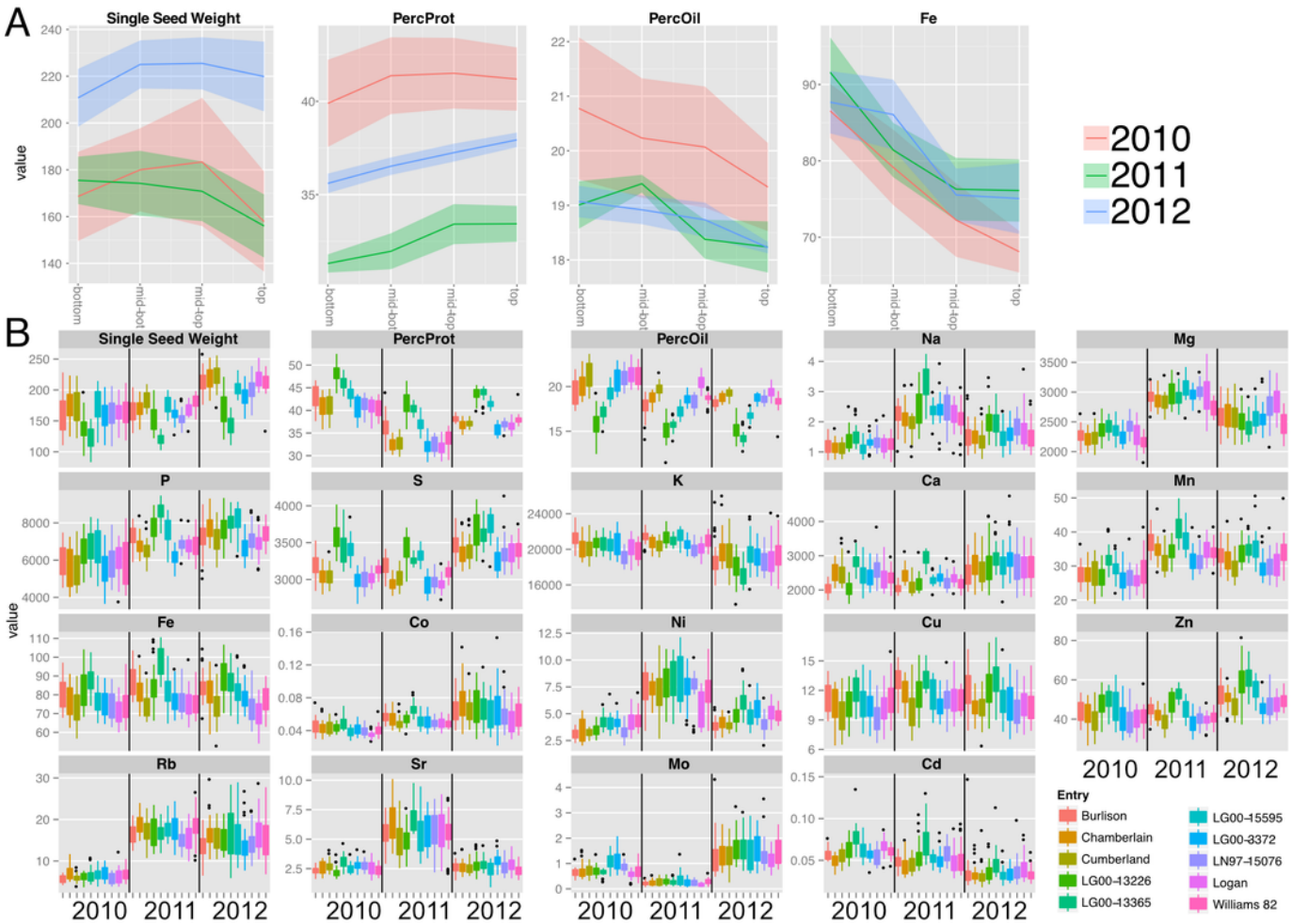


Figure 3

Canopy Gradients of Seed Composition Traits.

For each trait, the data was normalized to the plot average to remove the effect of environment and genotype. The plots display the quadrant average as a line with the 95% confidence interval calculated using standard error as the ribbon. A) Percentage protein, percentage Oil and single seed weight. B) Elements with a significant ($p < 1 e -10$) effect of gradient in an ANOVA analysis that included Entry, Year and Position.

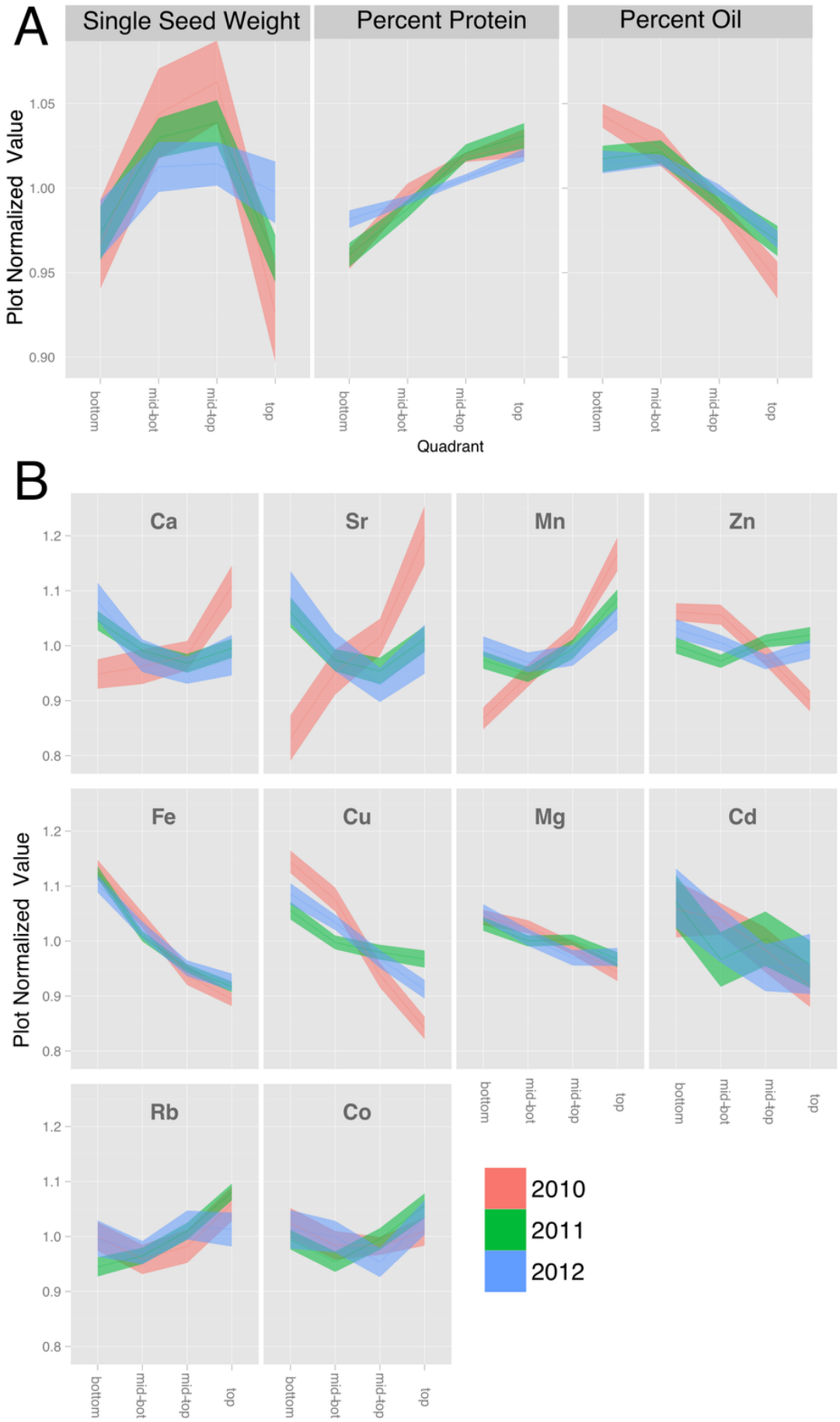

A
Single Seed Weight
Percent Protein
Percent Oil
Plot Normalized Value
1.05
1.00
0.95
0.90
bottom
mid-bot
mid-top
top
Quadrant
B
Ca
Sr
Mn
Zn
Fe
Cu
Mg
Cd
Rb
Co
1.2
1.1
1.0
0.9
0.8
2010
2011
2012

Figure 4

Correlation Plot among Composition Traits

Pearson correlation values between compositional traits. **A**) Correlation across 832 quadrants normalized to the plot average. **B**) Correlation across 208 plot means.

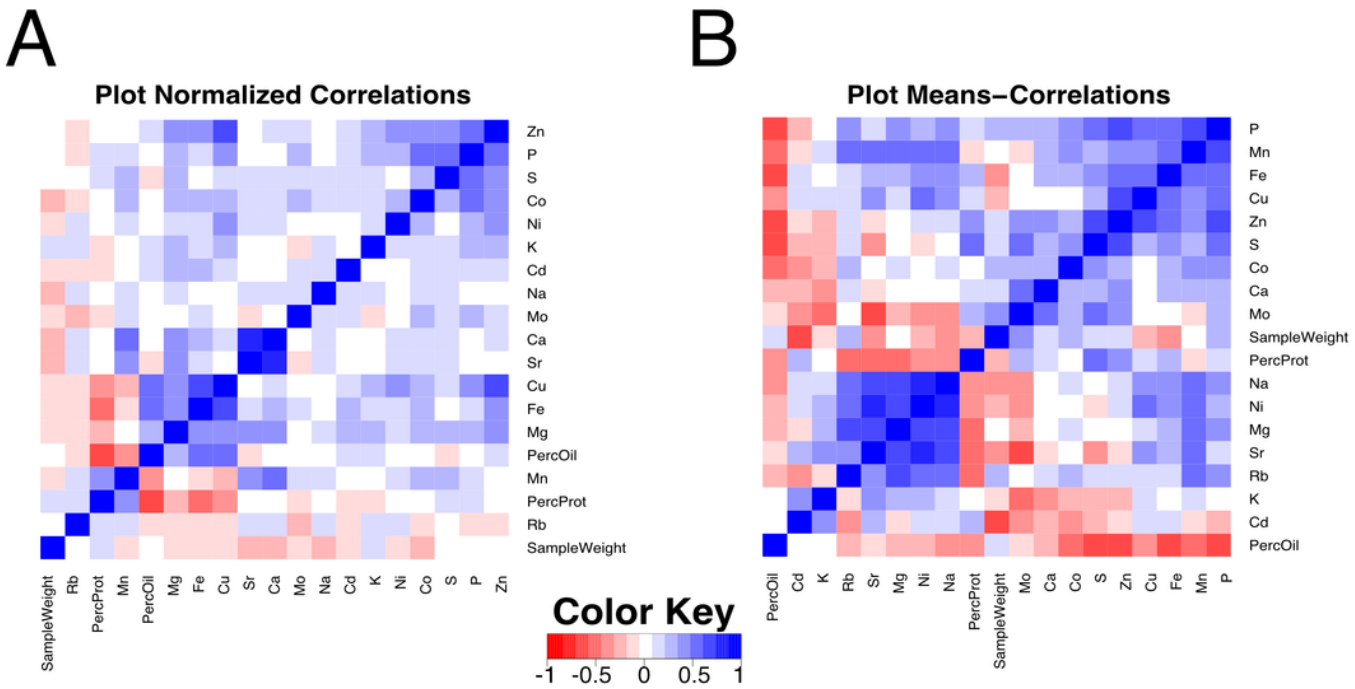


Figure 5

Effect of Thinning on Compositional Traits

For each trait, the data was normalized to the plot average to remove the effect of environment and genotype. The plots display the quadrant average as a line with the 95% confidence interval calculated using standard error as the ribbon. **A**) Percentage protein and percentage Oil in 2010. **B**) Elements (from 2010 and 2012) with a significant ($p < 1 e -10$) effect of gradient in an ANOVA analysis that included Entry, Year, Position and thinning.

A
Percentage Protein
Percentage Oil
Plot Normalized Value
1.05
1.02
0.99
0.96
0.93
bottom
mid-bot
mid-top
top
Quadrant
Dense
Thinned

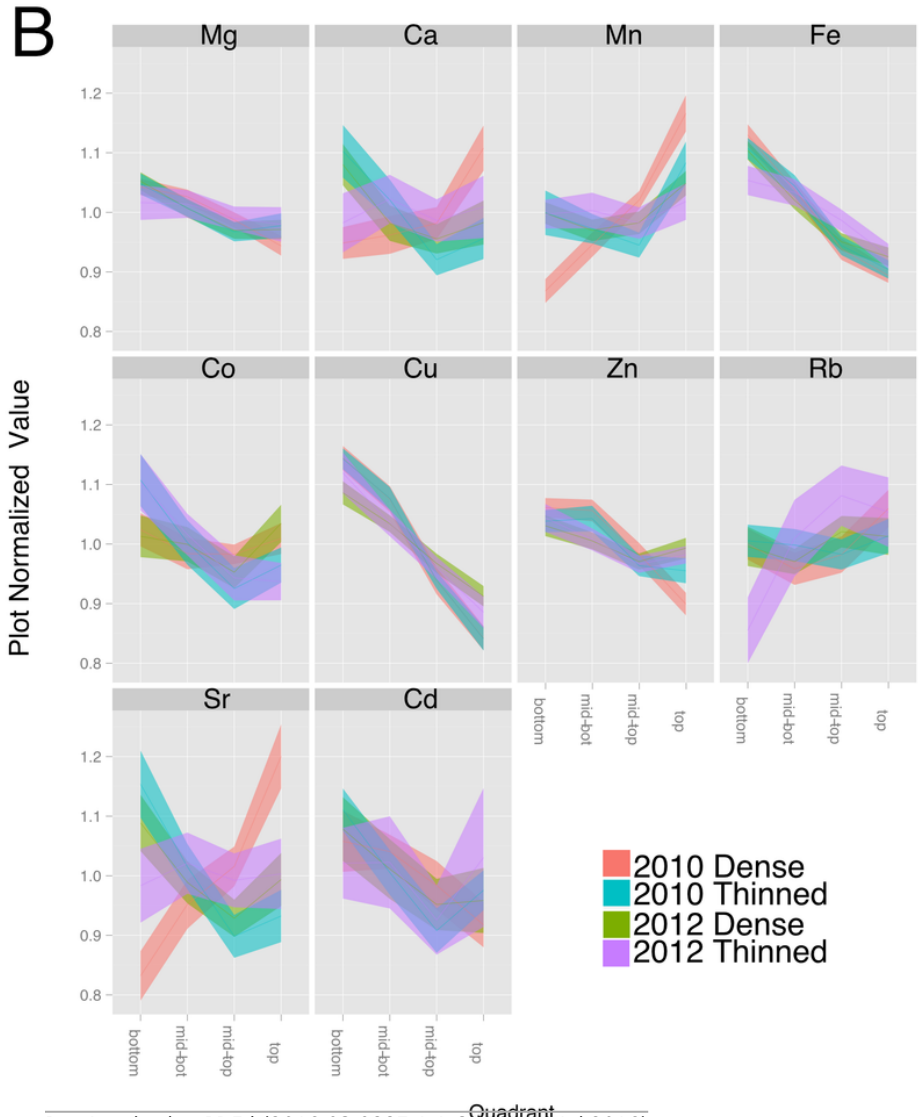

B
Mg
Ca
Mn
Fe
Co
Cu
Zn
Rb
Sr
Cd
Plot Normalized Value
1.2
1.1
1.0
0.9
0.8
bottom
mid-bot
mid-top
top
Quadrant
2010 Dense
2010 Thinned
2012 Dense
2012 Thinned

Figure 6

Difference in Top/Bottom composition traits is not correlated with Seed Fill Period

The difference in plot normalized composition between the top quad and the bottom quad for protein, oil and iron plotted versus the difference in seed fill period for 51 plots in 2012.

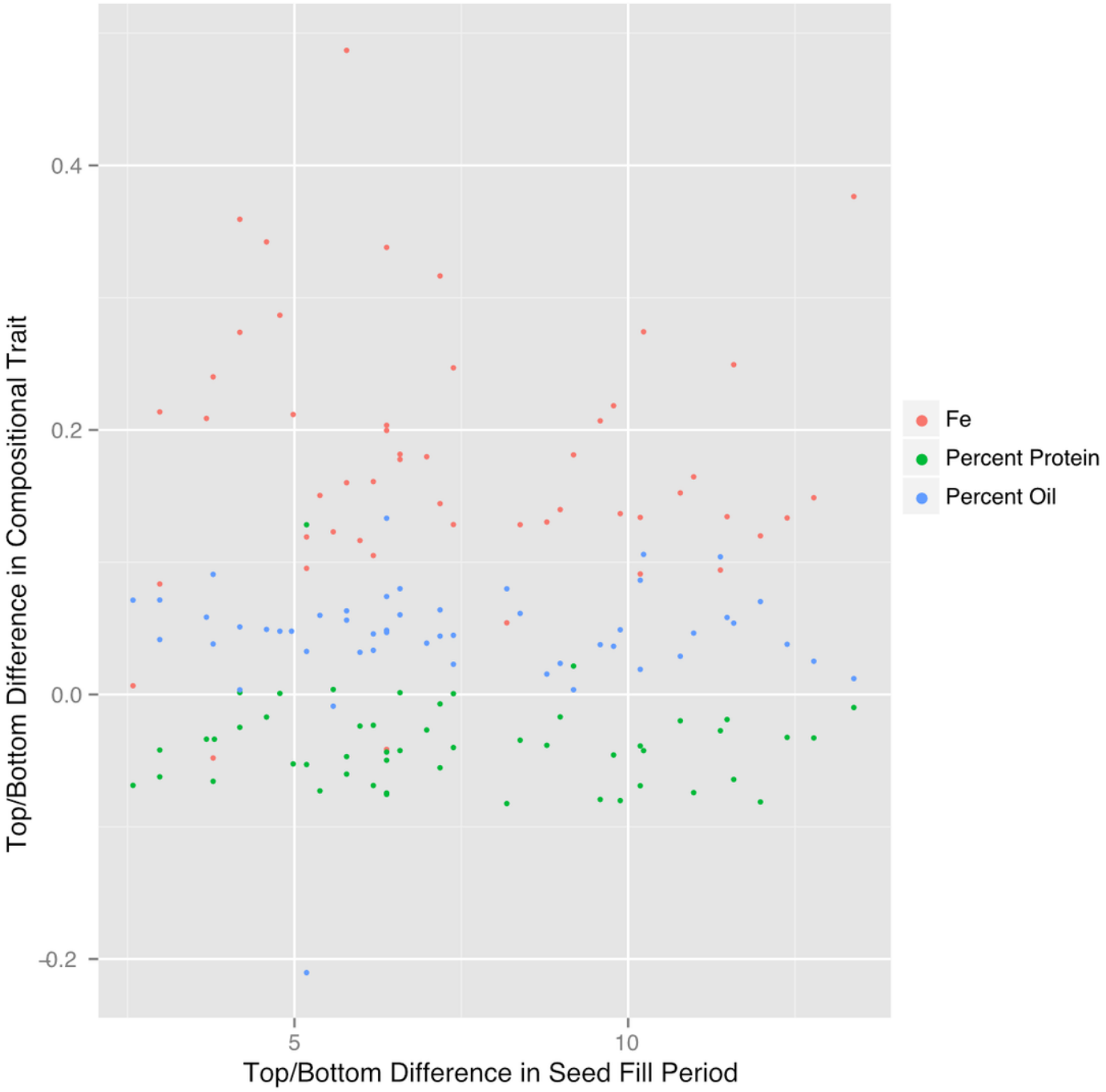


Figure 7

Canopy differences in Iron are reflected in food products.

Fe content of the products from 3 replicates of seven lines (flour) and four lines (Milk and Okara). Boxplots display the five number summary (median, 25, and 75% percentile define the box, with whiskers extending to 1.5 x interquartile range).

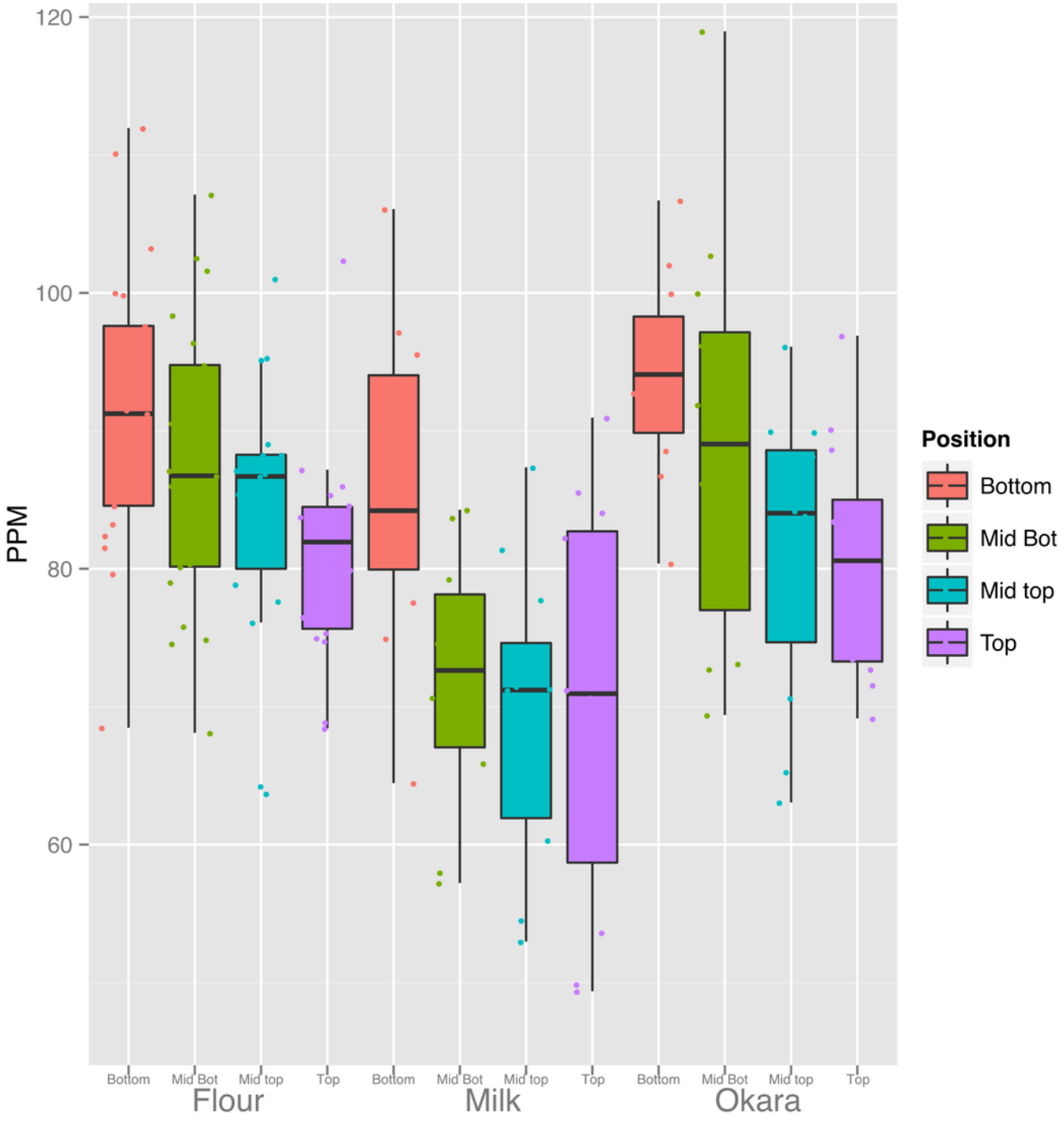

120
100
80
60
PPM
Bottom
Mid Bot
Mid top
Top
Flour
Milk
Okara
Position
Bottom
Mid Bot
Mid top
Top

Figure 8

Canopy Gradients of Leaf Composition Traits

For each trait, the data was normalized to the plot average to remove the effect of environment and genotype. The plots display the quadrant average as a line with the 95% confidence interval calculated using standard error as the ribbon. Elements with a significant ($p < 1 e -10$) effect of gradient in an ANOVA analysis that included Entry, Collection Date and Position

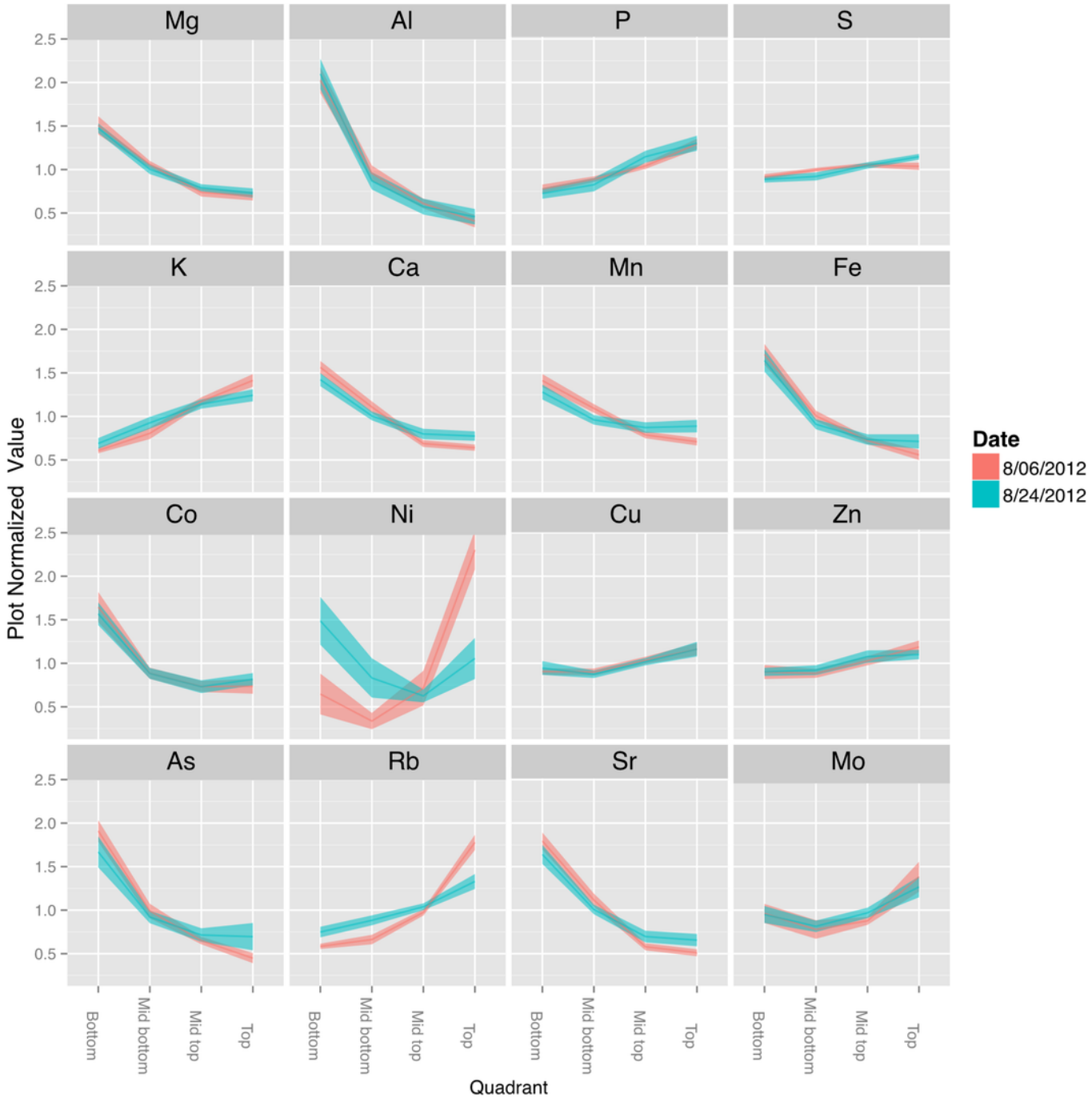

Mg
Al
P
S
K
Ca
Mn
Fe
Co
Ni
Cu
Zn
As
Rb
Sr
Mo
Plot Normalized Value
2.5
2.0
1.5
1.0
0.5
Bottom
Mid bottom
Mid top
Top
Quadrant
Date
8/06/2012
8/24/2012

Figure 9

Concentrations of selected primary metabolites in developing seeds of cultivar 'Williams 82'

A, Suc; **B**, citrate; and **C**, Asn. Boxplots display the five number summary (median, 25, and 75% percentile define the box, with whiskers extending to 1.5 x interquartile range) for three replicates at each sampling time: 7AM (7) , 12N (12) , 7PM (19) and the following morning at 7AM (31). The black vertical bars represent the intervening night period. Values are µg (g DW).

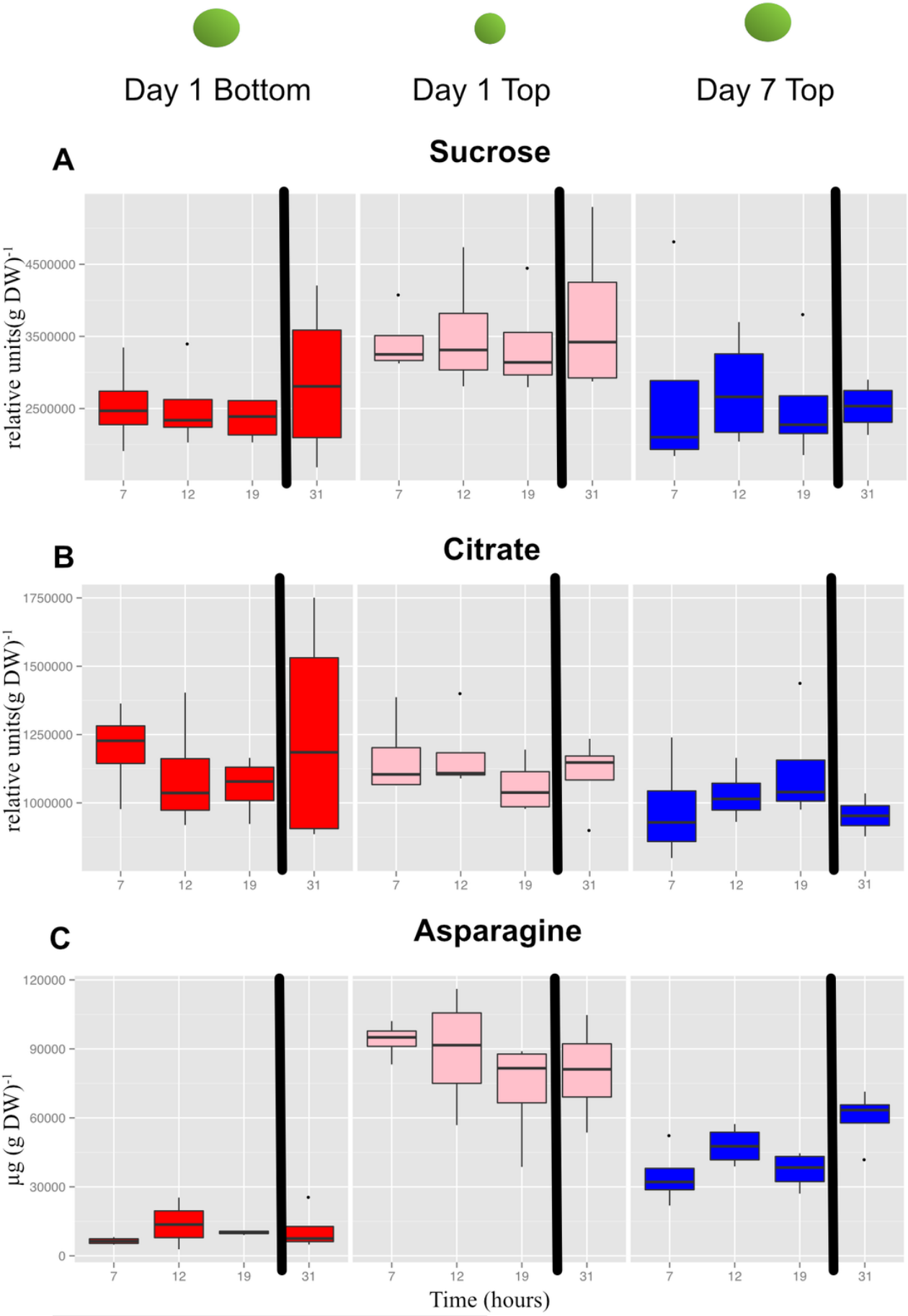
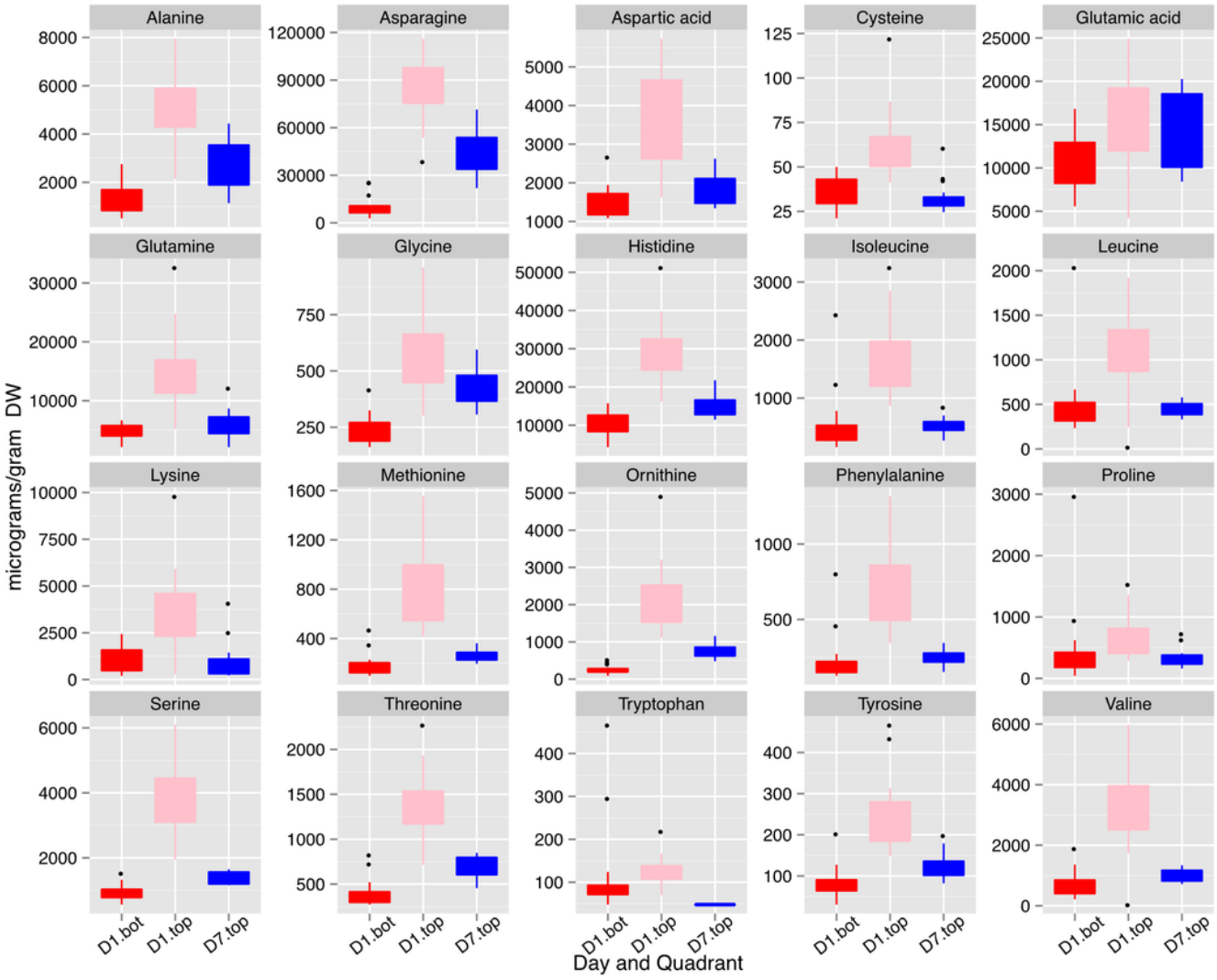

Day 1 Bottom
Day 1 Top
Day 7 Top
A
Sucrose
relative units(g DW)-1
4500000
3500000
2500000
B
Citrate
relative units(g DW)-1
1750000
1500000
1250000
1000000
C
Asparagine
µg (g DW)-1
120000
90000
60000
30000
0
7
12
19
31
Time (hours)

Figure 10

Concentrations of free amino acids in developing seeds

Boxplots display the five number summary (median, 25, and 75% percentile define the box, with whiskers extending to 1.5 x interquartile range) for values from each sampling interval (3 replicates and 4 time points are merged within each box) and nodal position. Ornithine levels reflect both ornithine and arginine as arginine is converted to ornithine during sample prep for GC-MS. D1.bot, D1.top and D7.top refer to the samples collected on day one top and bottom quadrants and the day seven top quadrant respectively.