Kansas Soybean Commission (FINAL ANNUAL Report 2017/2018)

Principle investigators: Krishna Jagadish, Raju Bheemanahalli and William Schapaugh

Project Title: "High-throughput Platform to Enhance Quality of Beans and Add Value to Kansas Soybean Breeding Program"

Checkoff Organization: Kansas Soybean Commission

Categories: Breeding & genetics

NCSRP, USB, QSSB Project Code: 1778

Amount of funding: $35,000.00

Project Year: 2017/2018

Department Head: Michel Ransom (interim)

Project Summary

Soybean breeding programs are successful in increasing yield potential, but progress in breeding for optimizing seed quality composition such as protein, oil, fatty acids, amino acids has not received similar attention. This is primarily because of the lack of a rapid screening tool to capture the complex tradeoffs between yield and quality parameters. Hence, there is a real need to develop a simple, robust and high-throughput platform for quantifying quality parameters and the seed compositional changes in response to a range of environmental conditions. The major quality determining components, i.e. amino acids, protein, oil, and fatty acids compositions are often negatively impacted by harsh environmental conditions (high temperatures and low or erratic rainfall) during pod filling stage. Enhancing the quality of the beans is emerging as a major priority that needs to be addressed, to obtain an edge in the domestic and international market. Till date, assessing the diversity in protein, oil, fatty acids, amino acids, and other quality composition in soybean grown in Kansas has not been systematically attempted. Thus, with the establishment of Near-infrared spectroscopy (NIRS) - a rapid and high-throughput tool in the Agronomy Department will strengthen Kansas soybean breeding program to develop beans with higher quality. Progress achieved in the project will increase additional income generation potential among Kansas growers.

Project Objectives

1) Develop and standardize a high-throughput approach to quantify genetic diversity in beans protein, amino acids, oil, oleic acid etc. from germplasm generated by the Kansas Soybean Breeding Program
2) Estimate the spatial and temporal impact of Kansas climatic variability on soybean quality with emphasis on essential amino acids, oleic acid etc.
3) Integrate the technology into Kansas soybean breeding program to enhance breeding efficiency towards developing high-quality beans

Project Deliverables

1) Genetic variability in protein, amino acids, oil, oleic acid compositions of native, exotic and advanced soybeans breeding lines profiled

2) Location-specific climatic impact on soybean seed quality quantified

3) A high throughput platform to determine trade-offs between yield and quality parameters established

4) NIRS spectral curves developed and standardized for supporting the soybean breeding program during and beyond the time-frame of the proposed project

Genetic resources and Methodologies:

***Plant materials*:** A diverse set of exotic soybeans (Soybean Association Panel; 249 accessions) were grown at Manhattan in 2016 (by Prof. Schapaugh and team), was used to determine the reliability and robustness of the NIRS in estimating beans (whole and ground seeds) quality composition. Further, a representative subset of Soybean Association Panel (SAP) and two lines with high oleic acids were used to validate the existing NIRS calibration curves for protein, moisture, oil, and fatty acids. On the other hand, we have also scanned soybean entries from private seed companies, certified growers, and agricultural experiment stations grown in different ecological regions (Cherokee, Manhattan, Ottawa, and Parsons) or multi-location production trials in Kansas. For validation of sugars, two soybean seed sets (cultivated in 2015 and 2016) were obtained from two experiments (Bradford Research and Education Center (BREC_2015) and Delta Research Center (DC_2016) conducted at the University of Missouri. Lastly, an experiment was conducted to develop new calibration curves for moisture and fatty acids.

NIRS scanning procedure: Seeds of each entry and replicate were profiled for composition traits such as protein, oil, fatty acids (oleic acid, linoleic acid, and linolenic) and key amino acids (lysine, methionine, threonine, and tryptophan) using NIRS (DA 7250 NIR analyzer, Perten Instruments), along with moisture content on percent dry basis. Scanning of soybean was done on whole and ground beans. Seeds were ground into a fine powder using a SPEX mixer/mill (SPEX Industries, Inc. Metuchen, NJ, USA). NIR spectra of whole seeds were collected using a small black rotating cup (volume of 125 mL) at room temperature. Ground seeds were analyzed using the small plastic cup (25 mL). Each sample was scanned for 6 seconds (15 spectra/sec) with a wavelength ranging from 950 to 1650 nm (optical resolution ~7 nm). Each genotype was scanned four times (2 repacks and 2 repeats).

Laboratory-based analysis to validate Perten calibration models

- Protein: Dry combustion (LECO C/N analyzer) method was followed to determine % dry basis protein using the ground samples that were dried at 130 °C for 1h.
- Oil: Extraction of oil with petroleum ether (Soxhlet apparatus)
- Moisture: Drying of whole soybean seeds at 103°C for 72 h. Some whole and ground seeds were moistened/dried manually at different levels to obtain samples with a wide range of moisture content.
- Fatty acids (FA): The GC-MS was tested for its ability to provide reliable results on FA analysis. Initially, a standard fatty acid methyl ester mix with 37 component mix

(Supelco® 37 Component FAME Mix) as tested on a DB-5MS column (30 m in length and 0.25 µm in diameter). For the quantification of fatty acids, we needed a column capable of separating soybean FAs (linoleic acid, linolenic acid, oleic acid, stearic acid, and palmitic acid). DB-5MS column showed a co-elution of oleic, linolenic, and linoleic fatty acids at 33.92 min, confirming its unsuitability for quantification of soybean FAs.

- The experiment was carried out using HP-INNOWax column (30 m in length and 0.25 µm in diameter) which showed a better resolution of the FAME 37 component mix. The extraction of fatty acids from soybean seeds was carried out using the method described in Obour et al. (2017) and the quantification was done using HP-INNOWax column.
- Sugars: Sugars were determined using HPLC at the University of Missouri (2015 and 2016).

Development of new calibration models

For the validation experiment, dark-colored seeds were not used as NIRS predicted data indicated high M-distances for colored seeds as those samples are not well represented in Perten calibration models. This proved the necessity of developing a new calibration model including colored seeds (black and brown). Calibration models were constructed for moisture and fatty acids using the partial least squares (PLS) regression technique of the GRAMS/AI 8.0 software.

Results

Objective 1: Develop and standardize a high-throughput technique for estimating beans quality composition

Genetic diversity in beans quality composition using high-throughput method: The exotic accessions (n=133, excluding black and brown coated beans) showed a wide range of diversity in soybean seed composition (Fig. 1). Protein content in soybeans ranged from 37.8 to 46.4% (Fig. 1A), oil content ranged from 18.1 to 23.6 % (Fig. 1B), with a narrow range of moisture (8.8 to 9.6%; Fig. 1C). Further, there were large genetic variability in fatty acids (oleic acid, linoleic acid, and linolenic acid) and a few major amino acids (lysine, methionine and threonine) [Fig. 1D-I]. Oleic acid (% dry basis) varied from 16.6 to 29.1%; linoleic acid ranged from 47.3 to 57.7%, linolenic acid ranged from 5.5 to 9.9% on a dry basis (Fig.1D-F). Similarly, we investigated some of the major soybean amino acids such as lysine (2.51 to 2.96 %), methionine (0.51 to 0.61%) and threonine (1.44 to 1.7%), which displayed relatively narrow variation but sufficient quantitative distribution (Fig. 1G-I). There was a strong negative correlation between protein and oil ($r = -0.72$, $p < 0.001$), and a weak positive correlation between protein and moisture ($r = 0.51$, $p<0.05$; Fig. 1J). A strong negative correlation between oleic acid and both linoleic acid ($r= -0.88$, $p<0.001$) and linolenic acid ($r=-0.80$, $p<0.001$) was recorded. All other relationships between specific to fatty acids and amino acids are presented in Figure 1J.

Reproducibility and reliability of NIRS output: The reproducibility of NIRS scanning was tested with two independent runs of the same set of 133 SAP accessions [oil ($r=0.97$, $p<0.001$), protein ($r=0.98$, $p<0.001$), and oleic acid ($r=0.89$, $p<0.001$)]. Most of the key parameters were significantly ($p<0.001$) consistent, indicating the accuracy of the high throughput instrument in detecting quality compositions in soybeans (Fig. 2). To understand the reliability of NIRS in detecting the seed compositional traits, a correlation matrix between sample types (Whole-W and Ground-G) was

developed along with yield (Fig. 3). The relationships between whole and ground beans for key composition traits were significantly correlated (Fig. 3). A positive strong correlation for protein ($r=0.95$, $p<0.001$) and oil ($r = 0.93$, $p< 0.001$) were noticed for whole and ground beans. These results indicated the accuracy and reliability of the high-throughput approach in capturing the genetic variability of composition traits (Fig. 3).

Validation of calibration curves (NIRS vs laboratory-based analysis): Based on the level of diversity (n=133) observed for the protein, oil, and fatty acids, a continuum of samples capturing the diversity were selected (49 for protein, 24 for oil and 38 for moisture) and validated using relevant laboratory-based methods for protein, oil, moisture, and fatty acids (Figs. 4 and 5). NIRS predicted protein content and laboratory-estimated protein content showed a significantly strong relationship (Fig. 4A) with each other for both whole (R^2=0.93, n=49) and ground beans (R^2=0.98; n=49). A strong association for oil was noticed between NIRS predicted and laboratory-estimated (Fig. 4B). Moisture content showed a strong accuracy between the NIRS predicted and laboratory-estimated moisture content (R^2=0.95; n=38; Fig 4C). Further, NIRS predicted and laboratory-estimated oleic, linoleic and linolenic contents, are shown in Fig. 5. The association values (R^2) for oleic and linoleic content were 0.91, and 0.98 using the ground samples, indicating the high degree of fitting in regression of the laboratory and NIRS predicted values (shown in Fig. 5 A, C). Whereas such relationship was not noticed for linolenic acid, palmitic, and stearic acid contents either with the ground or whole beans, an example of linolenic acid presented in Fig. 5E. For sugars (sucrose, raffinose, and stachyose), HPLC and NIRS predicted data were not in agreement with each other (data not shown) due to prolonged storage time (HPLC analysis was done in 2015/2016, and the NIRS scanning was done in 2018).

New calibration models for moisture and fatty acids (including colored seeds): Moisture calibrations including whole and ground beans indicated good statistics with R^2 >0.98 and SECV > 0.72%. New calibration models for fatty acids did not indicate good statistics (Table 1). This could reflect a high error in the standard method of the constituent analysis and should be confirmed with further research. Standard method should be able to estimate the fatty acid contents within 0.34, 0.31, 0.90, 1.65, and 0.22% for the palmitic, stearic, oleic, linoleic and linolenic acids, respectively, because the error of the reference method of fundamental analysis should be less than 5% of the constituent range.

Objective 2: The spatial impact of Kansas climatic variables (temperature) on soybean quality

Impact of climatic variables: Soybean entries grown in a broader range of climatic conditions were obtained from Kansas performance trials (from Prof. Schapaugh's and Jane Lingenfelser team), and the impact of climate variables were tested. The daily average maximum temperature (T_{Max}; Fig. 6A) during the pod-filling stage (August to October) ranged from 27.1 (Cherokee) to 28.3 ^{o}C (Parsons). Further, large variation in pod yield (bushel/acre; Fig. 6B), protein (Fig. 6C), and oil (Fig. 6D) were noticed, indicating the influence of growing environment across locations on yield and quality. Soybean grown in Cherokee had the lowest pod yield (44.1 bushel/acre, an average of 27 entries), but the protein level was highest (40.6%) compared to other locations (Fig. 6C). While, soybean entries grown in Ottawa regions recorded the highest pod yield (75.2 bushel/acre, averaged across 80 entries), while the protein level decreased by 2.1% compared to Cherokee. Soybeans collected from Manhattan region contained 2.85% low protein compared to Cherokee. To understand the *tradeoff* and/or relationships between seed compositions and yield in

soybean, the correlation matrix is developed using all the entries grown across all the four locations (Fig. 6E). A negative relationship between oil and both yield ($r=-0.34$, $p<0.05$) and protein ($r = -0.42$, $p< 0.05$) were noticed, suggesting that the soybean entries grown in these regions contained a higher level of oil by virtue of lower pod yield (bushels/acre). Meanwhile, the pod yield was significantly and positively correlated with an amino acid such as tryptophan ($r=0.67$, $p<0.001$). As expected, protein showed significant positive correlations with most of the amino acids including lysine and methionine. Averaged daily maximum temperature showed a positive association with protein ($R^2=0.415$) and a weak negative response with pod yield ($R^2=-0.875$). Yield data obtained from Prof. Schapaugh and his team on Soybean Association Panel was used to understand the tradeoff between yield and composition traits (particularly protein and oil). Soybean seed yield (Fig. 7A) had a negative relationship with protein content (*$r=-0.69$, $P<0.05$*) and a positive correlation with oil content (*$r= 0.61$, $P<0.05$*) (Fig. 7B).

Estimating spatial impact of day-time temperature on beans composition:

Year 1 ***Soybean Varietal Performance Test (2016)*****:** Seed samples of eight popular soybean cultivars grown in a wide range of climatic conditions (Cherokee, Manhattan, Ottawa, and Parsons) were obtained from the Kansas Soybean Performance Trials for estimating spatial impact day-time temperature on composition. During seed-fill (in August 2016), these regions (Cherokee-32.2^{o}C, Manhattan-30^{o}C, Ottawa-30^{o}C, and Parsons-31.6^{o}C) often experienced maximum day temperatures greater than 30^{o}C (Fig. 6A). The average growing maximum day temperature in Ottawa (in August 2016) was 1.8^{o}C lower than in Cherokee (Fig 6A). The soybean genotypes grown in Ottawa recorded the highest seed yield (76.7 bushels/acre, an average of 8 different entries), but had the lowest protein. The same soybean genotypes grown in Cherokee had the lowest seed yield (43.8 bushels/acre, an average of 8 different entries, and oil (20.8 % dry basis), but had 3% higher protein than in Ottawa (Fig. 8). A large variation in seed yield (bushel/acre), protein, and oil were noticed in response to variations in climatic conditions at each location during the pod-fill. The observed trade-off between yield and protein among 8 popular soybean cultivars across locations (Fig. 8) indicated a possible negative influence of abiotic stress (especially temperature).

Year 2 ***Soybean Varietal Performance Test (2017):*** Seeds of nine common soybean genotypes (AG 3432, AG 4232, KS3618Ngr, KS4117Ns, MG 3.5, MG 3.9, MG 4.2, MG 4.5, and S14-9051R) from six different locations (Ottawa, Colby, Rossville_dry, Rossville_Irri, Parsons, and Onaga) were analyzed to determine the effect of climatic variables on the quality composition. During the pod-fill stage (August through early September) daily average maximum temperature for Parsons, Rossville, and Colby were reported as 28.2 °C, 28.1 °C, and 30.1 °C, respectively. However, the number of days that exceeded a critical level of 29.4°C was highest in Colby (27 days) followed by Rossville (15 days) and Parsons (12 days). Seed yield had a significant positive correlation to protein (Fig. 9A; $r=0.29$; $p<0.001$) and negative correlation to oil (Fig. 9B; $r=0.39$; $p<0.001$). Seed protein and oil showed a significant negative correlation (Fig. 9C; $r=0.72$; $p<0.001$). Seed yield (Fig. 10A), protein (Fig. 10B), and oil (Fig. 10C) showed larger variations across location, indicating the influence of the environment during grain filling. Lowest yield (Fig. 10A) was observed in Parsons (43.1 bushels acre^{-1}) and highest was found in Rossville_Dry (87.1 bushels acre^{-1}). Soybean grown in Rossville_dry had a significantly high protein (Fig. 10B; 41.2% dry basis) compared to Onega (39.1 % dry basis). Soybeans that were grown in a hotter climate

(at Colby) had the lowest oil (Fig. 10C) compared to all other locations, which indicate need for breeding soybean for heat stress tolerance (high day-time temperature) during pod-filling.

Objective 3: Integrate the technology into Kansas soybean breeding program to enhance breeding efficiency

We aimed at quantifying the oleic acid in a large, diverse population and the parental lines used in the KSU soybean breeding program for increasing oleic acid (Fig. 11). Interestingly, our existing calibration curves were able to capture significantly high oleic acid levels in the two-known high oleic parental lines, compared to the other 133 diverse germplasm (Fig. 11). The two parental lines on average had 66% more oleic acid compared to the diverse panel (Fig. 11). Soybeans seeds produced from the KSU breeding experiments were analyzed to support Prof. Schapaugh and the team to select high-quality beans for the subsequent season. The details on the experiments and different genetic resources scanned through the NIRS platform totaling to more than 4000 samples are listed in Table 2. The objective of the breeding experiment is to develop beans that produce 47.5%-48.5% protein (at 13% moisture) with the possible highest percentage of oil. Prof. Schapaugh and the team continue using the NIRS platform to ensure that the selected advanced breeding lines produce the required minimum soybean meal protein (48%) together with high oil content (Fig. 12). Figure 12 shows that some of the selected breeding lines using the NIRS showed a very high protein (>48.5 % soybean meal protein) along with high oil. Moreover, it guarantee that some of the high protein selections meet the criteria for special purposes and food-use applications.

Knowledge dissemination:

Raju BR, Shetty NJ, Lingenfelser JE, Schapaugh W, Jagadish SVK. 2017. Soybean yield and quality trade-offs. Soybean Breeders Workshop: Physiology & Agronomy, February 13-15, St. Louis, MO, USA (poster presented).

Final Project Results

NIRS is a promising high-throughput phenotyping tool help to capture composition rapidly, accurately and at low cost. The high-throughput phenotyping platform has now been routinely using in the KSU breeding programs (Prof. Schapaugh's and Jane Lingenfelser team) to phenotype the segregating mapping populations and to understand the effect of microclimate on quality (protein, oil and oleic acid) while advancing location specific soybeans for growers. Recent emphasis on increasing the oleic acid in the breeding pipelines to target specialty markets has been the rationale at the global level. Our results showed that integrating the high throughput technology would enable quick identification and development of oleic acid beans through breeding programs. On the other hand, the output of the project is currently helping soybean breeding program to devise a target trait development (high yielding beans with quality). New calibration models should be included with black and brown seeds.

Benefit to Soybean Farmers

Enhance quality of soybeans grown to provide Kansas soybean growers an edge in domestic and international market and provide newer opportunities to enhance their revenue. High yielding beans coupled with quality expected to enhance the market value or opportunities to increase the income of the soybean growers.

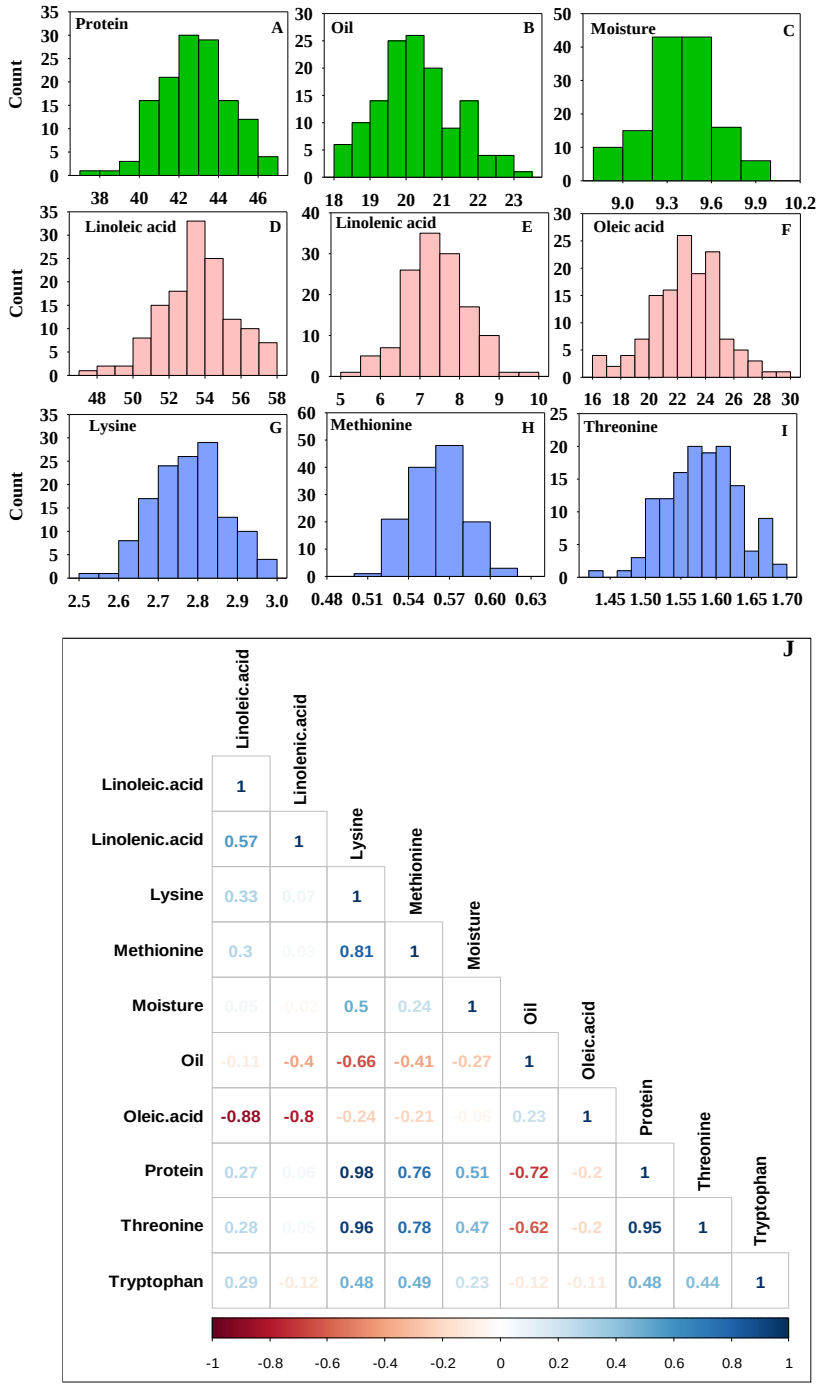

Figure 1: Histogram of seed composition traits variability and their relationships in soybean germplasm collection (*n*=133) grown in Manhattan during 2016.

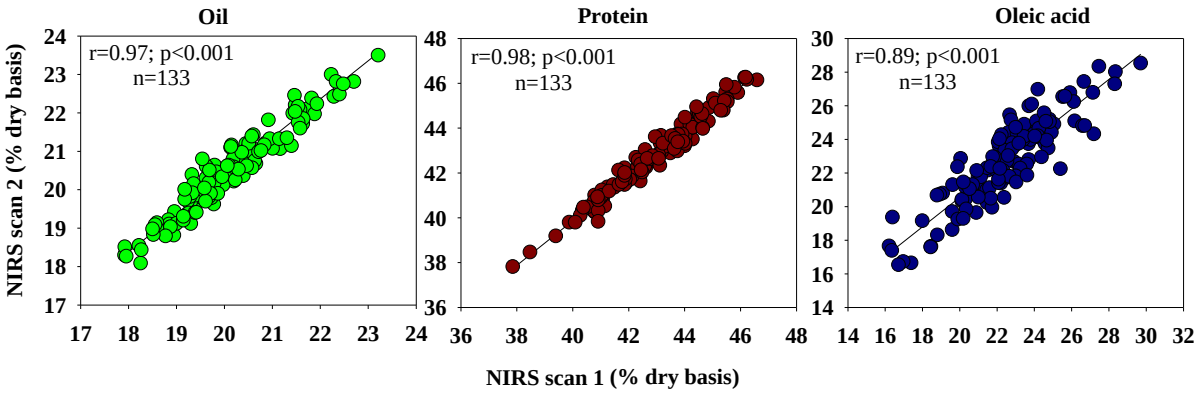

Figure 2: Reliability of the high-throughput platform (NIRS scanning) between scans (runs) on detecting soybean seed compositions (oil, protein and oleic acid) in a diverse collection (*n*=133) grown in the Manhattan during 2016.

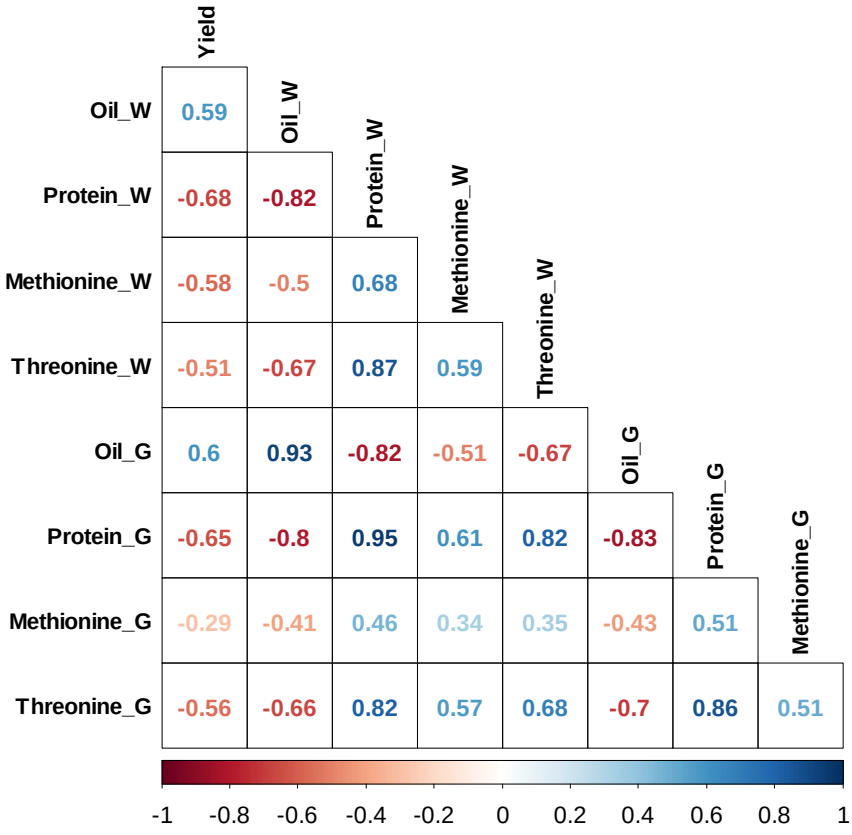

Figure 3: Reliability of the high-throughput platform (NIRS scanning) between whole beans (W) and ground (G- fine powder) beans on detecting composition traits in a diverse collection (*n*=133) grown in Manhattan in 2016.

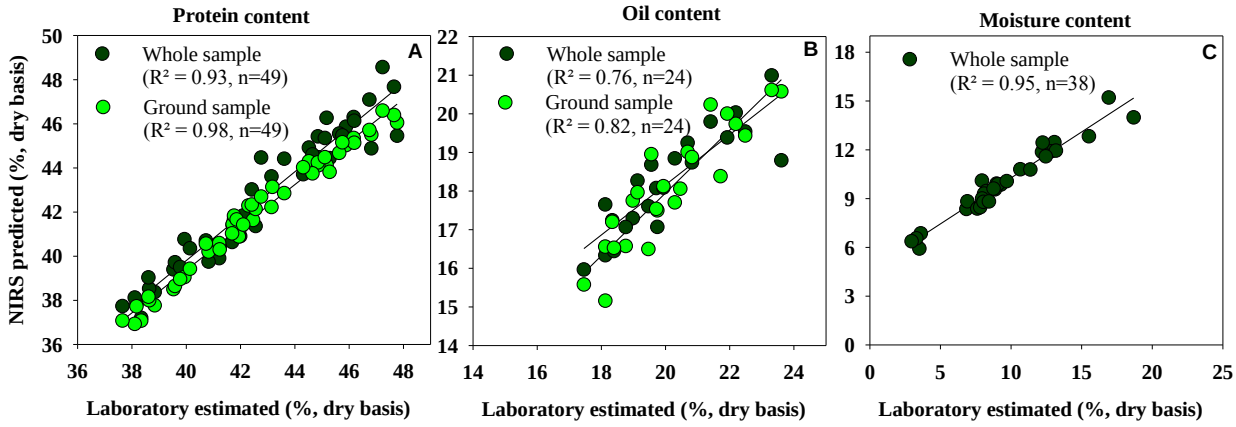

Figure 4: Association between NIRS predicted and laboratory estimated protein content (A), oil content (B) and moisture content (C) in % dry basis.

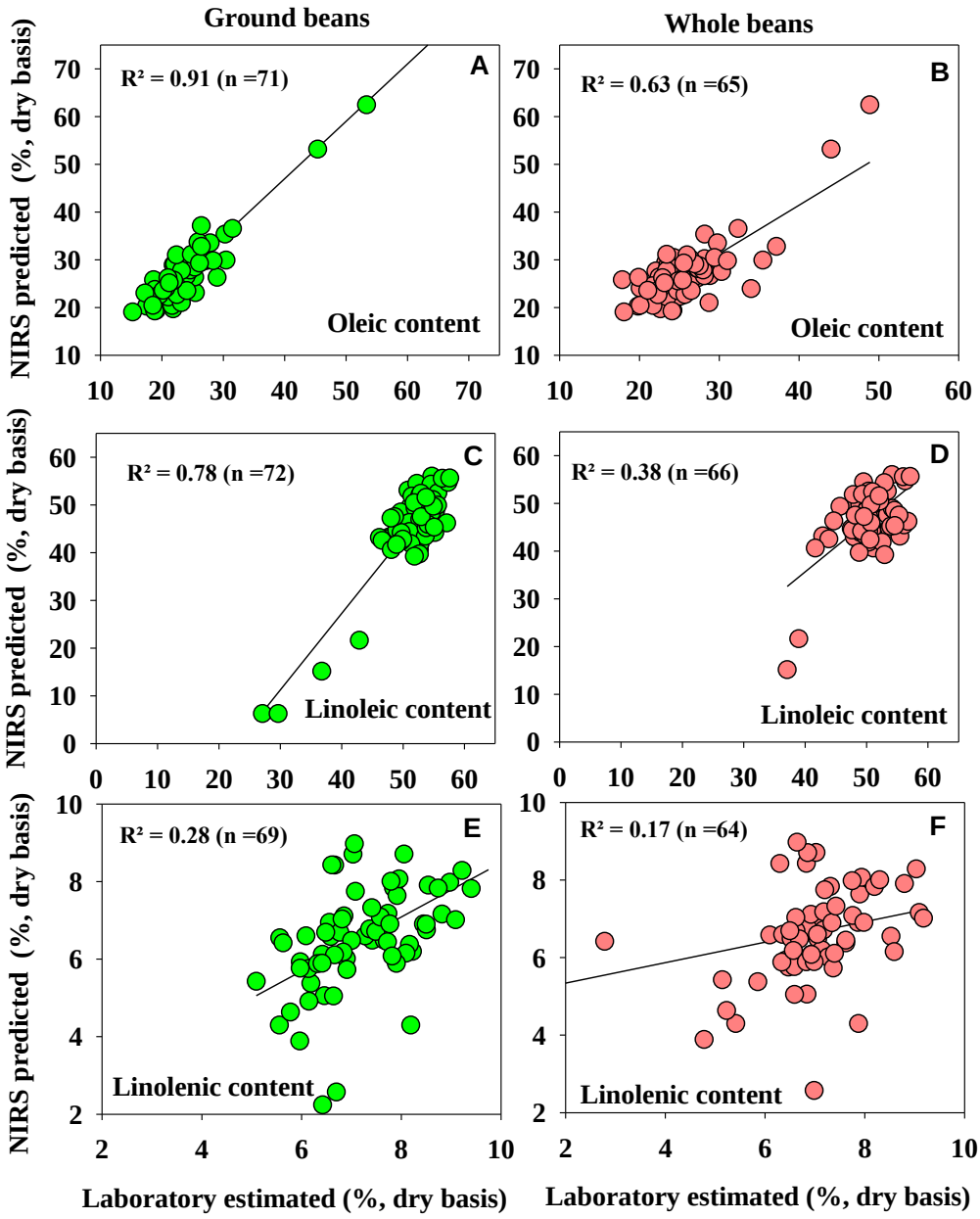

Figure 5: Association between NIRS predicted and laboratory estimated fatty acids acid (oleic, linoleic and linolenic) in ground (green, A, C and E) and whole seeds (orange, B, D and F).

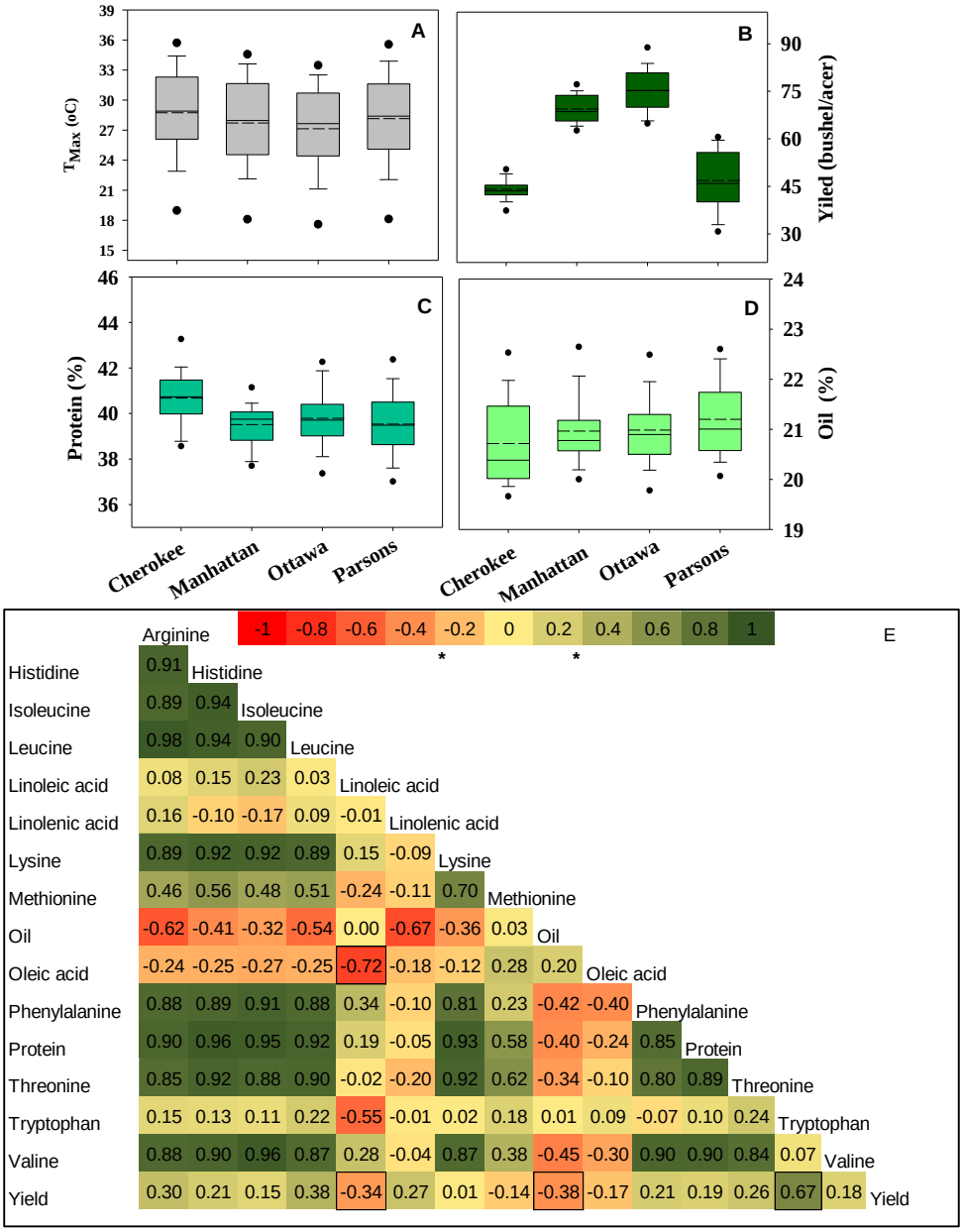

Figure 6. Variations in maximum temperature (T_{Max}, ^{o}C; A), yield (bushel/acre; B), protein (%; C), and oil (%; D), across Kansas soybean performance trials (*n*=215) and a correlation matrix between seed compositions and yield among soybean entries (E).

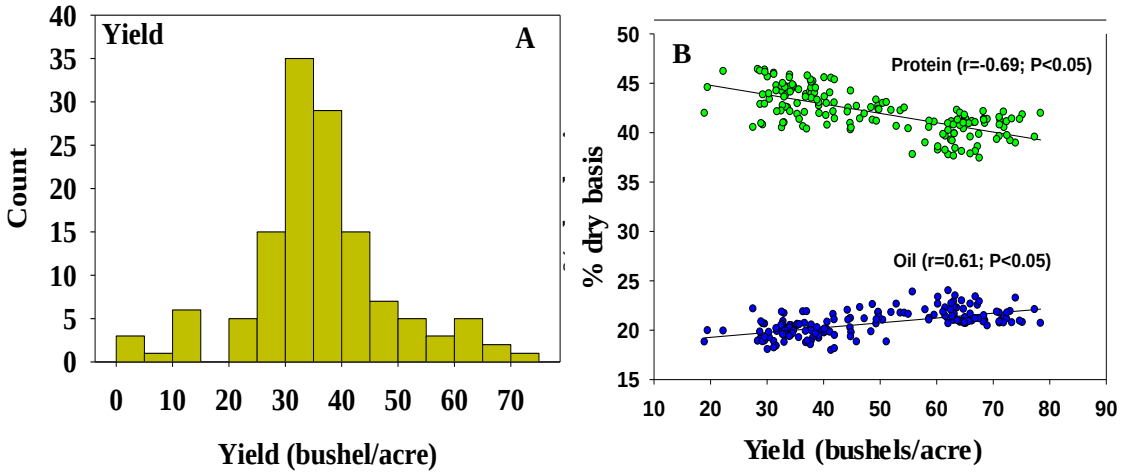

Figure 7. Variation in soybean yield (A) and its relationship with protein and oil (B) under non-stress condition. Yield data is based only on non-shattered entries.

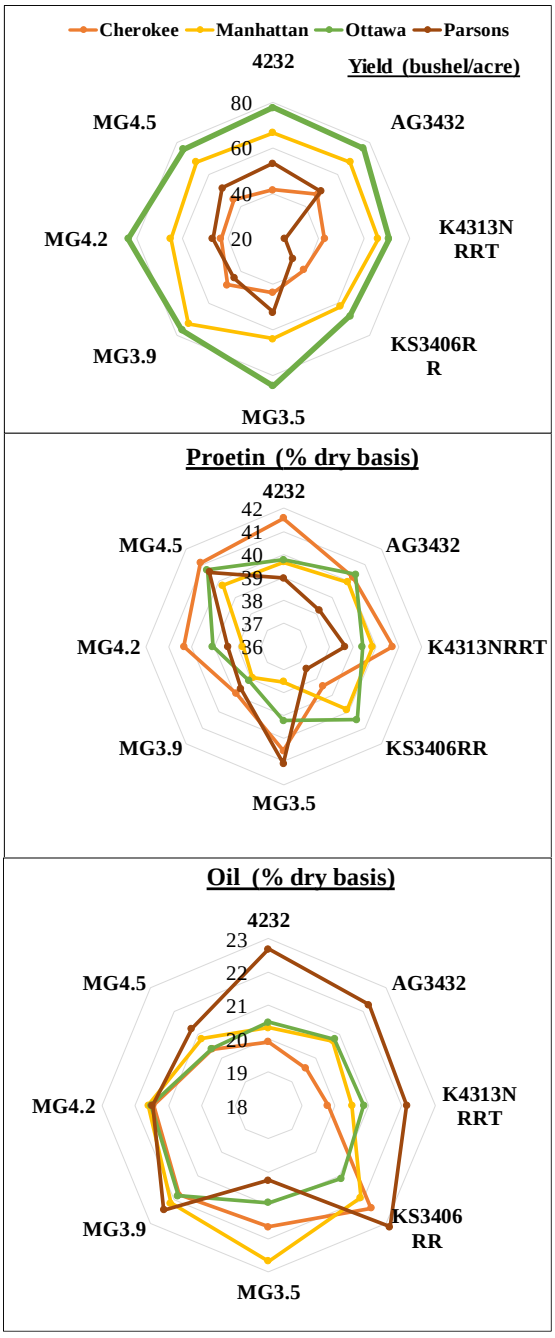

Figure 8. Soybean yield (bushel/acre) protein and oil trade-offs among popular varieties grown across four different locations in Kansas (2016). Seed samples of eight popular soybean genotypes grown in a wide range of climatic conditions (Cherokee, Manhattan, Ottawa, and Parsons) were obtained from the Kansas Soybean Performance Trials (from Prof. Schapaugh's and Jane Lingenfelser team) for estimating spatial impact day-time temperature on composition.

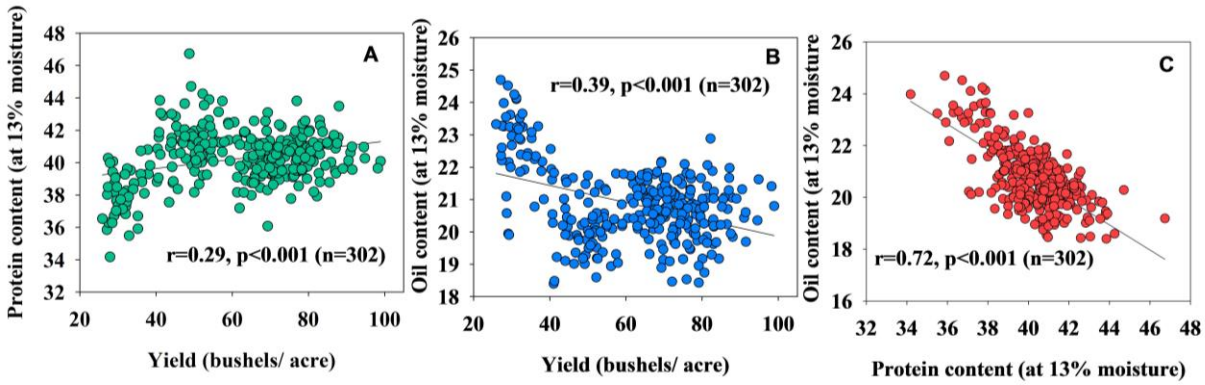

Figure 9: Correlations between yield and protein (A), yield and oil (B), and protein and oil (C) in seeds from six different locations grown in Kansas (2017). Seed samples of nine popular soybean genotypes grown in a wide range of climatic conditions (Ottawa, Colby, Rossville_dry, Rossville_Irri, Parsons, and Onaga) were obtained from the Kansas Soybean Performance Trials (from Prof. Schapaugh's and Jane Lingenfelser team).

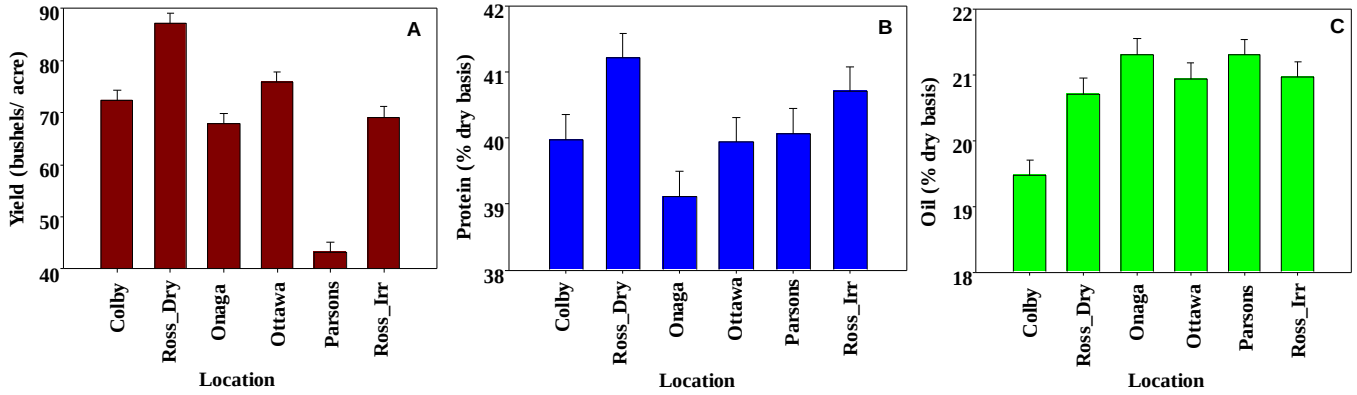

Figure 10: Significant effect of location on yield (A), protein (B) and oil (C) of nine common genotypes grown in six different locations in Kansas (2017).

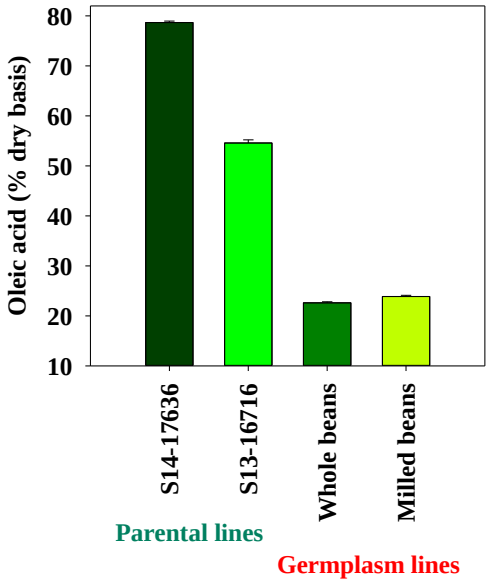

Figure 11. Comparison of high oleic acid parental line with 133 diverse germplasm accessions. Bars indicate ±SE.

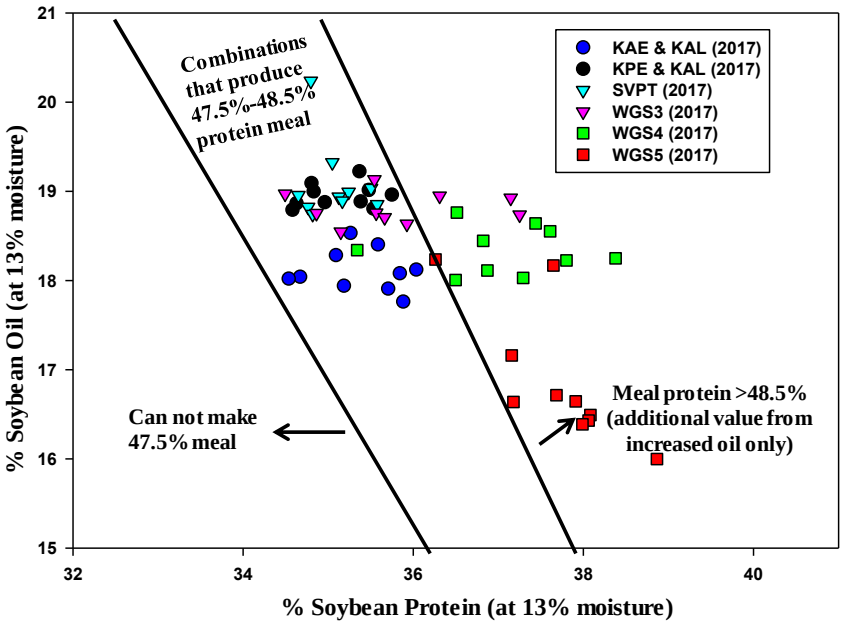

Figure 12. Genotypes capable of producing high soybean meal protein (47.5-48.5%). Black lines indicate the margins of combinations of soybean protein and oil that yield high soybean meal protein (47.5-48.5%). Adapted from Brumm and Hurburgh (2006).

Table 1. Summary statistics for new calibration models for whole and ground beans.

	Palmitic	Stearic	Oleic	Linoleic	Linolenic
Whole Beans					
PLS Factors	6	6	18	14	6
SECV	1.31	1.33	3.14	5.58	0.92
R^2	0.3	0.13	0.46	0.55	0.27
Ground Beans					
PLS Factors	6	0	9	12	10
SECV	1.38	-	3.43	5.38	0.88
R^2	0.23	-	0.26	0.55	0.36

Note: standard error of cross-validation

Table 2: Details of different genetic /breeding resources used in the project.

Year	Experiment	Details	No. of samples
2017	Kansas Advance test (early and late maturity groups): KAE and KAL	35 genotypes, 2 reps	70
2017	Kansas Performance Test (early and late maturity groups): KPE and KPL	130 genotypes across 6 locations	450
2017	SA_Missouri drought	74 genotypes, 3 reps,1 location (Salina)	222
2017	OT_ Missouri drought	74 entries, 3 reps, 1 location (Ottawa)	222
2017	Soybean Varietal Performance Test: SVPT	131 genotypes across 6 locations	337
2016	Whole genetic sequence: WGS3	200 entries, 2 reps, 2 locations (Manhattan and Salina)	795
2016	Whole genetic sequence: WGS4	254 entries, 2 reps, 2 locations (Manhattan and Salina)	801
2016	Whole genetic sequence: WGS5	90 entries, 2 reps, 1 location (Salina)	178
2017	Whole genetic sequence: WGS3	200 entries, 2 reps, 1 location (Salina)	400
2017	Whole genetic sequence: WGS4	254 entries, 2 reps, 1 location (Salina)	500
2017	Whole genetic sequence: WGS5	90 entries, 2 reps, 1 location (Salina)	180