Atlas Journal of Plant Biology 2020, pp. 96–105
doi: 10.5147/ajpb.v0i0.215

Evaluation of Yield Performance of Soybean Mutant FM6-847 in North Carolina

Jiazheng Yuan [1], Nacer Bellaloui[2], Naoufal Lakhssasi[3], Samer M. AbuBakr[1], Safiah Kassem[4], Zakariah Kassem[4], Sakinah Kassem[5], Courtney Barnes[1], Aeisha McLelland[1], Barbara Brown[1], Wendy Adams[1], Tarik El-Mellouki[6], Khalid Meksem[3], and My Abdelmajid Kassem[1*]

[1] Plant Genomics and Biotechnology Lab, Department of Biological and Forensic Sciences, Fayetteville State University, Fayetteville, NC, USA; [2] Crop Genetics Research Unit, USDA–ARS, 141 Experiment Station Road, P.O. Box 345, Stoneville, MS 38776, USA; [3] Department of Plant, Soil, and Agricultural Systems, Southern Illinois University, Carbondale, IL 62901, USA; [4] Athens Drive High School, 1420 Athens Dr, Raleigh, NC 27606, USA; [5] Dillard Drive Middle School, 5200 Dillard Dr, Raleigh, NC, USA; [6] Johnson County Community College, Overland Park, KS 66210, USA.

Received: December 16, 2019 / Accepted: March 30, 2020

Abstract

Soybean is a well-known crop for its protein, oil, fatty acids, minerals, isoflavones, and other bioactive compounds. The high yielding mutant FM6-847 was developed through ethyl methanesulfonate (EMS) mutagenesis. FM6-847 was derived from soybean cultivar Forrest and the yield performance of FM6-847 was not evaluated in North Carolina. The objective of this study was to evaluate the yield performance of the mutant FM6-847 compared to three USDA reference lines LD00-2817, LD06-7620, and LD07-3395 in a field trial in Fayetteville, NC over a period of two years (2016–2017). We compared plant height (PH), pod numbers (Pod#), seed number (Seed#), 100-seed weight (100-SW), and total seed weight (TSW) of the mutant and USDA reference lines. In addition, we investigated the presence of Rhizobia, nitrogen fixing bacteria that help in increasing the yield of legumes through enriching nutrients by nitrogen fixation. The results showed that the mean TSW of mutant line was significantly higher ($P<0.05$) than all of the USDA reference lines in 2017 trial. The yield parameters of PH, pod#, seed#, and 100-SW were also significantly different between the soybean mutant line with more than one USDA reference lines. The contribution of yield parameters to the TSW was also analyzed and these parameters were significantly contributing to the TSW based on linear fixed model. The first two principal components explained more than 70% phenotypic variation among variables in the dataset based on the results of principal component analysis (PCA). Finally, a total of 13 bacterial strains including nitrogen fixation bacteria *Rhizobium giardinii* were identified in the soil of the field trial.

Keywords: Soybean, mutant FM6-847, total seed weight, yield parameters, multiple comparison.

* Corresponding author: mkassem@uncfsu.edu

Atlas Journal of Plant Biology - ISSN 1949-1379. Published By Atlas Publishing, LP (www.atlas-publishing.org)

Introduction

Seeds of soybean [*Glycine max* (L.)] are rich in protein, oil, fatty acids, minerals, isoflavones, and many other bioactive compounds (Brim and Burton, 1978; Dornbos and Mullen, 1992; Meksem et al., 2001; Ojo et al., 2002; Bellaloui et al., 2008, 2010a,b). Soybean is an important cash crop in the US including NC and the world. However, nitrogen is a limiting nutrient for plants growth (Nie et al., 2011). Soil is a very complex environment with a high diversity of bacteria (Fierer and Jackson, 2006) among which some have the ability to fix atmospheric nitrogen. Most of biologically fixed nitrogen comes from the symbiotic relationship between legumes and Rhizobia species (Vance, 1998). Through this symbiotic relationship, these bacteria can fix atmospheric nitrogen and plants take it up for their growth and development (Ouma et al., 2016). The US produced 123.7 Million metric tons of soybean in 2018 which make it the world's first soybean producer (34%) followed by Brazil (32%, 117 Million metric tons), Argentina (15%, 55 Million metric tons), China (4%, 15.9 Million metric tons), India (3%, 11 Million metric tons), Paraguay (2.6%, 9.5 Million metric tons), and Canada (2%, 7.3 Million metric tons). All other countries produced about 6% or 21.6 Million metric tons (Soystats, 2019). In NC, soybeans were planted on 669 thousand hectares (ha), yielded about 1.5 Metric Tons, and generated estimated revenues of $464 Million for farmers in 2018 (Soystats, 2019).

It is well established that seed yield, yield components, and seed composition traits, including plant height, seed weight, protein, oil, fatty acids, amino acids, isoflavones, and sugars contents depend on the genotype and environmental factors such as temperature, drought conditions, presence or absence of pests, row spacing, application of nutrients, and many other factors (Luedders, 1977; Wolf et al., 1982; Maestri et al., 1998; Meksem et al., 2001; Dalley et al., 2004; Lightfoot et al., 2005; Jacobson et al., 2007; Ivey et al., 2011; Ragin et al., 2012, 2014; Akond et al., 2012, 2013a,b, 2014a,b, 2015; Bellaloui et al., 2008; 2011; 2012; 2013a,b, 2014, 2016; Khandaker et al., 2015; Ambrocio et al., 2016). The 'Essex' by 'Forrest' recombinant inbred line (RIL) population (ExF, n=100) was studied for many traits, including yield and yield components, seed isoflavone, protein, and oil contents (Meksem et al., 2001; Yuan et al., 2002; Kassem et al., 2004a,b, 2006; Lightfoot et al., 2005). The results showed that Essex cultivar had a mean maturity of 104 days and mean seed yield of 3.6 metric tone/ha while Forrest cultivar had a mean maturity of 111 days and mean seed yield of 3.3 metric tone/ha. The RIL population had a mean maturity of 108 days and mean seed yield of 3.4 metric tone/ ha (Lightfoot et al., 2005). Subsequently, Jacobson et al. (2007) found that plant height ranged from 28.1 to 93.3 cm in the ExF RILs. Other studies found that plant height is positively correlated with seed yield (Ivey et al., 2011; Ragin et al., 2012). In ExF RIL population, the RILs showed higher seed weight than the parents (Essex and Forrest) and 46% of the RILs outperformed the high yielding parent Forrest. Seed weight showed the highest variations for year of planting and genotype but it was negatively correlated with plant height (Ivey et al., 2011). In the 'PI 438489B' by 'Hamilton' RIL population (PIxH, n=50), plant height was highly correlated with 100-seed weight, total seed weight, and pod and seed numbers. Days to flowering showed a high correlation with plant height but low correlation with total seed numbers (Ragin et al., 2012). Akond et al. (2012) grew the PIxH RIL population (n=50) in two row spaces (25 cm vs. 50 cm) and their results showed high variations for all traits studied including plant height, pod and seed numbers, 100-seed weight, and total seed weight. Moreover, Plant height showed a high correlation with pod and seed numbers as well as total seed weight in both row spaces (Akond et al. 2012). In order to study the correlation among soybean quality traits and trait improvements, a new monofunctional ethylating agent called ethyl methane sulfonate (EMS) can often be used to create the EMS mutagenized soybean population and thereby select improved quality traits of soybean lines with superior quality traits. Studies suggested that the EMS mutants increase the protein content as high as 50% compared to that of control lines (Espina et al., 2018). Similarly, an increase in the quality of oil production has been indicated resulting from the use of EMS and displayed a positive correlation of oxidative stability index (OSI) and oleic acid content (Patil et al., 2007). The high yielding mutant FM6-847 was developed through EMS mutagenesis at Southern Illinois University Carbondale (SIUC) and the mutant line was originally derived from one of parental lines of ExF RILs, soybean cultivar Forrest. The yield performance and quality trait assessment of soybean EMS mutant FM6-847 were not conducted in the field of North Carolina.

To date, linear fixed-effective model and PCA are widely used statistical approaches in agriculture research to investigate the response of the phenotypic parameters to total yield (Yuan et al. 2016) and a few studies have explored the effectiveness of a number of field parameters of crop yield in various environmental conditions and the interrelationship among these field parameter were assessed in different crops (Gopal 1999; Afuape et al. 2011; Gana et al. 2013; Placide et al. 2015; Yuan et al., 2016). The objective of this study was to investigate the agronomic performance of the FM6-847 mutant compared to three USDA reference lines (LD00-2817, LD06-7620, and LD07-3395) in two field trials and over two years (2016 and 2017) in Fayetteville, NC. We compared PH, pod#, seed#, 100-SW, TSW, protein, and oil of the mutant and reference lines within two consecutive years. A soil microbial community analysis was also performed.

Materials and Methods

Plant Material

A high-yielding mutant line (FM6-847) and three USDA reference lines (LD00-2817, LD06-7620, and LD07-3395) have been used in this study. The mutant line was developed at the Illinois Agricultural Experiment Station in Southern Illinois University at Carbondale. The high yielding mutant FM6-847 was originally derived from soybean cultivar Forrest through EMS mutagenesis in Southern Illinois University at Carbondale (SIUC). LD00-2817 is a late F5 maturity group IV line

derived from a cross between 'Ina' and 'Dwight' cultivars, is highly resistant to soybean cyst nematode (SCN, Heterodera glycines races 1 and 3). It yields on average 4.6 metric ton/ha and has a lodging score of 2.6, a plant height of 43 cm, a 100-seed weight of 13.6 g, a protein content of 33.7%, and an oil content of 20% on average (Scofield et al., 2016). LD06-7620 is a maturity group IV line derived from a cross between 'IA3023' and 'LD00-3309' lines, and is not resistant to SCN race 1 but resistant to race 3. It yields on average 4.4 metric ton/ha, and has a lodging score of 2.3, a plant height of 39 cm, a 100-seed weight of 14.7 g, a protein content of 35.4%, and an oil content of 19.3% on average (Scofield et al., 2016). LD07-3395 is a late maturity group III line derived from a cross between 'Maverick' and 'Dwight' cultivars, and is highly resistant to SCN race 3 and moderately resistant to SCN race 1. It yields on average 4.6 metric ton/ha, and has a lodging score of 2.2, a plant height of 37 cm, a 100-seed weight of 16 g, a protein content of 32.4%, and an oil content of 20.6% on average (Scofield et al., 2016). These three USDA checks have exhibited better yield than that of soybean cultivar Forrest.

Development of an EMS Mutagenized 'Forrest' Population

The soybean cultivar 'Forrest' was used to develop an EMS-mutagenized population. The seeds were mutagenized with 0.6% (v/v) EMS, as described previously (Meksem et al., 2008) and planted at the Horticulture Research Center (HRC) of Southern Illinois University at Carbondale. A total of 1,588 M2 family seeds were harvested in 2011, then successively advanced to M3, M4, M5 and M6 generations in 2012, 2013, 2014 and 2015, respectively. In total, 1037 mutant families (lines) were harvested. Then, seeds were threshed at the ARC, packaged, and stored at 4°C for short time storage and at -20°C for long time storage. FM6 plants including the FM6-847 mutants were harvested from M5 plants in 2015.

Growing Conditions

Seeds of the mutant FM6-847 and USDA reference lines (LD00-2817, LD06-7620, and LD07-3395) were sown directly in a field in Fayetteville, NC (35.0527° N, 78.8784° W and 262 feet above sea level; Cumberland County) in a randomized complete block design (RCB) with four replicates and three row spaces (25 cm, 50 cm, and 75 cm) between the seeds. The plants were growing in the rain feed plots with occasionally watered and kept in the field until maturity (121 days, May–September 2016 and 2017). No pesticide, herbicide, or fertilizer was applied.

Soil Microbial Community Analysis

In order to assess the nitrogen fixation microbes in the rain feed and no fertilizer field, soil samples were collected from the field and serially diluted. Briefly, an autoclaved trowel was used to remove 5 shallow scoops of soil from different points in the field from approximately the top four inches of soil. Soil samples were mixed in an autoclaved beaker using autoclaved spoons. The homogenized soils were then serially diluted in sterile isotonic (0.85%) saline. A 100 µl from each dilution was plated on a nutrient agar plate using a glass spreader. The nutrient agar plates were incubated at room temperature (23-25°C) for 2-3 days. After incubation, different colonies were chosen from different plates followed by at least three rounds of streaking for single colonies on nutrient agar plates. The purified strains were used for further analysis.

DNA Extraction from Bacterial Strains

Genomic DNA was isolated from the purified strains using PureLink™ Genomic DNA Mini Kit (Invitrogen, CA). Briefly, each strain was streaked on a nutrient agar plate and incubated two to three days at room temperature until good colony growth was observed. A single colony from each plate was used to inoculate 5 ml nutrient broth and incubated at room temperature until turbid growth was observed. A sufficient volume of culture (1 to 2 ml) was centrifuged at 10,000 g for 2 minutes to provide a pellet size of approximately 10^9 bacteria. PureLink™ Genomic DNA Mini Kit (Invitrogen, CA), was used to extract DNA from the cells according to the manufacturer's directions. Extracted DNA was stored at −20°C. A negative control was included consisting of sterile water rather than cells to confirm absence of contamination.

16S rRNA Gene, DNA Sequencing, and Annotation

Universal bacterial primers for 16S rRNA were used to amplify a fragment of approximately 1,400 bp. Extracted DNA was used as the template in PCR using the primer pair 27F/1492R (Wilson et al.,1990). The PCR products were visualized using agarose gel documentation system (Bio-Rad, USA). The nucleotide sequences of the PCR products were determined at Eton Bioscience, Durham, NC using 27F/1492R (Wilson et al., 1990) primers for 16S rRNA gene sequence. Sequence chromatograms were assembled to generate the consensus sequences using Phred-Phrap (Machado et al., 2011). Taxonomy affiliation was deduced using BLASTn (Altschul et al., 1990).

Trait Measurements

Plant height (PH) and pod numbers (PN) were recorded in the field just before harvest and seed numbers (SN) were recorded in the lab just after harvest. 100-seeds (100-SW), and total seed weight (TSW) were recorded in the lab after harvest. The means of the agronomic traits studied here are show in Table 1.

Statistical Data Analysis

The One-way ANOVA combining R package multcomp with Tukey function was used to compare the mean of the field data of different lines within two-year interval and the year was treated as a block effect (Mangiofico, 2015). The R function

Table 1. Multiple comparisons of yield parameters in the two dataset. The R package multcomp was used to conduct the analysis. PH: plant height; Pod#: total pod number; Seed#: total seed number; 100-SW: 100 seed weight; TSW: total seed weight.

Line	Year	PH	Pod#	Seed#	100-SW	TSW	Protein	Oil
LD00-2817 - FM6-847	2016							
LD06-7620 - FM6-847	2016	.			**		***	
LD07-3395 - FM6-847	2016	**						
LD00-2817 - FM6-847	2017							
LD06-7620 - FM6-847	2017	*	.	*		.		
LD07-3395 - FM6-847	2017	*		*				
LD00-2817 - FM6-847	2016 2017							
LD06-7620 - FM6-847	2017 2017			.	***		**	
LD07-3395 - FM6-847	2018 2017	*			.			

Note: *<0.05, **<0.001, ***<0.001, · ≈0.05.

of boxplot (www.r-project.org) was used to display the mean TSW. The linear mixed-effects model (Starkweather, 2010) was implemented to handle the fixed and random effects and to determine if the independent (explanatory) variables both fixed effects and random effects, altered the dependent (response) variables. The analytic tools in R (www.r-project.org) including that of linear mixed-effects model fit by maximum likelihood, correlation, and PCA were employed in the data analysis. Linear mixed-effects fit by maximum likelihood built in R package lme4 was used to analyze the independent variables including both fixed effects and random effects (Bates et al., 2014). Except for TSW, other yield parameters including year, PH, pod#, seed#, 100-SW, and quality traits of protein, and oil were considered as independent variables. Thus, the total seed weight (TSW) was considered as the dependent variable. The soybean lines were treated as a random effect. Year (YR) as an important environmental aspect was considered to be a factor with two levels while other parameters were designed to be continuous explanatory variables. The R package car (type II Wald chi-square tests) was integrated to display the significant level of the assessment. PCA was also used to analyze the dataset as a complementary analytic tool to validate the results of the mixed model. PCA analysis was carried out following the prcomp algorithm plus ggbiplot package for the visual display (Vu and Lei 2012, www.r-project.org). The correlation and barplot were carried out using R code described by King (2013). The following abbreviated terms were used in the R script including TSW presented total seed weight; Plht denoted as plant height; Pod presented total pod number; Seed was for total seed number; SW100 was for 100 seed weight.

Results and Discussion

Yield Performance

The one-way ANOVA combing the R package multcomp with function of Tukey was employed to analyze the yield parameters between soybean mutant FM6-847 with USDA reference lines. As presented in Table 1, the total seed weight, plant height, total pod number, and total seed number were significantly different at least from one of USDA reference lines using the field data in 2017. Compared the plant height and total seed number of FM6-847 to the USDA reference line LD07-3395, the significantly mean differences were observed in the field data collected in 2017. However, only plant height and 100-seed weight between FM6-847 and USDA reference lines LD06-7620 and LD07-3395, respectively were significantly different in the field data collected in 2016. The severe drought was experienced in the growth season of the field trial in 2016 and thereby significantly affected the yield performance among lines in the rain feed trial. The box plot was clearly displayed the mean total seed weight among the lines analyzed in 2016 and 2017, respectively (Figure 1). The mean of total seed weight of the mutant was among the highest in 2017 and nearly to be the lowest one compared to the USDA reference lines assessed in 2016 at the rain feed field plot (Figure 1). The block function of ANOVA was used to compare the two year yield data and of the seven yield parameters, the plant height, 100-seed weight, and protein content were significantly different compared the mutant to the USDA reference line LD06-7620 when year was treated as the block effect. The significant differences of plant height and 100 seed weight were also identified between FM6-847 and LD07-3395 in the two year dataset combined (Table 1). In contrast to the means of other yield parameters that were significantly higher in FM6-847 than in USDA reference lines, the content of protein was higher in USDA line LD06-7620 than the mutant FM6-847.

Correlation and Contribution of Yield Parameters to TSW

The soybean mutant FM6-847 line that was analyzed in this study had wide ranges of phenotypic differences for all major components of yield parameters (Table 2). The coefficients of variation (CV) of pod number (83.43%), seed number (61.35%) and total seed weight (61.25%) were high among the lines suggested that the higher variability of these traits was identified in the assay. In contrast, the CVs of protein (5.44%) and oil (4.7%) were very narrow. The linear mixed-effects model fit by maximum likelihood was used to conduct the analyses of the dataset from field trials within two consecutive years in order to esti-

mate the contribution of the yield parameters among soybean lines toward soybean total seed weight (Table 3). For the fixed-effects of the linear mixed model, PH, seed#, 100-SW were all were significantly contributed to TSW ($P < 0.001\%$). However, no association between the total protein content with TSW was identified. The higher significant coefficients of variation were observed for PH, seed#, and 100-SW (Table 3). Based on the mixed-effects algorithm, pod# also significantly contributed to TSW ($P<0.01$) while year and oil content had marginally been considered as significant factor for TSW of soybean lines ($P\approx0.05$). The relationship among the field traits was observed in the correlation analysis. Seed# and pod# had the highest correlation (r=0.99 and r=0.98, respectively) with TSW (Figure 2). The TSW had a weak correlation with plant height (r=0.45) and 100 seeds weight (r=0.31) but had a moderate or high correlation with year (r=0.69) suggesting that the environmental factor played a significant role in the rain feed plot (Figure 2). Our results also suggest that the yield parameters contribute to TSW differently for the mutant line. Based on the algorithm of correlation, pod number and seed number were also strongly correlated. However, the relationships between oil and other parameters were excluded by the software due to the low correlation indicating that the factors that contributed to the TSW mainly from the components other than the contents of protein and oil.

Principal Component Analysis

As one of the most widely used multivariate statistical approaches in agriculture research, PCA was used to identify the major variance components among the correlated traits (Afuape et al. 2011; Gana et al., 2013; Placide et al., 2015; Yuan et al., 2016). Multivariate approach of PCA assigned a given set of variables into the principal components (PCs) and transferred the original correlated variables through linear combination to

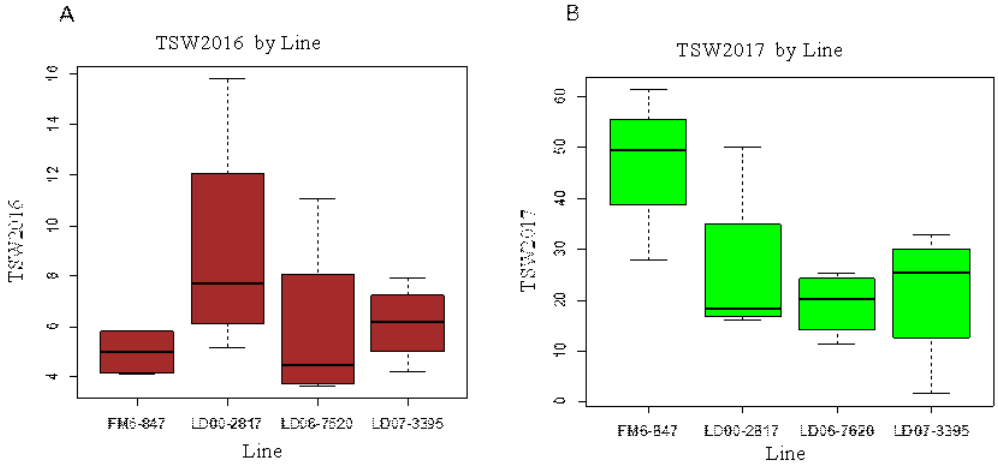

Figure 1. Box plots of total seed weight among the lines assessed. The values of the seed weight were log 2 transformed. A: Box plot of TSW of 2016 field trial; B. Box plot of TSW of 2017 field trial.

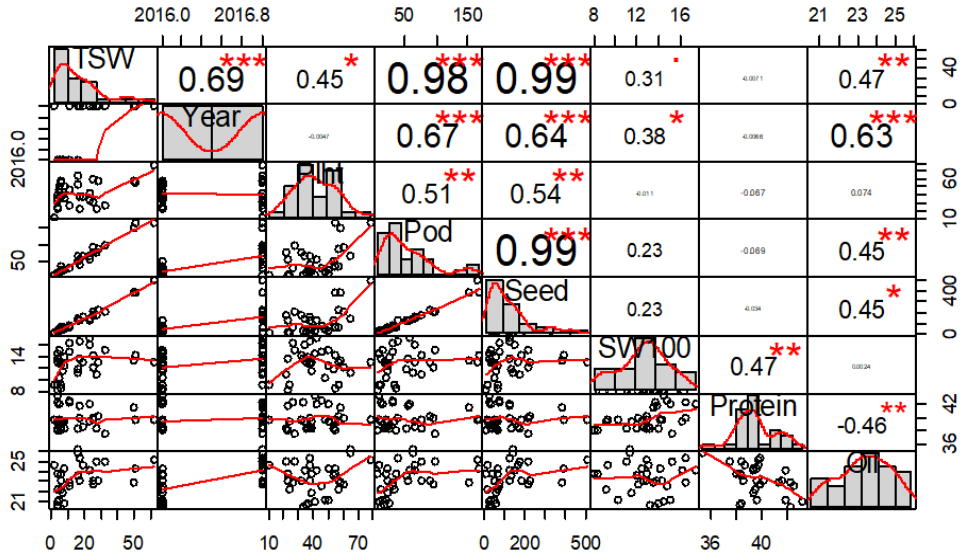

Figure 2. Correlation among yield parameters based on two-yield data collected in the field trials of Fayetteville, NC.

Table 2. Mean, range, standard error, and coefficient of variation (CV) for field parameters and protein and oil contents based on % dry weight of soybean lines assessed in North Carolina. PH: Plant height; Pod#: pod number; Seed#: seed number; 100-SW: 100 seed weight (g); TSW: total seed weight (g). Protein and oil contents are in %.

Trait	Mean	Range	CV	SE
PH	41.52	59	38.36	3.4
Pod#	130.86	454	83.43	23.28
Seed#	228.09	626	61.35	29.84
100-SW	14.48	8.08	13.72	0.42
TSW	32.82	88.56	61.25	4.29
Protein	41.43	7.6	5.44	0.48
Oil	23.35	3.6	4.7	0.23

Table 3. Deviance table of relationship of TSW with other field parameters. PH: plant height; Pod#: total pod number; Seed#: total seed number; 100-SW: 100 seed weight; TSW: total seed weight.

	Chisq	Df	Pr(>Chisq)	Significance
Year	3.735	1.000	0.053	.
PH	21.106	1.000	0.000	***
Pod#	7.597	1.000	0.006	**
Seed#	45.892	1.000	0.000	***
100-SW	14.363	1.000	0.000	***
Protein	0.864	1.000	0.353	
Oil	4.797	1.000	0.029	*

Note: *<0.05, **<0.001, ***<0.001, · ≈0.05. The linear mixed model fit by maximum likelihood was used to evaluate the two years field data. The linear fixed-effect model was carried out using lmer in lme4 package (Bates et al. 2014) and the significant values were displayed using R package car (www.r-project.org).

Table 4. Five principal component axes to variation in field parameters. PH: plant height; Pod#: total pod number; Seed#: total seed number; 100-SW: 100 seed weight; TSW: total seed weight.

	PC1	PC2	PC3	PC4	PC5
Year	0.74	0.23	-0.53	0.04	0.35
PH	0.52	-0.4	0.7	-0.18	0.23
Pod#	0.98	-0.12	-0.01	0.08	-0.1
Seed#	0.98	-0.12	0.04	0.09	-0.1
100-SW	0.36	0.79	0.06	-0.48	-0.08
TSW	0.98	-0.03	-0.03	0.07	-0.15
Protein	0.01	0.8	0.44	0.4	0.05
Eigenvalue	3.830	1.510	0.970	0.450	0.23
Proportion of variance	0.550	0.220	0.140	0.060	0.03
Cumulative proportion	0.550	0.760	0.900	0.960	1

Note: The prcomp algorithm of r (www.r-project.org) was used in the analysis.

produce a series of eigenvalues and associated eigenvectors (Johnson and Wichern, 2007). Eigenvectors are contained the correlations between original variables and PCs, which specify the axes of linear coefficients (Jolliffe, 2002) and the variance along direction of each PCs axis represents the eigenvalue in the analysis. To validate results of the mixed model analysis, PCA was applied to analyze the phenotypic components that contributed to the total seed weight in the dataset. As the complementary approach of the linear mixed model, PCA can handle the original variables through linear combination and produced a new set of uncorrelated PCs (Jolliffe, 2002; Johnson and Wichern, 2007). A total of 5 PCs were generated in the dataset (Table 4). The five PCs represent both main principal components and proportional contribution of PCs to the total variation of the data collected during 2016–2017 (Table 4). The first PC accounted for 55% of variation observed from 2016 to 2017. The second principal component contributed 22% of the total variation in the dataset. The results showed that except of protein, all other components were positively contributed to PC1. The second PC accounted for 22% of the total variation

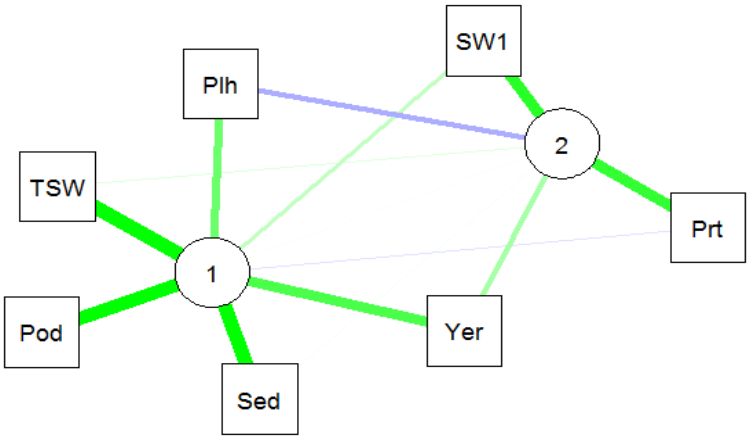

Figure 3. The relationship of PC1 vs. PC2 from principal component analysis (PCA) among the field parameters. Plh: plant height; Pod: total pod number; Prt: protein; Sed: total seed number; SW1: 100 seed weight; TSW: total seed weight. Green bar: positive contribution; Blue bar: Negative contribution. The size of the bar will be increased with the contribution.

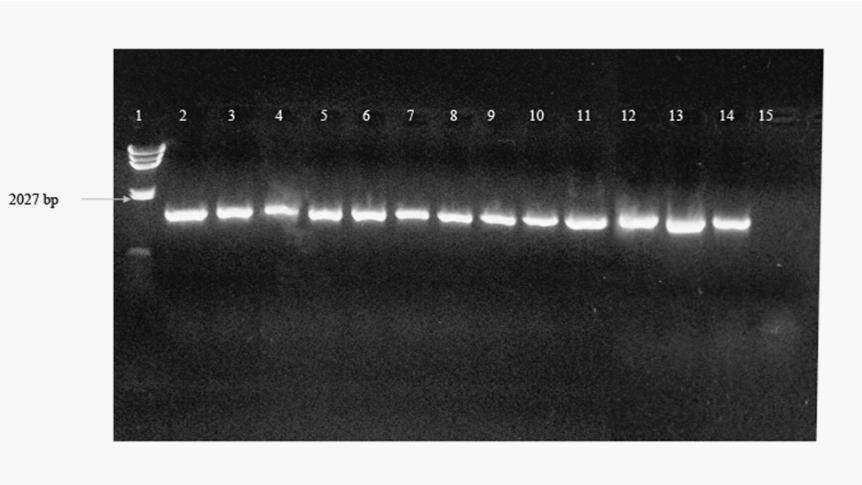

Figure 4. PCR amplicons using 16S rRNA primers (27F/1492R). Lane 1, Lambda DNA HindIII size marker; lanes 2-14, amplicons from strains AJK-1- AJK-13, respectively. Lane 15, negative control.

Table 5. Summary of annotation of strains and their GenBank matches.

Strain#	Strain Name (16S rRNA Accession#)	Genus (Class)	Closest GenBank Match Accession# (% similarity)
1	AJK-1 (MN715357)	*Brevundimonas* (Alphaproteobacteria)	MN704401 (100%)
2	AJK-2 (MN715358)	*Microbacterium* (Actinobacteria)	KY292472 (99.09%)
3	AJK-3 (MN715359)	*Nakamurella* (Actinobacteria)	NR_157706 (100%)
4	AJK-4 (MN715360)	*Flavobacterium* (Flavobacteriia)	EU057851 (100%)
5	AJK-5 (MN715361)	*Pseudomonas* (Gammaproteobacteria)	MH883923 (100%)
6	AJK-6 (MN715362)	*Frigoribacterium* (Actinobacteria)	KR922277 (100%)
7	AJK-7 (MN715363)	*Paenarthrobacter* (Actinobacteria)	MG686085 (100%)
8	AJK-8 (MN715364)	*Rhizobium* (Alphaproteobacteria)	MK611735 (100%)
9	AJK-9 (MN715365)	*Microbacterium* (Actinobacteria)	KX981237 (100%)
10	AJK-10 (MN715366)	*Bacillus* (Bacilli)	MK934384 (100%)
11	AJK-11 (MN715367)	*Bacillus* (Bacilli)	MN640965 (100%)
12	AJK-12 (MN715368)	*Aminobacter* (Alphaproteobacteria)	DQ401867 (100%)
13	AJK-13 (MN715369)	*Aminobacter* (Alphaproteobacteria)	MK382451 (99.92%)

to the phonotypical variation. Protein (0.8), total seed weight (0.79), and year (0.23) positively contributed to this principal component and the remaining variables contributed to the PC negatively. The last three PC contributed only 23% of the total variation in the population and PC3 contributed 14% to the total variation in the population (Table 4). The r packages (www.r-project.org) were used to visually display multivariate relationship of the first two principal components with the field yield parameters using variables of the two year trials (Figure 3). The green color bar represented positive contribution to the PCs while blue color bar showed negative contribution to the PC. The graph of two PCs described the contribution of field parameters to the corresponding PCs at the multivariate algorithm suggesting that the field parameters were mainly contributing to PCs positively in the EMS mutant line (Figure 3) and only plant height was negatively contributing PC2.

Gel Electrophoresis and Phylogenetic Analysis of Soil Bacteria

Thirteen bacterial strains were isolated from the field's soil and identified using 16S rRNA sequences (Table 5). Bands with the expected size were obtained from the agarose gel when amplicons of DAN samples of bacterial strains were amplified using the primers derived from 16S rRNA (Figure 4). The annotation of the sequenced amplicons suggested that the strains were identified as: 5 *Actinobacteria*, 4 *Alphaproteobacteria*, 2 *Bacilli*, 1 *Gammaproteobacteria*, and 1 *Flavobacteriia* (Table 5). Out of the four *Alphaproteobacteria*, one strain was identified as *Rhizobium giardinii* (Table 5). The 16S rRNA sequences were deposited in GenBank under accession numbers MN715357-MN715369. The presence of nitrogen fixation *Rhizobium giardinii* seemed to explain the beneficial effect to the soybean plants grown in the unfertilized rain feed field. It is believed that *Rhizobia* help in increased plant growth when interacting with legumes (Mabrouk et al., 2018) and maintained healthy growth of the soybean plants. In addition, previous studies have shown that some bacterial strains, besides *Rhizobium*, that we isolated from the field also have the potential to promote plant growth such as: *Brevundimonas* (Kumar and Gera, 2014), *Microbacterium* (Cordovez et al., 2018), *Pseudomonas*, and *Bacillus* (Di Benedetto et al., 2019).

In conclusion, we demonstrated that the soybean mutant FM6-847 performed outstanding in yield, yield components, and several seed composition traits (protein and oil contents) and is well adapted to North Carolina. We also isolated and identified several beneficial bacteria for soybean growth and development including the nitrogen-fixing *Rhizobium giardinii*, and other beneficial species such as *Brevundimonas*, *Microbacterium*, *Pseudomonas*, and *Bacillus* for beneficial effect on soybean plant growth.

Acknowledgments

We thank P. Ratcliff and the rest of the undergraduate student crew at FSU for taking care of the plants in the greenhouse and field. Also, we thank Dr. Fares Najar from The Chemistry/Biochemistry Bioinformatics Core (CBBC) of the University of Oklahoma for his help in the sequence assembly and analysis.

References

Afuape SO, PI Okocha, and D Njoku (2011) Multivariate assessment of the agro-morphological variability and yield components among sweet potato (*Ipomoea batatas* (L.) Lam) landraces. Afr. J. Plant Sci. 5: 123–132.

Akond A, S Liu, SK Kantartzi, K Meksem, N Bellaloui, DA Lightfoot, J Yuan, D Wang, J Anderson, and MA Kassem (2015) A SNP Genetic Linkage Map Based on the 'Hamilton' by 'Spencer' Recombinant Inbred Line (RIL) Population of Soybean [*Glycine max* (L.) Merr.] Identified QTL for Seed Isoflavone Contents. Plant Breeding 134: 580–588. https://doi.org/10.1111/pbr.12298.

Akond M, B Ragin, R Bazzelle, SK Kantartzi, K Meksem, and MA Kassem (2012) Quantitative Trait Loci Associated with Moisture, Protein, and Oil Content in Soybean [*Glycine max* (L.) Merr.]. Journal of Agricultural Science 4 (11): 16-25.

Akond M, B Ragin, R Bazzelle, W Clark, SK Kantartzi, K Meksem, and MA Kassem (2013a) Effect of Two Row Spaces on Several Agronomic Traits in Soybean [*Glycine max* (L.) Merr.]. Atlas Journal of Plant Biology 1 (2): 18–23.

Akond M, R Bazzelle, B Ragin, H Herrera, U Kaodi, C Akbay, SK Kantartzi, V Njiti, A Barakat, K Meksem, DA Lightfoot, and MA Kassem (2013b) Additional Quantitative Trait Loci and Candidate Genes for Seed Isoflavone Content in Soybean. J. Agricultural Science 5 (11): 20-33.

Akond M, S Liu, M Boney, SK Kantartzi, K Meksem, N Bellaloui, DA Lightfoot, and MA Kassem (2014a) Identification of Quantitative Trait Loci (QTL) Underlying Protein, Oil, and Five Major Fatty Acids' Contents in Soybean. American Journal of Plant Sciences 5 (1): 158-167.

Akond M, S Liu, SK Kantartzi, K Meksem, N Bellaloui, DA Lightfoot, J Yuan, D Wang, and MA Kassem (2014b) Quantitative Trait Loci for Seed Isoflavone Contents in MD96-5722 by 'Spencer' Recombinant Inbred Lines of Soybean. Journal Agriculatural Food Chem 62: 1464-1468.

Altschul SF, W Gish, W Miller, EW Myers, and DJ Lipman (1990) Basic local alignment search tool. J Mol Biol. 215: 403–410.

Ambrocio Z, S Kantartzi, K Meksem, and MA Kassem (2016) Influence of Drought Stress on Several Root Traits and their Correlation with Seed Protein and Oil Contents in Soybean. Atlas Journal of Biology 2016, pp. 267–273. https://doi.org/10.5147/ajb.2016.0140.

Bates B, M Mächler, MB Bolker, and CS Walker (2014) Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67: 1–51. https://doi.org/10.18637/jss.v067.i01.

Bellaloui N, A Mengistu, and DR Paris (2008) Soybean seed composition in cultivars differing in resistance to charcoal rot (*Macrophomina phaseolina*). J. Agric. Sci. 146: 667–675. https://doi.org/10.1017/S0021859608007971.

Bellaloui N, A Mengistu, and MA Kassem (2013a) Effects of Genetics and Environment on Fatty Acid Stability in Soybean Seed. Food and Nutrition Sciences 4 (9A1): 165–175.

Bellaloui N, A Mengistu, ER Walker, and LD Young (2014) Soybean seed composition as affected by seeding rates and row spacing. Crop Science 54: 1782–1795. https://doi.org/10.2135/cropsci2013.07.0463.

Bellaloui N, HA Bruns, AM Gillen, HK Abbas, RM Zablotowicz, A Mengistu, and RL Paris (2010a) Soybean seed protein, oil, fatty acids, and mineral composition as influenced by soybean-corn rotation. Agricultural Sciences 1 (3): 102-109. https://doi.org/10.4236/

as.2010.13013.

Bellaloui N, JR Smith, AM Gillen, and JD Ray (2010b) Effect of maturity on seed sugars as measured on near-isogenic soybean (*Glycine max*) lines. Crop Sci. 50: 1978–1987. https://doi.org/10.2135/cropsci2009.10.0596.

Bellaloui N, JR Smith, AM Gillen, and JD Ray (2011) Effects of maturity, genotypic background, and temperature on seed mineral composition in near-isogenic soybean lines in the early soybean production system. Crop Sci. 51: 1161–1171. https://doi.org/10.2135/cropsci2010.04.0187.

Bellaloui N, JR Smith, AM Gillen, DK Fisher, and A Mengistu (2012) Effect of shade on seed protein, oil, fatty acids, and minerals in soybean lines varying in seed germinability in the early soybean production system. Am. J. Plant Sci. 3: 84–95. https://doi.org/10.4236/ajps.2012.31008.

Bellaloui N, Y Hu, A Mengistu, HK Abbas, MA Kassem, and M Tigabu (2016) Elevated Atmospheric Carbon Dioxide and Temperature Affect Seed Composition, Mineral Nutrition, and 15N and 13C Dynam¬ics in Soybean Genotypes under Controlled Environments. Atlas J. Plant Biology 2016, pp. 56–65. https://doi.org/10.5147/ajpb.2016.0157.

Bellaloui N, Y Hu, A Mengistu, MA Kassem, and CA Abel (2013b) Effects of foliar boron application on seed composition, cell wall boron, and seed δ15N and δ 13C isotopes in water-stressed soybean plants. Frontiers in Plant Science, Plant Nutrition, Volume 4, Article 270: 1-12. https://doi.org/10.3389/fpls.2013.00270.

Brim CA and JW Burton (1978) Recurrent selection in soybeans. II. Selection for increased percent protein in seeds. Crop Sci. 19 (4): 494-498.

Cordovez V, S Schop, K Hordijk, HD de Boulois, F Coppens, I Hanssen, JM Raaijmakers, and VJ Carrión (2018) Priming of Plant Growth Promotion by Volatiles of Root-Associated Microbacterium spp. Appl Environ Microbiol 84. https://doi.org/10.1128/AEM.01865-18.

Dalley CD, JJ Kells, and KA Renner (2004) Effect of Glyphosate Application Timing and Row Spacing on Corn (Zea mays) and Soybean (*Glycine max*) Yields. Weed Technology 18 (1): 165-176.

Di Benedetto NA, D Campaniello, A Bevilacqua, MP Cataldi, M Sinigaglia, Z Flagella, and MR Corbo (2019) Isolation, Screening, and Characterization of Plant-Growth-Promoting Bacteria from Durum Wheat Rhizosphere to Improve N and P Nutrient Use Efficiency. Microorganisms 7 (11): 541. https://doi.org/10.3390/microorganisms7110541.

Dornbos DL and RE Mullen (1992) Soybean seed protein and oil contents and fatty acid composition adjustments by drought and temperature. J Am Oil Chem Soc 69 (3): 228–231.

Espina MJ, CMS Ahmed, A Bernardini, E. Adeleke, Z. Yadegari, P Arelli, and A Taheri, A (2018) Development and Phenotypic Screening of an Ethyl Methane Sulfonate Mutant Population in Soybean. Frontiers in Plant Science, 9. doi: 10.3389/fpls.2018.00394

Fierer N and RB Jackson (2006) The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci USA 103: 626-631.

Gana AS, SZ Shaba, and EK Tsado (2013) Principal component analysis of morphological traits in thirty-nine accessions of rice (*Oryza sativa* L.) grown in a rainfed lowland ecology of Nigeria. J. Plant Breed. Crop Sci. 5: 120–126.

Ivey S, K Ouertani, E Washington, P Lage, SK Kantartzi, K Meksem, DA Lightfoot, and MA Kassem (2011) Evaluation of Agronomic Traits in 'Essex' By 'Forrest' Recombinant Inbred Line Population of Soybean [*Glycine max* [L.] Merr.]. Atlas Journal of Plant Biology 1 (1): 13-17.

Jacobson J, A Alcivar, J Rainho, and MA Kassem (2007) Genomic Regions Containing QTL for Plant Height, Internodes Length, and Flower Color in Soybean [*Glycine max* (L.) Merr.]. BIOS A Quarterly Journal of Biology 78 (4): 119-126.

Johnson RA and WD Wichern (2007) Applied multivariate statistical analysis. Prentice-Hall, Upper Saddle River, NJ.

Jolliffe IT (2002) Principal component analysis. 2nd ed. Springer series in statistics. Springer-Verlag, New York, NY.

Kassem MA, K Meksem, CH Kang, VN Njiti, VY Kilo, AJ Wood, and DA Lightfoot (2004a) Loci Underlying Resistance to Manganese Toxicity Mapped in a Soybean Recombinant Inbred Line population of 'Essex' x 'Forrest'. Plant and Soil. 260 (1-2): 197-200.

Kassem MA, Meksem K, Iqbal MJ, Njiti VN, Banz WJ, Winters TA, Wood A, Lightfoot DA (2004b) Definition of Soybean Genomic Regions That Control Seed Phytoestrogen Amounts. J. Biomed & Biotech 1 (1): 52-60.

Khandaker L, M Akond, S Liu, SK Kantartzi, K Meksem, N Bellaloui, DA Lightfoot and MA Kassem (2015) Mapping of QTL Associated with Seed Amino Acids Content in 'MD 96-5722' by 'Spencer' RIL Population of Soybean using SNP Markers. Food and Nutrition Sciences 6: 974–984. https://doi.org/10.4236/fns.2015.611101.

King B (2013) Multiple regression. [Online]. Available: http://ww2.coastal.edu/kingw/statistics /R-tutorials/multregr.html.

Kumar V, and R Gera (2014) Isolation of a multi-trait plant growth promoting Brevundimonas sp. and its effect on the growth of Bt-cotton. 3 Biotech. 4 (1): 97–101.

Lightfoot DA, VN Njiti, PT Gibson, MA Kassem, JM Iqbal, and K Meksem (2005) Registration of the Essex by Forrest Recombinant Inbred Line (RIL) Mapping Population. Crop Science 45 (4): 1678-1781.

Luedders VD (1977) Effect of seed coat color on protein and oil content of soybeans. Journal of Heredity 68: 321–322.

Mabrouk, Y, I Hemissi, IB Salem, S Mejri, M Saidi, and O Belhadj (2018) Potential of Rhizobia in Improving Nitrogen Fixation and Yields of Legumes. Symbiosis. https://doi.org/10.5772/intechopen.73495.

Machado M, WC Magalhães, A Sene, B Araújo, AC Faria-Campos, SJ Chanock, L Scott, G Oliveira, E Tarazona-Santos, and MR Rodrigues (2011) Phred-Phrap package to analyses tools: a pipeline to facilitate population genetics re-sequencing studies. Investig Genet. 2: 3. https://doi.org/10.1186/2041-2223-2-3.

Maestri DM, DO Labuckas, JM Meriles, AL Lamarque, JA Zygadlo, and CA Guzman (1998) Seed composition of soybean cultivars evaluated in different environmental regions. J Sci Food Agric 77: 494-498.

Mangiafico SS (2015) An R Companion for the Handbook of Biological Statistics. Rutgers Cooperative Extension, New Brunswick, NJ.

Meksem K, VN Njiti, WJ Banz, MJ Iqbal, MA Kassem, DL Hyten, J Yuang, TA Winters, and DA Lightfoot (2001) Genomic regions that underlie soybean seed isoflavone content. J. Biomed. And Biotech. 1 (1): 38-44.

Nie M, Y Wang, J Yu, M Xiao, L Jiang, J Yang, C Fang, J Chen, and B Li (2011) Understanding plant-microbe interactions for phytoremediation of petroleum-polluted soil. PLoS One. 6:e17961.

Ojo DK, MA Adebisi, and BO Tijani (2002) Influence of environment on protein and oil contents of soybeans seed (*Glycine Max* (L.) Merril). Global J Agri Sci 1 (1): 27-32.

Ouma EW, AM Asango, J Maingi, and EM Njeru (2016) Elucidating the potential of native rhizobial isolates to improve biological nitrogen fixation and growth of common bean and soybean in smallholder farming systems of Kenya. International Journal of Agronomy. 1-7. http://dx.doi.org/10.1155/2016/4569241.

Patil A, SP Taware, MD Oak, SA Tamhankar, and VS Rao (2007) Im-

provement of Oil Quality in Soybean [*Glycine max* (L.) Merrill] by Mutation Breeding. Journal of the American Oil Chemists Society, 84(12), 1117–1124. doi: 10.1007/s11746-007-1146-1

Placide R, H Shimelis, M Laing, and D Gahakwa (2015) Application of principal component analysis to yield and yield related traits to identify sweet potato breeding parents. Trop. Agric. 92: 1–15.

Ragin B, M Akond, SK Kantartzi, K Meksem, H Herrera, C Akbay, DA Lightfoot, and MA Kassem (2014) Effect of Row Spacing on Seed Isoflavone Contents in Soybean [*Glycine max* (L.) Merr.]. American Journal of Plant Sciences 5 (26): 4003-4010.

Ragin B, R Bazzelle, W Clark, S Kantartzi, K Meksem, and MA Kassem (2012) QTL for Yield and Yield Components Mapped in the PI 438489B By 'Hamilton' Recombinant Inbred Line (RIL) Population of Soybean [*Glycine max* (L.) Merr.]. Journal of Agricultural Science 4 (9): 98-105.

Scofield S, BJ Schemerhorn, G Cai, and GL Nowling (2016) Uniform soybean tests northern region. USDA-ARS Crop Production and Pest Control Research Unit Purdue University, W. Lafayette, IN 47907-1150.

Soystats (2019) http://soystats.com.

Starkweather J (2010) Linear fixed effects modeling using R. [Online]. Available: www.unt.edu/rss/class/Jon/ : : : /Linear MixedModels_JDS_Dec2010.pdf.

Vance CP (1998) Legume symbiotic nitrogen fixation: Agronomic aspects. In: Spaink HP et al., editors. The Rhizobiaceae. Dordrecht: Kluwer Academic Publishers. pp. 509-530.

Vu V and J Lei (2012) Minimax rates of estimation for sparse PCA in high dimensions. In N. Lawrence and M. Girolami. Proc. Fifteenth International Conference on Artificial Intelligence and Statistics (AISTATS). JMLRW&CP, La Palma, Canary Islands.

Wilson KH, RB Blitchington, and RC Greene (1990) Amplification of bacterial 16S ribosomal DNA with polymerase chain reaction. J Clin Microbiol 28: 1942-1946.

Wolf RB, JF Cavins, R Kleiman, and LT Black (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 (5): 230–232.

Yuan J, V Njiti, K Meksem, J Iqbal, K Triwitayakorn, My A Kassem, G Davis, M Schmidt, and D Lightfoot (2002) Quantitative trait loci in two soybean recombinant inbred line populations segregating for yield and disease resistance. Crop Sci. 42:271-277.

Yuan JZ, AM Murphy, D De Koeyer, M Lague, and B Bizimungu (2016) Effectiveness of the field selection parameters on potato yield in Atlantic Canada. Can. J. Plant Sci. 96: 701–710. https://doi.org/10.1139/cjps-2015-0267.

Atlas Journal of Plant Biology - ISSN 1949-1379. Published By Atlas Publishing, LP (www.atlas-publishing.org)