Trends in Soybean Trait Improvement over Generations of Selective Breeding

Robert W. Bruce, Christopher M. Grainger, Alexandra Ficht, Mehrzad Eskandari, and Istvan Rajcan★

ABSTRACT

Improvement of agronomic and seed quality traits in soybean [*Glycine max* (L.) Merr.] occurs in breeding programs as a result of crossing elite soybean genotypes and selection of superior offspring in each cycle. The objective of this study was to evaluate changes in traits over 100 yr of selection in Ontario, Canada, in two soybean breeding programs at the University of Guelph: Guelph Campus (Maturity Group [MG] 0 and I) and Ridgetown Campus (MG II) programs. Fifty-six Guelph cultivars (released between 1983 and 2016), 46 Ridgetown cultivars (released between 1987 and 2015), and 37 Guelph-pedigree related historical cultivars (from 1913 to 1985) were grown in field trials over 3 yr (Guelph Campus) and 2 yr (Ridgetown Campus) at two locations per breeding program to measure 13 traits. Yield increased in the Guelph Campus cultivars by 17.1 kg ha^{-1} yr^{-1} and by 15.7 kg ha^{-1} yr^{-1} in the Ridgetown Campus cultivars, while yield per day of maturity per year of release significantly increased in elite cultivars. Protein in the Ridgetown Campus cultivars increased at 0.76 g kg^{-1} yr^{-1}, whereas the Guelph Campus cultivars showed no trend. In Ridgetown Campus cultivars 100-seed weight significantly increased at 0.1 g yr^{-1}. Within the historical cultivars, 100-seed weight was significantly correlated with both protein ($r = 0.50$) and oil ($r = -0.42$). Genotype × trait biplot analyses revealed differential patterns of trait relationships between the elite and historical cultivars. Oil and yield were associated within the Guelph Campus cultivars. The association of 100-seed weight and protein was only found in the historical cultivars. Yield has increased in both breeding programs while seed traits have been altered over time.

R.W. Bruce, C.M. Grainger, A. Ficht, and I. Rajcan, Dep. of Plant Agriculture, Crop Science Building, Univ. of Guelph, 50 Stone Road East, Guelph, ON, N1G 2W1; M. Eskandari, Dep. of Plant Agriculture, Agronomy Building, Univ. of Guelph–Ridgetown Campus, 120 Main St. East, Ridgetown ON, N0P 2C0. Received 3 Nov. 2018. Accepted 13 May 2019. ★Corresponding author (irajcan@uoguelph.ca). Assigned to Associate Editor Aaron Lorenz.

Abbreviations: BLUE, best linear unbiased estimator; DTM, days to maturity; MG, maturity group; NIR, near-infrared reflectance.

S OYBEAN [*Glycine max* (L.) Merr.] is a major worldwide oilseed crop accounting for >340 million metric tons in 2017 (USDA, 2019). Soybean in Canada, are grown annually, on average, on 2.9 million ha with an average yield of 2.6 t ha^{-1} in 2019. Soybean is also an important rotational crop, often following corn (*Zea mays* L.) and preceded by winter wheat (*Triticum aestivum* L.) in Ontario. Because of soybean's importance to Ontario agriculture and food security, resources are allocated to breeding efforts to increase crop yield within public and private soybean breeding programs. However, little effort has been made to study the retrospective progress of single breeding programs, so as to understand the historical trends at a breeding program's scale, and to characterize the phenotypic diversity used in current soybean breeding programs in Ontario, Canada.

Increasing interest and demand for food-grade soybeans has driven a niche market for specialty-trait identity-preserved soybean cultivars, for which the University of Guelph is an important source in Canada. For the development of specialty soybean cultivars, the selection process places a higher emphasis on seed quality traits such as seed size, fatty acid, and sugar profiles (depending on the end use of the crop) while still maintaining agronomic performance. Within breeding programs developing specialty-trait soybeans, phenotypic diversity for seed composition and size is required for making continued improvement in new cultivars. Little information

is available as to the historical changes through breeding for these traits, whether under direct selection or not. Data regarding ranges of phenotypic diversity at a breeding program scale for seed and agronomic traits could help to inform breeders of the historical diversity within breeding germplasm, and how selections have changed this diversity within breeding germplasm. This information can be used to leverage trends and trait correlations in the creation of new cultivars in the current era of inexpensive, fast, and reliable seed phenotyping technologies.

Yield has been a major trait for improvement in soybeans, with a number of agronomic studies conducted to quantify trait improvement. Yield increases (per year) range from 11 to ∼32 kg ha^{-1} yr^{-1} in various studies using germplasm from Maturity Group (MG) IV to MG 000. In 24 MG II soybean cultivars, Koester et al. (2014) identified a 26.5-kg ha^{-1} yr^{-1} improvement from 1923 to 2007, with extended growing season noted in modern cultivars compared with historical cultivars. Using 60 MG III soybean cultivars, Specht et al. (2014) identified a 23-kg ha^{-1} yr^{-1} yield increase from 1929 to 2010, with a higher rate of gain being since 1980 at 29.4 kg ha^{-1} yr^{-1}. The latter results was related to genetic and agronomic improvements in recent years as well as the impact of climate change (Specht et al., 2014). Since 1972, on-farm US soybean yields have increased at 31.4 ha^{-1} yr^{-1} compared with an overall yield increase of 22.6 ha^{-1} yr^{-1} (Specht et al., 1999).

Within short-season soybeans, Cober and Voldeng (2012) reported a 20-kg ha^{-1} yr^{-1} yield increase from 1971 to 2000. Exponential growth in the number of available cultivars was identified starting in the 1980s, although rate of genetic gain was linear over this period despite increasing public and private investment in soybean breeding efforts. Voldeng et al. (1997) reported that within 41 MG 0, 00, and 000 cultivars released from 1934 to 1992, yield increased at 11 kg ha^{-1} yr^{-1} before 1976, whereas since 1976, it increased by 30 kg ha^{-1} yr^{-1}. In the same period, seed oil concentration was increasing at 4 g kg^{-1} yr^{-1}, while protein concentration was decreasing at 4 g kg^{-1} yr^{-1}. Rincker et al. (2014) found that yield stability and seed oil concentration were increasing while seed protein was decreasing in MG II, MG III, and MG IV soybeans. Although overall agronomic trends have been identified in the literature, yield improvements in pedigree-related breeding cultivars have not been studied, nor have seed-trait-specific trends in a breeding program.

The University of Guelph, Guelph, ON, Canada, has two active public soybean breeding programs with unique cultivars and experimental genotypes available for study. The Ridgetown program (Ridgetown, ON, Canada) develops soybeans mainly for MG II, whereas the Guelph program (Guelph, ON, Canada) develops MG 0 and I soybean cultivars. Pedigree information for cultivars produced within the two breeding programs allows for the identification of historical germplasm related to the current, elite soybean cultivars. Although both programs have the ultimate goal of increasing seed yield, other seed traits make their soybeans attractive for food-grade end users. We hypothesize that selective breeding in each breeding program has resulted in genetic gain for yield and altered seed composition traits through years of selection.

The first objective of this study was to characterize and compare the changes in multiple agronomic and seed traits over decades of breeding within two breeding programs at the University of Guelph. The second objective was to understand relationships between traits within soybean breeding programs and how these relationships have changed over decades of breeding. The third objective was to compare trends in trait values over time between the Guelph and Ridgetown soybean breeding programs.

MATERIALS AND METHODS
Selection of Germplasm
A set of cultivars was selected based on pedigree relationships to modern soybean cultivars produced from the Guelph and Ridgetown breeding programs. From the Guelph Campus breeding program, 56 cultivars were selected with release dates ranging from 1983 to 2016. From the Ridgetown Campus, 46 cultivars were selected with release dates ranging from 1987 to 2015. Selection was first based on broad usage of the cultivar in Ontario and within the breeding program in crossing, then to cover the period for which each breeding program has produced cultivars, and finally cultivar selection was limited to those cultivars with available seed to generate the final cultivar set for study.

Using pedigree records from each breeding program as well as published cultivar descriptions (Supplemental Table S1) to trace modern Guelph cultivars to their founders, 37 pedigree-related historical cultivars were selected for the study with release dates ranging from 1913 to 1985. Only cultivars with maturity appropriate for study in Ontario, Canada, were selected to be grown in field trials.

Field Trials
The 56 Guelph Campus cultivars and 37 historical cultivars were grown in 2015, 2016, and 2017 at the Woodstock Research Station (Woodstock, ON; 43.145223° N, 80.782389° W; Guelph loam soil type [fine-loamy, mixed, active, mesic Haplic Glossudalfs]) and in St. Pauls, ON (43.338539° N, 81.152799° W; Huron clay loam soil type [Gray-Brown Podzolic Great Soil Group]), as part of a larger breeding program test of up to 180 accessions. An additional field year in 2017 for the Guelph Campus and historical cultivars was provided through additional funding for a separate project. The trials were planted as two-replicate randomized complete block design with 500 seeds per four-row plot, a row spacing of 40 cm, plot spacing of 45 cm, plot length of 5 m, and a 2-m alley. St. Pauls trials were planted on 21 May 2015, 20 May 2016, and 20 May 2017, whereas Woodstock trials were planted on 5 June 2015, 8 June 2016, and 8 June 2017.

The 46 Ridgetown Campus cultivars were grown at Chatham, ON (42.201274° N, 82.152175° W), and Ridgetown, ON (42.4392° N, 81.8871° W), in 2015 and 2016 as part of a larger breeding program test of 150 accessions. Both Ridgetown Campus locations have a clay loam soil type. The Ridgetown Campus trials were planted similarly to the Guelph breeding program trials. Due to entry-based sampling (replicates were pooled after harvest) in Chatham 2016, only one sample of seed per entry was available for analysis and treated as a single replicate in that environment.

Trait Measurement

In 2015, 2016, and 2017, the following standard agronomic traits were measured: yield (kg ha^{-1} at 13% moisture), days to maturity (DTM), and plant height (cm). Seed traits for each yield plot were measured on whole seeds using a Perten DA 7250 SD near-infrared reflectance (NIR) spectrometer (Perten Instruments) using calibrations provided by Perten Instruments (Kovalenko et al., 2006; Hurburgh, 2007; Butler et al., 2010; Bellaloui et al., 2014) with provided calibration statistics (Supplemental Table S2). A medium-sized Perten sample tray of seeds from each plot was randomly sampled, hand-screened, and cleaned before measurement. All NIR traits are reported on a dry basis (0% moisture). Traits recorded via NIR were protein, oil, linoleic acid, linolenic acid, oleic acid, palmitic acid, stearic acid, sucrose, fiber, and moisture. Hilum color was visually assessed and 100-seed weight was measured from the random seed sample.

Selected traits were defined as traits on which breeders have made selections for direct trait improvement, including yield, seed size, oil, and protein. Other traits studied here have been changed through indirect selection pressure within each breeding program.

Statistical Analysis

Trait data were examined in each environment (year × location combination) to identify environments with poor trial performance before combining trait data. Within the Guelph data, average yield at Woodstock 2015 was found to be 1371 kg ha^{-1} (data not shown), lower than all other Guelph testing environments with a CV of 34%; therefore, it was not included in the combined analysis. No weather or planting issues could be identified for this low yield, but the uniformity of the yield loss suggests a field-wide effect, such as higher drought pressure or weed pressure than other locations in 2015. Other traits from Woodstock 2015 had similar average values to the remaining environments and low CV and were therefore kept for further analysis. At the five remaining environments, the values of all agronomic and seed traits were consistent across the environments and kept for combined analysis.

Within the Ridgetown data, all environments had consistent agronomic and seed trait performance, and all data were kept for combined analysis. Due to sampling errors in Chatham 2016, only a single entry-based seed sample was available, and it was treated as a single block from this environment.

Plot data from each Woodstock and St. Pauls locations were spatially corrected per plot to account for field variation using radial smoothing in PROC GLIMMIX in SAS 9.4 (SAS Institute, 2013). The model included cultivar as a fixed effect, cov_sp (smoothing spline effect) as a covariate and longitude and latitude (representing physical field positions of each plot, in meters between plot centers including plot alleys) as random effects (Bowley, 2015). The effect statement of cov_sp = spline (lat long) defined the smoothing spline effect in the model. Adjusted plot values were used in further combined analyses. For Chatham and Ridgetown locations, plot data were processed through Agrobase (Agronomix Software, 2019) using a nearest-neighbor adjustment to generate adjusted plot values, as is standard in the Ridgetown Campus breeding program. The model in Agrobase is $yij = m + v_i + t_{ij} + e_{ij}$, where y_{ij} is the plot value of the ith variety of the jth block for a given trait, m is the trial mean for a given trait, v_i is the effect of the ith variety, t_{ij} is the effect of the trend with the ith variety in the jth block, and e_{ij} is the residual for the ith variety in the jth block, where the two adjacent plots are used in the calculation of the local trend.

Adjusted plot values were used in a combined analysis in PROC GLIMMIX in SAS 9.4 in a model with a fixed effect (cultivar) and random effects (environment, cultivar × environment, and block within environment). A best linear unbiased estimator (BLUE) was calculated for each trait per cultivar and used as the final trait value presented in this study for each cultivar. Yield at Woodstock in 2015 was not used in the combined Guelph analysis, as previously described. Using the combined trait data, a set of BLUEs were derived for each of the cultivars across all year and location combinations for each breeding program for each agronomic and seed trait from which minimums, means, and maximums are presented (Table 1). An index of yield per DTM (yield/DTM) was created using combined environment yield and DTM data for each set of cultivars.

Pearson correlation coefficients were calculated in PROC CORR in SAS 9.4 for each pair of traits, where significance difference of the correlation coefficient from zero was declared at $p = 0.05$. Normality for each trait was checked using PROC UNIVARIATE, with a Type I error rate of $\alpha = 5\%$. Six-environment field data were used for correlation analysis for Guelph cultivars, whereas Ridgetown calculations used four environments.

Genotype × trait biplot analyses were conducted using the GGEbiplot program (Yan et al., 2000), where data were formatted in four-way format. Six-environment data were used for biplot analysis for Guelph germplasm, whereas four-environment were used for Ridgetown. Genotype × trait biplots were generated using a SD-scaled method and tester centered. Missing values were not imputed for any traits.

Linear regression of traits by year of release (cultivar registration) was conducted in PROC REG in SAS 9.4. The model used was $y = mx + b + e$, where the trait was denoted as y, the slope was denoted by m, intercept was denoted by b, and error was denoted by e. Regressions were declared significantly different than zero at $p < 0.05$. Only traits and groupings with significant regressions are presented as linear regressions in the trait by year analysis. Guelph cultivars were split into MGs 0 and I, with MG I being defined as 115 to 122 DTM, and MG 0 defined as 107 to 114.5 DTM based on the 2016 Ontario Soybean and Canola Committee (OSACC) report for relative maturity conducted at the same locations as the Guelph Campus field trials.

Table 1. Seed and agronomic trait best linear unbiased estimator (BLUE) minimums, means, and maximums for 37 historical cultivars, 56 Guelph cultivars calculated from six-environment field trials, and 46 Ridgetown cultivars calculated from four-environment field trails. Traits are presented on a dry basis where applicable.

| | Woodstock and St. Pauls | | | | | | Chatham and Ridgetown | | |
| | Historical | | | Guelph cultivars | | | Ridgetown cultivars | | |
Trait	Min.	Mean	Max.	Min.	Mean	Max.	Min.	Mean	Max.
100-seed weight (g)	14.1	17.8	22.9	15.9	18.3	21.5	15.4	19.2	24.1
DTM (d)†	106.0	121.3	135.0	104.8	115.6	127.1	109.4	122.5	135.3
Fiber (%)	5.7	6.1	6.3	5.8	6.1	6.5	5.9	6.1	6.4
Height (cm)	52.9	82.0	112.3	53.7	73.1	97.0	83.1	96.4	113.5
Linoleic acid (%)‡	49.7	52.5	54.8	49.1	52.4	55.4	47.4	51.3	55.0
Linolenic acid (%)‡	8.1	8.5	9.5	7.0	8.1	10.1	5.3	7.0	8.1
Oil (g kg^{-1})	171	191	205	182	200	211	192	204	218
Oleic acid (%)‡	20.3	23.6	27.2	18.4	23.6	28.8	19.6	25.3	30.7
Palmitic acid (%)‡	11.6	12.1	12.4	11.5	12.1	12.6	11.0	11.8	12.3
Protein (g kg^{-1})	404	428	466	385	421	448	401	425	447
Stearic acid (%)‡	3.6	3.9	4.2	3.5	3.8	4.1	3.4	3.8	4.3
Sucrose (%)	6.1	7.1	8.0	5.8	6.9	8.1	5.5	6.5	7.3
Yield (kg ha^{-1})§	1079	2537	3572	1479	2710	3534	3456	4138	4852
Year of release	1913	1960	1985	1983	2002	2016	1987	2003	2015

† DTM, days to maturity.

‡ % of oil.

§ Yield at 13% moisture.

RESULTS

Trends in Guelph Cultivars

In the Guelph Campus cultivars, trends were identified for yield by year of release (17.1 kg ha^{-1} yr^{-1}) (Fig. 1E) and yield per DTM per year of release at 0.15 kg ha^{-1} yr^{-1} (Fig. 1F). Yields of Guelph Campus cultivars were split by estimated MG, where the yield of MG I cultivars increased at 13.8 kg ha^{-1} yr^{-1} (Fig. 2A), and for the MG 0 cultivars, the rate of increase was 18.4 kg ha^{-1} yr^{-1} (Fig. 2B).

Within the Guelph cultivars, yield was positively correlated with DTM ($r = 0.82$) and plant height ($r = 0.60$), whereas DTM was also found to be positively correlated with plant height ($r = 0.63$) (Table 2). No significant correlations were observed for 100-seed weight in the Guelph cultivars. Sucrose had a significant and negative correlation with protein ($r = -0.70$). Among the fatty acid components, linoleic acid was significantly positively correlated with linolenic acid ($r = 0.56$) while showing a significant negative correlation with oleic acid ($r = -0.88$) and stearic acid ($r = -0.72$). Linolenic acid was also significantly negatively correlated with oleic acid ($r = -0.76$), whereas stearic and oleic acid were positively correlated ($r = 0.57$). Protein and oil among Guelph cultivars were significantly negatively correlated ($r = -0.61$) (Table 2).

Using genotype × trait biplots to assess multitrait relationships in the Guelph cultivars, the negative oil and protein correlation previously shown is visible by opposing vectors (Fig. 3A). Yield was positively associated with oil, plant height, DTM, and sucrose due to the acute vector angles while being negatively associated with protein. No association was found with 100-seed weight

and any of yield, oil, or protein, similarly demonstrated through the Guelph cultivar trait correlations (Table 2).

When Guelph cultivars were split into MG 0 and MG I for biplot analysis (Supplemental Fig. S1), the overall trends were similar to those reported for the overall Guelph Campus cultivars, with yield being associated with DTM and plant height. In MG 0 cultivars, yield was associated with protein and 100-seed weight, which was not seen in MG I or overall Guelph Campus cultivars. This indicates strong selection for increasing protein levels in MG 0 vs. MG I soybeans.

Trends in Historical Cultivars

No significant trends were identified for historical cultivars over year of release (Fig. 1). Within the historical cultivars, oil and protein showed a significant negative correlation ($r = -0.75$), whereas yield had significant correlations with DTM ($r = 0.73$) and plant height ($r = 0.69$) (Table 2). Days to maturity and plant height were also significantly correlated ($r = 0.91$). Seed weight had significant correlations with protein ($r = 0.50$) and with oil ($r = -0.42$). Using genotype × trait biplots, yield was associated with plant height and DTM in the historical cultivars (Fig. 3B). Similarly, seed weight and protein were also associated. Oil and protein had a negative association, and both of these traits showed no association with yield.

Trends in Ridgetown Cultivars

Yield for Ridgetown cultivars was increasing at 15.7 kg ha^{-1} yr^{-1} across the range of maturity of the cultivars tested, with most cultivars being MG II (Fig. 1E). Yield per DTM at Ridgetown was also significantly increasing at 0.17 kg ha^{-1} yr^{-1} DTM^{-1} (Fig. 1F).
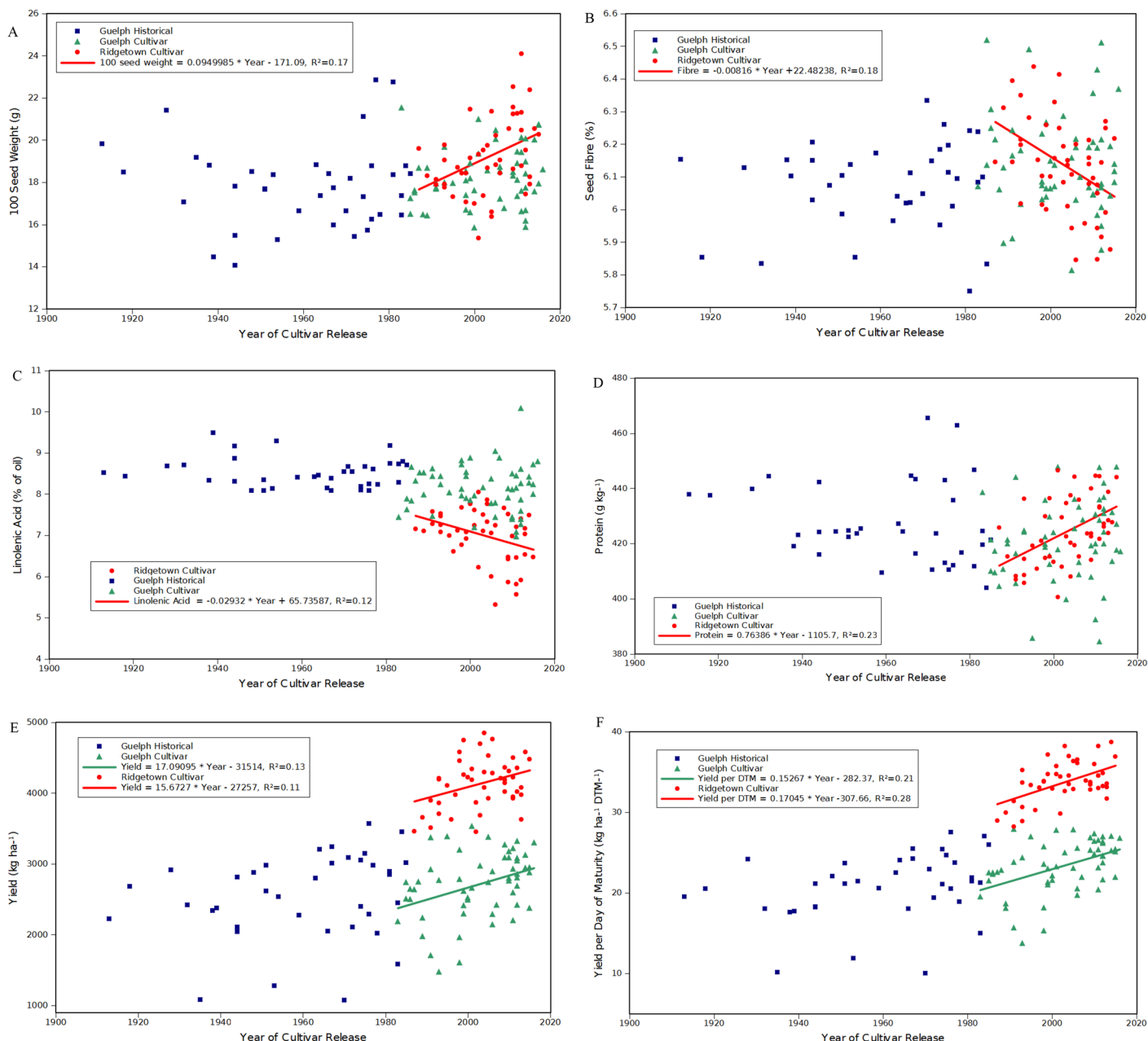
Fig. 1. Trait by year of release regression plots for (A) 100-seed weight (g), (B) seed fiber (%), (C) linolenic acid (% of oil), (D) protein (g kg^{-1}), (E) overall yield (kg ha^{-1}), and (F) yield per day of maturity (kg ha^{-1} day to maturity [DTM]$^{-1}$). Best linear unbiased estimator (BLUE) trait estimates from six environments (Guelph) and four environments (Ridgetown) were regressed against year of release, with significance tested at $\alpha = 0.05$. Regression lines are the same color as the germplasm they represent.

Trait values were regressed on year of release for Ridgetown Campus cultivars. Seed weight was significantly increasing in Ridgetown cultivars at 0.1 g yr^{-1} (Fig. 1A). Seed fiber (Fig. 1B) significantly decreased at 0.008% yr^{-1} among the Ridgetown cultivars, although the contribution of this component to seed composition is small. Linolenic acid decreased among the Ridgetown cultivars at a rate of 0.03% of oil per year (Fig. 1C). Seed protein was found to be significantly increasing for the Ridgetown cultivars at 0.76 g kg^{-1} yr^{-1} (Fig. 1D). This is expected, given the Ridgetown breeding objective of high-protein soybeans.

The oil and protein correlation among the Ridgetown cultivars was significant, at $r = -0.67$ (Table 3). Yield was significantly correlated with DTM ($r = 0.51$) and significantly correlated with plant height ($r = 0.35$). Protein had a significant negative correlation with DTM ($r = -0.30$) while having a significant positive correlation with seed weight ($r = 0.39$) (Table 3). A genotype × trait biplot analysis in the Ridgetown cultivars showed a negative association of oil and protein, whereas yield was closely associated with DTM (Fig. 3C). Seed weight was not found to be associated with any other seed or agronomic traits within the Ridgetown Campus cultivars.
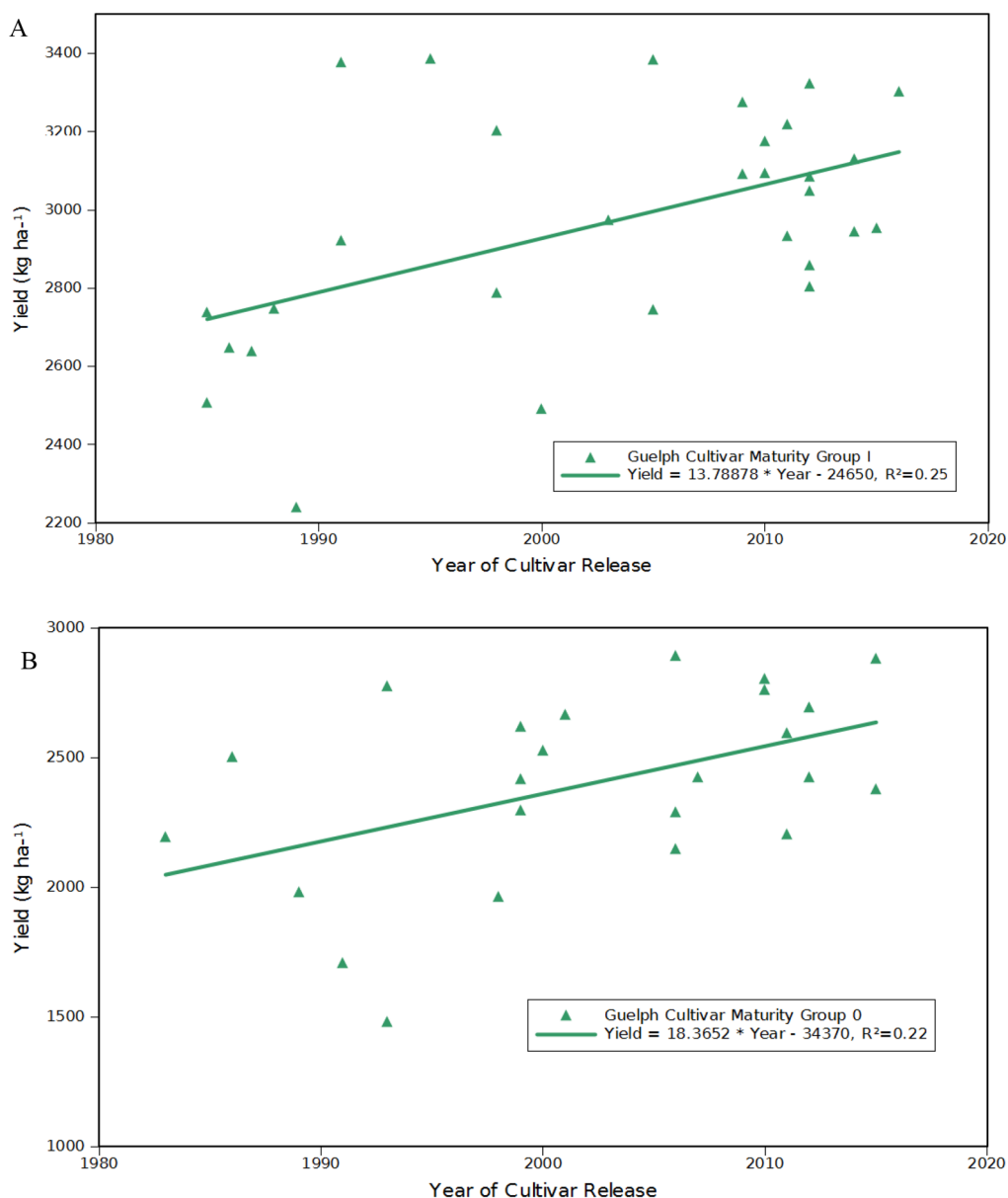
Fig. 2. Trait × year of release regression plots for (A) Maturity Group I yield (kg ha^{-1}) from 29 cultivars and (B) Maturity Group 0 yield (kg ha^{-1}) from 23 cultivars. Best linear unbiased estimator (BLUE) trait estimates from six environments (Guelph) were regressed against year of release with significance tested at $\alpha = 0.05$.

Comparison of Historical Cultivars to Elite Cultivars

No significant trait × year of release trends were identified for DTM (Supplemental Fig. S2A), plant height (Supplemental Fig. S2B), linoleic acid (Supplemental Fig. S2C), oil (Supplemental Fig. S2D), oleic acid (Supplemental Fig. S2E), palmitic acid (Supplemental Fig. S2F), stearic acid (Supplemental Fig. S2G), or sucrose (Supplemental Fig. S2H). The differentiation between the two breeding programs is apparent for DTM by year of release, whereas in recent years, later maturing cultivars have been released from the Guelph Campus program while the Ridgetown Campus has released several earlier maturing cultivars (Supplemental Fig. S2A).

Across all three germplasm groups, the fatty acid associations were similar, with stearic acid and oleic acid being positively associated, linoleic acid and linolenic acid being positively associated, and each of these pairs showing negative associations. Sucrose had a positive association with yield, though it has not been under direct selection at either the Guelph or Ridgetown Campus breeding programs. The range for linolenic acid among Guelph historical cultivars is narrower compared with cultivars from the two breeding programs.

DISCUSSION

The yield increases in breeding germplasm are in line with those reported earlier and summarized by Specht et al. (1999). The rates of gain were similar at both breeding programs, although overall yield was higher within the Ridgetown germplasm due to the later MG of the

Table 2. Pearson correlations of six-environment derived best linear unbiased estimator (BLUE) values for seed and agronomic traits in 56 Guelph cultivars (top diagonal) and six-environment derived BLUE values for seed and agronomic traits in 37 Guelph pedigree historical cultivars (bottom diagonal).

Trait	Yield	DTM†	100-seed weight	Height	Sucrose	Fiber	Linoleic acid	Linolenic acid	Oleic acid	Palmitic acid	Stearic acid	Oil	Protein
Yield	–	0.82*	0.25	0.60*	0.31*	0.02	0.03	-0.01	0.11	-0.05	0.02	0.07	-0.26
DTM	0.73*	–	0.17	0.63*	0.35*	0.08	0.02	0.05	0.08	-0.04	0.05	0.15	-0.37*
100-seed weight	0.06	-0.015	–	0.1	-0.15	-0.19	-0.05	-0.20	0.14	-0.18	-0.10	-0.05	0.2
Height	0.69*	0.91*	-0.18	–	0.38*	0.14	-0.05	0.11	0.05	-0.07	0.05	-0.07	-0.22
Sucrose	0.42*	0.39*	-0.37*	0.43*	–	0.36*	-0.13	0.35*	-0.02	0.27*	0.06	0.12	-0.70*
Fiber	-0.07	-0.12	-0.27	-0.17	0.19	–	-0.41*	0.12	0.16	0.59*	0.48*	0.12	-0.63*
Linoleic acid	0.33*	0.33*	-0.13	0.37*	0.18	-0.27	–	0.56*	-0.88*	-0.17	-0.72*	0.02	0.25
Linolenic acid	0.09	0.33*	-0.27	0.36*	0.51*	-0.20	0.63*	–	-0.76*	0.47*	-0.23	-0.38*	-0.07
Oleic acid	0.16	0.14	0.09	0.1	-0.04	0.16	-0.73*	-0.60*	–	-0.13	0.57*	0.07	-0.11
Palmitic acid	-0.45*	-0.49*	-0.18	-0.44*	-0.03	0.38*	-0.29	-0.15	-0.02	–	0.43*	-0.21	-0.31*
Stearic acid	-0.04	0.3	-0.08	0.26	0.25	-0.02	-0.27	0.1	0.28	0.04	–	-0.11	-0.24
Oil	0.24	-0.10	-0.42*	-0.07	0.24	0.28	0	-0.29	0.03	0.22	0.08	–	-0.61*
Protein	-0.39*	-0.23	0.50*	-0.21	-0.70*	-0.46*	0.05	-0.08	-0.13	-0.15	-0.29	-0.75*	–

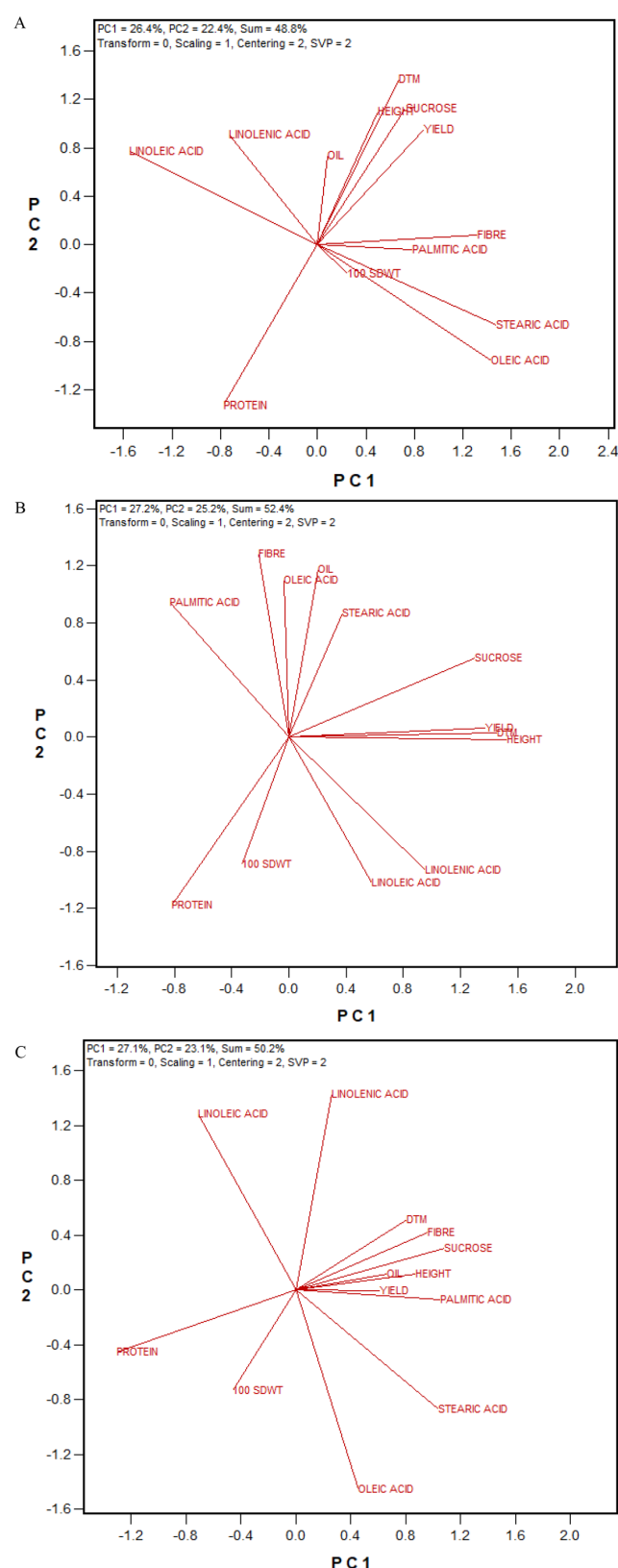* Significant at the 0.05 probability level.

† DTM, days to maturity.

Fig. 3. Genotype × trait biplot of (A) 56 Guelph cultivars with release years from 1983 to 2016 and (B) 37 historical cultivars with pedigree relationships to the Guelph cultivars with release years from 1913 to 1985 grown in two replicate trials at Woodstock and St. Pauls in 2015, 2016, and 2017, and (C) 46 Ridgetown cultivars from 1987 to 2015 grown in two replicate field trials at Ridgetown and Chatham in 2015 and 2016. DTM, days to maturity; 100 SDWT, 100-seed weight; PC, principal component; SVP, singular value partitioning.

Table 3. Pearson correlations of four-environment derived BLUE values for seed and agronomic traits in 46 Ridgetown cultivars.

Trait	DTM†	100-seed weight	Height	Sucrose	Fiber	Linoleic acid	Linolenic acid	Oleic acid	Palmitic acid	Stearic acid	Oil	Protein
Yield	0.51*	0.04	0.35*	0.32*	−0.17	−0.08	0.13	0.11	0.2	0.15	−0.05	−0.05
DTM		−0.23	0.50*	0.56*	0.12	−0.04	0.37*	0.06	0.19	0.23	−0.18	−0.30*
100-seed weight			−0.09	0.03	−0.19	−0.17	−0.33*	0.17	−0.04	−0.09	−0.26	0.39*
Height				0.40*	0.28	−0.11	0.04	0.12	0.14	0.16	0.15	−0.38*
Sucrose					0.25	−0.26	0.31*	0.09	0.47*	0.27	−0.03	−0.57*
Fiber						−0.24	0.29*	−0.10	0.41*	0.28	0.2	−0.61*
Linoleic acid							0.53*	−0.82*	−0.34*	−0.73*	0.07	0.16
Linolenic acid								−0.72*	0.31*	−0.16	−0.10	−0.28
Oleic acid									0.14	0.54*	−0.12	0.08
Palmitic acid										0.51*	−0.03	−0.36*
Stearic acid											0.17	−0.37*
Oil												−0.67*

* Significant at the 0.05 probability level.

† DTM, days to maturity.

breeding program and testing locations. The trend of increasing yield over time has been clearly established by breeding efforts in soybean. A remaining question is how this selection has influenced other traits in soybean that are not under direct breeder selection. From the biplots of Guelph cultivars, yield has been associated with increased DTM, plant height, and oil content at the expense of seed protein, similar to the oil and protein relationship described by Rincker et al. (2014). Through selection for yield within a given number of growing days, the plant efficiency at producing yield per day of maturity has also increased. No direct selection has occurred for lower linolenic acid in breeding germplasm in Ridgetown cultivars; however, decreases in this trait may be due to demands for increases in other fatty acid components, which also benefit end users due to increased oxidative stability of the soy oil low in linolenic acid.

Among the agronomic traits, the historical cultivars had a wide range for many of the traits, as the cultivars had a diverse range of origins, end uses, and targeted growing environments. Cultivars from both breeding programs showed a narrow range for parameters as they underwent rigorous testing before being selected for public release. From the field notes, historical cultivars had higher rates of seed loss due to pod shatter and lodging than modern elite cultivars (data not shown). The lack of trend for yield improvement among historical cultivars may be partly attributed to the diverse origins of the original founding PI genotypes and their derived cultivars, lack of selection pressure for improved yield, as well as their mixed end use (for forage production, not grain) and wide environmental distribution.

The two breeding programs had narrower ranges for 100-seed weight, DTM, height, oil, and yield than the Guelph historical cultivars. Average protein was lower (421 and 425 g kg⁻¹) in the Guelph and Ridgetown programs, respectively, while oil was higher (200 and 204 g kg⁻¹), compared with Guelph historical cultivars with 428 g kg⁻¹

protein and 191 g kg⁻¹ oil. The ranges of oil and protein values across the germplasm groups were similar to those reported in germplasm surveys, such as Qin et al. (2014) with an oil range of 142 to 228 g kg⁻¹ across 127 Chinese cultivars, and Song et al. (2016) with an oil range of 142 to 240 g kg⁻¹ across 763 Chinese cultivars. These two studies reported 318 to 498 g kg⁻¹ and 312 to 498 g kg⁻¹ protein, respectively. The maximum protein levels were similar, though the breeding cultivars in the present study had higher minimum protein levels. These two Chinese studies may have a lower reported minimum protein than this study due to the inclusion of later maturing cultivars than those present within the Guelph and Ridgetown breeding programs. The similar maximum ranges for protein indicate that the breeding programs have effectively maintained high levels of protein over generations of selection within their respective cultivars.

The Guelph cultivars had the lowest minimum protein levels among the three groups of germplasm while maintaining a wide range of protein levels. This may be due to the selection for cultivars with a range of end uses in the oilseed and food type markets. Although oil and protein were negatively correlated, the protein levels by year of release were decreasing at a significant rate. This may be due to specialty cultivars with trait-specific breeding objectives, such as high protein, rather than solely yield improvement. As the genetic basis of the oil and protein relationship is complex and still poorly understood, a number of questions remain to be addressed (Patil et al., 2017).

The absence of a trend for 100-seed weight in Guelph cultivars may be due to the food-grade specialty cultivar breeding efforts in recent years in Guelph, with specific seeds being desired for some food-type processes (e.g., natto or torfu). This is supported by the increasing range of seed weights in recent cultivars. No trend for linolenic acid was observed among the Guelph released cultivars,

as trait values among these cultivars fall into the typical range of linolenic acid for commercial cultivars.

One of the long-term goals in soybean oil research has been to reduce the linolenic acid component (Rahman and Takagi, 1997) to improve the flavor, frying, and stability characteristics (Dutton et al., 1951; Warner and Fehr, 2008). The minimum value for linolenic acid (5.3% of total oil) reported in this study for Ridgetown cultivars is higher than those achieved through mutagenesis or genetic modification. An example is Hoshino et al. (2014) using a TILLING (targeting induced local lesions in genomes) approach where linolenic acid levels <2% have been reported. As the range of low-linolenic-acid germplasm naturally occurring in soybean is small, future breeding efforts may require mutagenesis or other approaches to find suitable germplasm if lower linolenic acid levels are desired in cultivars.

Plant height and DTM were correlated in all sets of cultivars, though these appear weaker in the breeding programs. Sucrose had a similar range across the three germplasm groups, which may be due to the lack of active selection on this trait during soybean breeding. Days to maturity for both breeding programs showed a narrower range than for historical cultivars, as each breeding program focuses on a specific set of MGs while the historical cultivars encompass a wider range of MGs including later maturing cultivars.

Correlations of oil and protein have been repeatedly shown as negative in soybean cultivars in field-grown experiments (Openshaw and Hadley, 1984). Correlations for oil and protein in this study were similar in magnitude and direction as reported by Cober and Voldeng (2000), who estimated the correlation for single cross experimental genotypes at $r = -0.84$ and for experimental backcross genotypes at $r = -0.86$. In another study using 127 Chinese genotypes, a correlation of $r = -0.716$ was reported between oil and protein (Qin et al., 2014).

These differential relationships between Guelph cultivars and historical cultivars show that the correlation of oil and yield in elite cultivars has occurred in recent breeding efforts due to selection for high-yielding cultivars with high oil content. Breeders have also removed the association of 100-seed weight and protein in elite cultivars, possibly due to selection for high yield and high oil, thus making all elite cultivars lower in protein.

The data presented may support the hypothesis that specialty-trait breeding objectives (such as food-grade high protein) have expanded the range of oil and protein within the breeding germplasm to meet market niches outside of standard breeding objectives, such as solely for yield increases. It would be informative to investigate these relationships within other public and private breeding programs to determine trends in trait selections for non–food-grade soybeans.

CONCLUSION

Overall, this study has uncovered important trends in trait improvement in short-season soybean breeding at the University of Guelph, which may be present in other North American soybean breeding programs. Yield and seed oil content were increased through breeding efforts over the period studied. Trait correlations were similar to those previously reported in the literature, with the typical negative oil and protein correlation also present in the germplasm studied. The association of yield and oil was identified in cultivars from Ridgetown and Guelph and not in historical cultivars, indicating that this association is the result of selection within each breeding program. High protein levels in recent Ridgetown cultivars are the result of selection for food-grade soybeans, whereas in Guelph cultivars, a wider range of protein and oil has been maintained to satisfy diverse market requirements in new cultivars. Overall, yield increases were identified in both breeding programs over decades of selection, with similar improvements in the yields per given maturity. Soybean breeding programs may use the understanding of changes in agronomic and seed traits over decades of selection to help reverse or exploit the identified trends in the development of new cultivars.

Supplemental Material

Supplemental material is available online for this article.

Conflict of Interest

The authors declare that there is no conflict of interest.

References

Agronomix Software. 2019. Agrobase Generation II. Version 38.28.1. Agronomix Software, Winnipeg, MB, Canada.

Bellaloui, N., A. Mengistu, E.R. Walker, and L.D. Young. 2014. Soybean seed composition as affected by seeding rates and row spacing. Crop Sci. 54:1782–18795. doi:10.2135/cropsci2013.07.0463

Bowley, S. 2015. A hitchhiker's guide to statistics in biology. Generalized linear mixed model edition. Plants et al., Kincardine, ON, Canada

Butler, J.A., J.L. De Bruin, and P. Pedersen. 2010. Plant density effect on reduced linolenic acid soybean cultivars. Agron. J. 102:348. doi:10.2134/agronj2009.0167

Cober, E.R., and H.D. Voldeng. 2000. Developing high-protein, high-yield soybean populations and lines. Crop Sci. 40:39–42. doi:10.2135/cropsci2000.40139x

Cober, E.R., and H.D. Voldeng. 2012. A retrospective look at short-season soybean cultivar development in Ontario. Can. J. Plant Sci. 92:1239–1243. doi:10.4141/cjps2012-032

Dutton, H.J., C.R. Lancaster, C.D. Evans, and J.C. Cowan. 1951. The flavor problem of soybean oil. VIII. Linolenic acid. J. Am. Oil Chem. Soc. 28:115–118. doi:10.1007/BF02612206

Hoshino, T., S. Watanabe, Y. Takagi, and T. Anai. 2014. A novel GmFAD3-2a mutant allele developed through TILLING reduces α-linolenic acid content in soybean seed oil. Breed. Sci. 64:371–377. doi:10.1270/jsbbs.64.371

Hurburgh, C.R. 2007. Measurement of fatty acids in whole soybeans with near infrared spectroscopy. Lipid Technol. 19:88–90. doi:10.1002/lite.200600029

Koester, R.P., J.A. Skoneczka, T.R. Cary, B.W. Diers, and E.A. Ainsworth. 2014. Historical gains in soybean (*Glycine max* Merr.) seed yield are driven by linear increases in light interception, energy conversion, and partitioning efficiencies. J. Exp. Bot. 65:3311–3321. doi:10.1093/jxb/eru187

Kovalenko, I.V., G.R. Rippke, and C.R. Hurburgh. 2006. Measurement of soybean fatty acids by near-infrared spectroscopy: Linear and nonlinear calibration methods. J. Am. Oil Chem. Soc. 83:421–427. doi:10.1007/s11746-006-1221-z

Openshaw, S.J., and H.H. Hadley. 1984. Selection indexes to modify protein concentration of soybean seeds. Crop Sci. 24:1–4. doi:10.2135/cropsci1984.0011183X002400010001x

Patil, G., R. Mian, T. Vuong, V. Pantalone, Q. Song, P. Chen, et al. 2017. Molecular mapping and genomics of soybean seed protein: A review and perspective for the future. Theor. Appl. Genet. 130:1975–1991. doi:10.1007/s00122-017-2955-8

Qin, P., W. Song, X. Yang, S. Sun, X. Zhou, R. Yang, et al. 2014. Regional distribution of protein and oil compositions of soybean cultivars in China. Crop Sci. 54:1139–1146. doi:10.2135/cropsci2013.05.0314

Rahman, S.M., and Y. Takagi. 1997. Inheritance of reduced linolenic acid content in soybean seed oil. Theor. Appl. Genet. 94:299–302. doi:10.1007/s001220050414

Rincker, K., R. Nelson, J. Specht, D. Sleper, T. Cary, S.R. Cianzio, et al. 2014. Genetic improvement of U.S. soybean in Maturity Groups II, III, and IV. Crop Sci. 54:1419–1432. doi:10.2135/cropsci2013.10.0665

SAS Institute. 2013. The SAS system for Windows. Release 9.4. SAS Inst., Cary, NC.

Song, W., R. Yang, T. Wu, C. Wu, S. Sun, S. Zhang, et al. 2016. Analyzing the effects of climate factors on soybean protein, oil contents, and composition by extensive and high-density sampling in China. J. Agric. Food Chem. doi:10.1021/acs.jafc.6b00008

Specht, J.E., B.W. Diers, R.L. Nelson, J.F.F. de Toledo, J.A. Torrion, et al. 2014. Soybeans. In: S. Smith, et al., editors, Yield gains in major U.S. field crops. ASA, CSSA, and SSSA, Madison, WI. p. 311–356. doi:10.2135/cssaspecpub33.c12

Specht, J.E., D.J. Hume, and S.V. Kumudini. 1999. Soybean yield potential: A genetic and physiological perspective. Crop Sci. 39:1560–1570. doi:10.2135/cropsci1999.3961560x

USDA. 2019. World agricultural production. Circ. Ser. WAP 6-19. USDA Foreign Agric. Serv., Washington, DC. https://apps.fas.usda.gov/psdonline/circulars/production.pdf (accessed 8 July 2019).

Voldeng, H.D., E.R. Cober, D.J. Hume, C. Gillard, and M.J. Morrison. 1997. Fifty-eight years of genetic improvement of short-season soybean cultivars in Canada. Crop Sci. 37:428–431. doi:10.2135/cropsci1997.0011183X003700020020x

Warner, K., and W. Fehr. 2008. Mid-oleic/ultra low linolenic acid soybean oil: A healthful new alternative to hydrogenated oil for frying. J. Am. Oil Chem. Soc. 85:945–951. doi:10.1007/s11746-008-1275-1

Yan, W., L.A. Hunt, Q. Sheng, and Z. Szlavnics. 2000. Cultivar evaluation and mega-environment investigation based on the GGE biplot. Crop Sci. 40:597–605. doi:10.2135/cropsci2000.403597x