Phenotypic evaluation of Canadian × Chinese elite germplasm in a diversity panel for seed yield and seed quality traits

Authors: Priyanatha, Chanditha, and Rajcan, Istvan

Phenotypic evaluation of Canadian × Chinese elite germplasm in a diversity panel for seed yield and seed quality traits

Chanditha Priyanatha and Istvan Rajcan ⓘ

Department of Plant Agriculture, University of Guelph, Guelph, ON, N1G 2W1, Canada

Corresponding author: **Istvan Rajcan** (email: irajcan@uoguelph.ca)

Abstract

Despite mounting concerns regarding the narrowness of the genetic base of soybean ((*Glycine max* (L.) Merr.) in North America and the challenges that it may pose in the changing global environment and climate, exotic germplasm remains seldom used by breeders owing to various concerns. The objective of this study was to evaluate a Genome-Wide Association Study (GWAS) genomic diversity panel of 200 soybean genotypes for seed yield, seed quality, and agronomic trait performance. The GWAS panel consisted of lines derived from several generations of bi-parental crosses between elite Canadian and elite Chinese cultivars (CD–CH), elite Canadian cultivars (CD), and exotic elite Chinese cultivars (CH) evaluated at Elora and Woodstock, ON, in 2019 and 2020. In the combined analysis of variance, the CD–CH group showed a significant increase in seed yield, although the performance of this group was otherwise comparable or inferior to the adapted elite Canadian cultivars. Canadian cultivars were superior to both CD–CH and elite, exotic Chinese cultivars in seed oil and seed protein concentration. The yield potential of the exotic-derived soybean lines observed in this study provide a great source of novel genetics for soybean breeders interested in introgressing novel alleles from exotic sources to improve yield to help combat climate change.

Key words: Genome-wide Association Study (GWAS), exotic soybean germplasm, seed yield, yield improvement

Résumé

Malgré les préoccupations grandissantes que suscite l'amenuisement de la base génétique du soja [(*Glycine max* (L.) Merr.] en Amérique du Nord et les difficultés que cela pourrait poser en raison des changements climatiques et environnementaux dans le monde, diverses craintes empêchent toujours les obtenteurs de recourir plus au plasma germinal exotique. Les auteurs voulaient évaluer la diversité des 200 génotypes de soja rassemblés dans le cadre d'une étude d'association à la grandeur du génome (GWAS) pour le rendement grainier, la qualité de la graine et la performance des paramètres agronomiques. Le groupe de la GWAS consistait en lignées issues de plusieurs générations de croisements biparentaux entre des cultivars élites canadiens et élites chinois (CD-CH), des cultivars élites canadiens (CD) et des cultivars élites chinois exotiques (CH) évalués à Elora et à Woodstock (Ontario), en 2019 et en 2020. Selon l'analyse combinée de la variance, le groupe CD-CH a enregistré une hausse significative du rendement grainier, même si sa performance reste comparable ou est inférieure à celle des cultivars élites canadiens après acclimatation. La teneur en huile et en protéines des cultivars canadiens dépasse celle des cultivars CD-CH et des cultivars élites chinois exotiques. À cause de leur rendement potentiel, observé durant la présente étude, les lignées de soja issues des cultivars exotiques constituent une excellente source de gènes neufs pour les obtenteurs qui souhaiteraient ajouter des allèles d'origine exotique à leurs cultivars par introgression, en vue d'en accroître le rendement et de les rendre plus résilients au changement climatique. [Traduit par la Rédaction]

Mots-clés : étude d'association à la grandeur du génome (GWAS), plasma germinal exotique du soja, rendement grainier, hausse du rendement

Introduction

Although global soybean (*Glycine max* (L.) Merill.) production has grown over twofold in the past decade, climate change, shifting weather patterns, and growing global human population are expected to pose major challenges to developing environmentally resilient, high-yielding cultivars that are able to readily withstand the increasing biotic and abiotic stresses (Sneller et al. 2005; Fox et al. 2015; Wang et al. 2017; Kofsky et al. 2018; Gaire et al. 2020; Kilian et al. 2020). Due to the recurrent use of elite cultivars in breeding programs, genetic bottlenecks, and low diversity of founding populations, there is a growing concern regarding the

genetic variation found within breeding programs (Gizlice et al. 1993, 1994; Kisha et al. 1998; Cui et al. 2000; Sneller et al. 2005; Hyten et al. 2006; Iquira et al. 2010; Friedrichs et al. 2016). Low genetic diversity can result in diminished yield gains and leave the crop vulnerable to biotic and abiotic stresses (St. Martin and Asiam 1986; Bilyeu and Beuselinck 2005;). Exotic or under-utilized germplasm has emerged as a genetic reservoir that may hold potentially beneficial genes to help address these issues (Sneller et al. 2005; Fox et al. 2015; Wang et al. 2017; Kofsky et al. 2018; Gaire et al. 2020; Kilian et al. 2020). The following drawbacks are most cited as reasons for avoiding the use of exotic germplasm sources: yield-drag, excessive selection cycles, and time-cost constraints (; Bernardo 2009; Kim et al. 2012). Despite these challenges, there is much literature to support contributions of exotic germplasm (Palomeque et al. 2009a, 2009b; Kim et al. 2011; Rossi et al. 2013; Akpertey et al. 2014; Bellaloui et al. 2017), which highlighted the importance of this under-used germplasm source.

Whereas seed yield, disease resistance, and seed quality traits are often the main breeding objectives, other agronomic traits such as plant height, lodging, and seed weight are also important as they affect yield and can determine the end use of the crop (Zhe et al. 2010; Liang et al. 2016; Teng et al. 2017; Zatybekov et al. 2017; Zhang et al. 2018). The properties of soybean seed such as protein, essential amino acids, oil, metabolizable energy, flavonoids, and other secondary metabolites impart great commercial value to the crop (Willis 2003; Mittal et al. 2004; Patil et al. 2017; Sah 2018). With increasing evidence reported in the literature of the many benefits to human health and nutrition, as well as ongoing need to produce nutritious foods to combat world hunger and food shortage, the demand for soybean and soybean food products is predicted to grow significantly (Friedman and Brandon 2001; McCue and Shetty 2004; Mittal et al. 2004; Lule et al. 2015). Therefore, it becomes imperative to elucidate and thoroughly understand the genetic basis for soybean protein, essential amino acids, oil, metabolizable energy, flavonoids, and other secondary metabolites. Furthermore, the underlying phenotypic correlations among traits should also be investigated as the optimal physical and chemical characteristics of seed required vary for different food processors and manufacturers based on end use (Kinney and Clemente 2005; Whiting et al. 2020).

Soybean seed yield is a complex trait that is often positively correlated with seed weight in that high yield is often associated with late maturity, due to a longer seed-fill period, as well as plant height; however, these relationships have been shown to be environment or population dependent (Posadas et al. 2014; Contreras-Soto et al. 2017). Though simultaneous selection for both seed yield and seed-quality traits is most desirable, the negative correlations between some of these traits, polygenic or quantitative nature of the traits and considerable environmental influence, make the task challenging for breeders to improve the traits concurrently (Zhang et al. 2018).

Previous studies conducted on recombinant inbred soybean line (RIL) populations derived from crosses between exotic elite Chinese and adapted elite Canadian cultivars reported significant variation for seed yield (Palomeque et al. 2009a, 2009b; Rossi et al. 2013). Furthermore, these studies reported that a differential performance was observed among the RILs between Canadian and Chinese mega-environments (Palomeque et al. 2009a, 2009b; Rossi et al. 2013). As these studies were conducted only on RILs, the inclusion of a genomic-diversity panel may help to assess their performance against high-yielding, adapted cultivars to better estimate their agronomic and yield capacity. The objective of this study was to test a diverse panel of soybean cultivars consisting of advanced progeny lines derived from crosses between elite adapted Canadian × elite exotic Chinese cultivars against elite Canadian and Chinese cultivars to determine the potential of improving yield and seed quality traits using alleles from elite exotic germplasm from northeast China. Such an evaluation facilitates direct comparisons between the potential of the exotic-derived lines against high-yielding elite commercial cultivar representatives from both the Canadian and Chinese parental groups, thus facilitating better use of available germplasm sources for crop improvement (Lee et al. 2013; Thivierge et al. 2015; Stasko 2018; Torres et al. 2019). Building further upon the findings of Rossi et al. (2013) and Palomeque et al. (2009a, 2009b), we report results from agronomic evaluations of a diverse panel of 200 soybean lines conducted in 2019 and 2020 at Elora and Woodstock, ON, Canada.

Materials and methods

Plant material and experimental design

A genetic diversity panel consisting of 200 soybean genotypes of elite Canadian (CD) cultivars ($n = 59$), elite Chinese (CH) cultivars ($n = 53$), and Canadian × Chinese (CD–CH) progeny lines ($n = 88$) (Table S1) was evaluated in yield trials at the Elora Research Station (43°64′104.4″N; 80°40′567.4″W), Elora, ON, and Woodstock Research Station (43°08′44.8″N; 80°47′02.5″W), Woodstock, ON, during 2019 and 2020 field seasons. The diversity panel consisted of soybean cultivars developed by Agriculture and Agri-Food Canada (AAFC), Centre de recherche sur les grains (CÉROM), experimental breeding lines and cultivars from the University of Guelph Soybean Breeding Program, elite Chinese cultivars from Northeast China, as well as progeny derived from the Canadian × Chinese crosses were included. Two replications were evaluated per environment in a randomized complete block design (RCBD) with soybean lines randomly assigned.

Due to low number of seeds available and poor germination, 20 of the Chinese cultivars were grown as a seed increase panel at the Woodstock Research Station, Woodstock, ON, in 2019 (Table S1). In 2019, a total of 180 genotypes were evaluated for yield at the Elora Research Station, Elora, ON, in 2019, whereas 147 genotypes were evaluated at Woodstock, ON. For Elora, 250 seeds per plot were planted as two-row plots using 0.35 m between-row spacing for the final plot size of 5.5 m². The latter modified plot size (two-row vs. four-row plot) was used in Elora to accommodate the low seed numbers that were available for some of the Chinese cultivars and make the plot size equal for all entries in the trial. The Woodstock 2019 and 2020 field trial, as well as Elora trial in 2020,

were grown as regular four-row plots of plot size of 8.25 m² with 0.35 m row spacing. Due to an edging error, the plot size in Woodstock, ON, in 2020 was 7.5 m².

In both years, seeds for each genotype were screened to remove off-types as needed, packaged by field trial and arranged in the planting order. Hilum colour was recorded and used to cross-reference between records to ensure uniformity of seeds.

In 2019, yield trials were planted on 11 June at Elora, ON, and harvested on 23 October. Woodstock yield trials were planted on 14 June 2019 and harvested on 23 November 2019. In 2020, the yield trials were planted on 22 May at Elora and on 26 May at Woodstock, ON, and harvested on 6 November, at Elora, ON, and on 8 October at Woodstock, ON. Planting was done using a precision Wintersteiger research planter. Harvesting was conducted with Almaco (Nevada, IA, USA) and Wintersteiger (Ried im Innkreis, Austria) research plot combines. Due to combine error, two plots were lost in Elora in 2020 (Table S2). Yield was recorded as kg/ha and adjusted to 13% moisture.

Phenotypic data collection

Seedling emergence score was recorded for each plot 3 weeks after planting, based on the number of plants observed. A scale of 0–10 was used where 0 corresponded to no emergence and 10 to 100% emergence. Pubescence colour, flower colour, and leaf morphology were recorded subsequently. Flower colour was recorded at the R1 stage (one flower at any node) and full maturity date was recorded at R8 stage, whereas 95%–100% of pods turned brown and lodging was scored at maturity using a scale ranging from 1 to 5, where 1 means plants fully upright and 5 means plants fully prostrate. Plant height was measured as the distance between the terminal node and the ground, measured in cm (Ernpig and Fehr 1971).

After harvest, seeds were stored in cold storage. Seed quality traits were measured using Perten diode array 7250 near infrared spectroscopy (Springfield, USA) machine and following the manufacturer's guidelines. Seeds were screened to remove off-types, dirt, and other impurities. 100-seed weight was measured with a regular commercial scale. Randomly selected genotypes were re-run to ensure that the readings were consistent. Within the soybean seed-quality traits, only the protein and oil concentration (expressed as % on a dry seed basis), and 100-seed weight (g) were retained for analysis. One entry each from Elora 2019 and Woodstock 2019 was removed from analysis due to machine error (Table S2).

Statistical analysis

Analysis of variance (ANOVA) of seed yield, seed quality, and agronomic traits were conducted using the PROC GLIMMIX procedure in Statistical Analysis Systems (SAS) version 9.4 (SAS Institute Inc., Cary, NC, USA) for RCBD, with nearest neighbour covariance adjustment. "Genotype", "environment", and "genotype-by-environment" were considered fixed effects and "block (environment)" was considered random effect. Due to the unbalanced number of genotypes in 2019 at Elora and Woodstock, data were sorted by year and

environment and separated into three sets: 2019 (with 147 genotypes), 2020 (200 genotypes), and combined years (147 genotypes). The least squares mean (LSMEANS) values were calculated for seed yield, protein concentration, and oil concentration for both combined environments and individual environments. PROC UNIVARIATE was utilized to conduct Shapiro–Wilk tests to determine the distribution of residuals, whereas PROC PLOT was used to examine the normality of residual distribution. Homogeneity of error variance was tested by conducting Levene's test on the absolute residuals.

Tukey–Kramer multiple comparison test was invoked along with the LINES statement to generate statistically significant differences between comparison groups with comparisons made between environments, genotypes, and genotypes by environments. CONTRAST and ESTMATE statements were used to test the statistical differences, if any, between CD, CH, and CD–CH groups. Woodstock 2019 was not included in this analysis since Chinese cultivars were excluded from the yield trials that year, due to low germination and (or) poor seed availability.

Pearson's correlation coefficients were calculated with PROC CORR to establish the magnitude, direction, and statistical strength of relationships between seed yield, seed-quality traits, and agronomic traits. Correlation analysis was also conducted for the four different environments for seed yield, seed protein content, and seed oil content. The type I error (α) was set at 0.05.

Results

Seed Yield (kg/ha), seed protein concentration (%, dry basis), and seed oil concentration (%, dry basis) between the genotypic groups

The range of yield observed across all locations was 126–4985 kg/ha (Table 1). Both genotype and genotype × environment interactions were shown to be significant (Table 2). The CD–CH progeny lines outperformed both CD and CH cultivars in terms of seed yield in combined environment analysis (Table 3) with a mean yield and standard error of 2582 ± 75.0 kg/ha, whereas Canadian cultivars yielded 2289 ± 76.8 kg/ha and Chinese cultivars yielded 2263 ± 79.5 kg/ha (Table 3).

The range for protein concentration across locations was 34.0%–46.8% (dry basis) (Table 1). For this trait, there was no difference observed between Canadian genotypic group and CD–CH genotypic group (Table 3). However, there was a significant difference between these two groups and the Chinese genotypes, with the latter group containing the lowest protein concentration at 40.6 ± 0.12% (se). Canadian genotypic group had the highest protein concentration at 41.2% to 1 ± 0.11%, with CD–CH group containing 41.1 ± 0.09% (Table 3).

The mean seed oil concentration differed significantly between the three genotypic groups (Table 3). Seed oil concentration across locations varied from 14.9% to 23.1% (dry basis) (Table 1). Canadian genotypic group performed the best of the three for this trait with a seed oil concentration of 20.4% ± 0.07% (dry basis), followed by the CD–CH genotypic

Table 1. Summary of soybean seed yield (kg/ha), seed protein, and seed oil concentration for the four environments: Elora 2019, Elora 2020, Woodstock 2019, and Woodstock 2020.

Trait	Environment	N	Mean (kg/ha)	Std dev	Range
Seed yield	Elora 2019	359	2262	654.5	348–3931
	Elora 2020	398	2723	507.2	1248–4334
	Woodstock 2019	293	3219	671.3	126–4805
	Woodstock 2020	400	2306	678.8	826–3900
Protein	Elora 2019	359	42.2	1.64	36.1–46.3
	Elora 2020	398	41.0	1.38	36.9–44.9
	Woodstock 2019	293	41.1	2.11	35.5–46.8
	Woodstock 2020	400	39.8	1.88	34.0–45.2
Oil	Elora 2019	359	19.0	1.18	14.9–21.4
	Elora 2020	398	20.0	1.14	16.1–22.8
	Woodstock 2019	293	19.2	0.80	17.2–21.4
	Woodstock 2020	400	20.7	0.98	17.7–23.1

Table 2. Type III tests of fixed effects for seed yield (kg/ha), seed protein concentration (%, dry basis), and seed oil concentration (%, dry basis) for combined environment analysis for Elora, ON, and Woodstock, ON, for 2019 and 2020.

Trait	Effect	Num DF	Den DF	F Value	$Pr > F$
Soybean seed yield	Genotype	146	584	4.14	<0.0001
	Environment	3	4.005	14.62	0.0127
	Genotype × environment	438	584	1.39	<0.0001
Seed protein	Genotype	146	584	14.92	<0.0001
	Environment	3	3.995	87.15	0.0004
	Genotype × environment	438	584	1.8	<0.0001
Seed oil	Genotype	146	584	11.81	<0.0001
	Environment	3	3.999	54.67	0.0011
	Genotype × environment	438	584	1.57	<0.0001

Table 3. Least square means for soybean seed yield (kg/ha), seed protein, and seed oil concentration (dry basis, %) and test of significance of differences for the different genotypic groups of the soybean panel across combined environment analysis for trials conducted at Elora, ON, and Woodstock, ON, for 2019 and 2020.

Trait	Genotypic group*	Estimate (kg/ha)	Standard error	T-K LSM[†]
Seed Yield	Canadian	2289	76.8	B
	Chinese	2263	79.5	B
	CD–CH	2582	75	A
Protein	Canadian	41.2	0.11	A
	Chinese	40.6	0.12	B
	CD–CH	41.1	0.09	A
Oil	Canadian	20.4	0.07	A
	Chinese	19.1	0.08	C
	CD–CH	19.9	0.06	B

*The constitution of the GWAS panel was such that the number of soybean lines within each genotypic group is as follows: Canadian ($n = 59$), Chinese ($n = 53$), and CD–CH ($n = 88$).
[†]Tukey–Kramer Grouping for environment least squares means ($\alpha = 0.05$). LSM followed by the same letter are not significantly different.

group at 19.9% ± 0.06%, with the Chinese genotypic group bearing the lowest oil content at 19.1 ± 0.08% (Table 3).

Correlation analysis between seed yield, seed-quality traits, and agronomic traits

Correlations were calculated to evaluate the relationships between seed yield, seed-quality traits, and select agronomic traits (Table 4). The following traits were evaluated: seed yield, protein concentration, oil concentration, seed weight (g), height (cm), days to maturity, and lodging score (1–5).

Yield was positively correlated with height ($r = 0.47$; $p < 0.0001$) and lodging score ($r = 0.28$; $p < 0.0001$). Yield and oil concentration ($r = -0.12$; $p < 0.0001$), as well as yield and protein concentration ($r = -0.08$; $p = 0.0035$) showed significant negative correlations. Protein concentration was negatively correlated with yield ($r = -0.08$; $p = 0.0035$), oil concentration ($r = -0.40$; $p < 0.0001$), height ($r = -0.23$; $p < 0.0001$), and days to maturity ($r = -0.15$; $p < 0.0001$). A significant positive relationship was observed between protein concentration and seed weight ($r = 0.18$; $p < 0.0001$).

Oil concentration showed significant negative relationships with seed yield ($r = -0.12$; $p < 0.0001$), protein content ($r = -0.40$; $p < 0.0001$), height ($r = -0.05$; $p = 0.0411$), seed weight ($r = -0.12$; $p < 0.0001$), days to maturity ($r = -0.26$; $p < 0.0001$) and lodging score ($r = -0.24$; $p < 0.0001$).

Discussion

Soybean seed yield and seed-quality traits continue to be of economic importance and are the primary targets for improvement by breeders. However, due to the recurrent use of elite cultivars in breeding programs, coupled with genetic bottlenecks and low diversity of founding populations, there is growing concern regarding the genetic variation found within breeding programs (Gizlice et al. 1993, 1994; Kisha et al. 1998; Cui et al. 2000; Sneller et al. 2005; Hyten et al. 2006; Iquira et al. 2010; Friedrichs et al. 2016). Low genetic diversity can result in diminished yield gains and leave the crop vulnerable to biotic and abiotic stresses (St. Martin and Asiam 1986; Bilyeu and Beuselinck 2005).

The use of exotic germplasm has been highlighted as a potential solution to this growing concern and has been done with success in other plant species (Ehlers and Foster 1993; Concibido et al. 2003; Sneller et al. 2005; Bernardo 2009; Palomeque et al. 2009a; Diers et al. 2018; Gaire et al. 2020). Exotic germplasm is defined as germplasm that is unadapted to the breeder's target environment and can be a great source of novel genetic variation that can potentially contribute resilience to both biotic and abiotic stresses (Hallauer 2007; Mickelbart et al. 2015; Hoffmann et al. 2018). Despite the challenges associated with the use of exotic germplasm, such as yield drag, deleterious alleles, and the need for excessive backcrossing, there is growing evidence to support the use of exotic germplasm (Concibido et al. 2003; Kabelka et al. 2004; Guzman et al. 2007; Palomeque et al. 2009a; Palomeque et a. 2019b; Rossi et al. 2013; Akpertey et al. 2014; Hegstad et al. 2019).

Since China is the centre of origin of soybean and adapted elite Chinese cultivars may potentially contribute beneficial alleles and help improve seed yield, both Palomeque et al. (2010) and Rossi et al. (2013) evaluated populations derived from Canadian × Chinese crosses to evaluate the behaviour and inheritance of seed yield quantative trait loci (QTL) in different environments. Building further upon previous works mentioned above, this current study aimed to evaluate a diverse panel of 200 soybean genotypes that consisted of University of Guelph released cultivars and experimental breeding lines, AAFC cultivars, elite Chinese cultivars, and Canadian × Chinese progeny lines derived from different crosses. It was observed that the exotic elite Chinese cultivars had the lowest seed yield compared to the adapted Canadian ones (Table 3). This finding is in line with other studies that reported on evaluations of RIL populations with exotic parentage vs adapted elite cultivars (St. Martin and Asiam 1986; Palomeque et al. 2010; Rossi et al. 2013). Canadian × Chinese RILs, on the other hand, outperformed both the exotic Chinese and adapted elite Canadian cultivars (Table 3). Hegstad et al. (2019) reported on a study that evaluated the performance of exotic germplasm-derived soybean lines against the performance of adapted elite cultivars. They reported that across three field years, the highest yielding line yielded 280 kg/ha more than the adapted elite parent (Hagstad et al. 2019). The average yield difference observed in this study between the Canadian genotypic group and the Canadian × Chinese derived genotypic group was similar to that of Hagstad et al. (2019), with a difference of 293 kg/ha. Kim et al. (2012) reported on a study that included a Chinese cultivar (Tiefeng No. 8), which was used as a donor parent with repeat backcrossing. They reported that a yield increase was observed in at least one of the populations being evaluated (Kim et al. 2012). Diers et al. (2018) reported that in their evaluation of different soybean lines, the exotic founder germplasm lines yielded the lowest compared to adapted parent lines, which is similar to the current observation of exotic elite Chinese cultivars. None of these studies, however, reported on evaluating exotic-derived germplasm against a diverse group of soybean germplasm to assess seed yield, seed quality traits, and agronomic traits. Moreover, Posadas et al. (2014) reported an average yield gain of 567 kg/ha with each cycle of recurrent selection in a population developed from seven unadapted plant introductions. Their study only included the evaluation of a single population derived from three Japanese, three South Korean, and one Chinese plant introduction (Posadas et al. 2014). To the best of our knowledge, the current study is the first to directly evaluate Canadian × Chinese progeny in a diverse pool of soybean lines. The abovementioned studies together with ours provide further support toward the potential for contributing beneficial alleles from exotic germplasm sources for continued crop improvement.

Though soybean seed oil and protein are two major focal points for crop improvement for breeders, the complicated and often quantitative nature of these traits make direct improvement quite challenging. Previous studies have identified similar relationships to those observed in the current study (Table 4) between seed yield, seed quality traits, and agronomic traits (Palomeque et al. 2009a, 2010; Rossi

Table 4. Pearson correlation coefficients for soybean seed yield, seed oil concentration, seed protein concentration, and agronomic traits for a panel of 200 soybean lines evaluated from 2019 to 2020 at Elora, ON, and Woodstock, ON.

	Protein	Oil	Seed weight	Height	Days to maturity	Lodging
Yield	−0.08 0.0035[†]	−0.12 < 0.0001[*]	0.47 < 0.0001[*]	0.03 0.2918	−0.003 0.8986	0.28 < 0.0001[*]
Protein		−0.40 < 0.0001[*]	−0.23 < 0.0001[*]	0.18 < 0.0001[*]	−0.15 < 0.0001[*]	0.010.6314
Oil			−0.12 < 0.0001[*]	−0.05 0.0411[°]	−0.26 < 0.0001[*]	−0.24 < 0.0001[*]
Seed weight				−0.040.1776	0.47 < 0.0001[*]	0.26 < 0.0001[*]
Height					0.010.6361	−0.04 0.1556
Days to maturity						−0.08 0.0023[°]

[*]Significant at 0.0001.
[†]Significant at 0.001.
[°]Significant at 0.01.

et al. 2013; Contreras-Soto et al. 2017). Contreras-Soto et al. (2017) reported that soybean seed yield, plant height, and days to maturity showed positive correlations, with a general trend of later maturity and taller plants were associated with higher yield. Additionally, Posadas et al. (2014) reported that the magnitude, direction, and significance for plant height, lodging, and seed yield correlations were population dependent.

A significant negative correlation was observed between seed oil and protein concentration (Table 4), which was consistent with previous reports (Hyten et al. 2014; Ficht et al. 2022; Lee et al. 2019). Patil et al. (2018) also reported similar results to those reported herewith regards to the relationships between seed yield and seed-quality traits and concluded that soybean seed composition is significantly affected by the environment. This sentiment is consistent with reports elsewhere in literature (Zhe et al. 2010). It is possible that this was the case in the current study; therefore, expanded environments and further testing will be needed to establish the precise nature of these relationships.

Conclusion

In conclusion, the findings of this study are consistent with previous reports and show the promising potential of using exotic soybean germplasm, specifically from China, in improving yield potential and seed-quality traits in Canada and potentially northern USA. Further studies across multiple years and environments will be needed to verify the genotypic performance of the exotic-derived lines and to rule out environmental or other effects that may have affected seed yield and seed-quality traits during the period of the study. Furthermore, since the first year of field evaluations did not include all 200 cultivars at each environment, inclusion of the full panel across all years and environments would help overcome limitations of an uneven sample size experienced in the current study.

Acknowledgements

Article information

Copyright

Data availability

All data used in this study are stored on a OneDrive repository at the University of Guelph and may be available from the corresponding author upon request.

Author information

Author ORCIDs

Istvan Rajcan https://orcid.org/0000-0001-5156-2482

Author contributions

IR conceptualized the study, received funding and provided resources to carry it out, contributed to the experimental design, reviewed, edited, and helped interpret the data. CP co-designed the experiments, carried them out, collected and analysed the data, wrote the manuscript, and edited it to produce the final version.

Competing interests

The authors declare there are no competing interests.

Funding information

Supplementary material

Supplementary data are available with the article at https://doi.org/10.1139/cjps-2022-0016.

References

Akpertey, A., Belaffif, M., Graef, G.L., Rouf Mian, M.A., Grover Shannon, J. Cregan, P.B., et al. 2014. Effects of selective genetic introgression from wild soybean to soybean. Crop Sci. **54**: 2683–2695. doi:10.2135/cropsci2014.03.0189.

Bellaloui, N., Smith, J.R., Mengistu, A., Ray, J.D., and Gillen, A.M. 2017. Evaluation of exotically-derived soybean breeding lines for seed yield, germination, damage, and composition under dryland production in the midsouthern USA. Front. Plant Sci. **8**: 1–20. doi:10.3389/fpls.2017.00176. PMID: 28220127.

Bernardo, R. 2009. Genomewide selection for rapid introgression of exotic germplasm in maize. Crop Sci. **49**: 419–425. doi:10.2135/cropsci2008.08.0452.

Bilyeu, K.D., and Beuselinck, P.R. 2005. Genetic divergence between North American ancestral soybean lines and introductions with resistance to soybean cyst nematode revealed by chloroplast haplotype. J. Hered. **96**: 593–599. doi:10.1093/jhered/esi087. PMID: 15947084.

Concibido, V.C., La Vallee, B., Mclaird, P., Pineda, N., Meyer, J. Hummel, L., et al. 2003. Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. Theor. Appl. Genet. **106**: 575–582. doi:10.1007/s00122-002-1071-5. PMID: 12595984.

Contreras-Soto, R.I., Mora, F., de Oliveira, M.A.R., Higashi, W., Scapim, C.A., and Schuster, I. 2017. A genome-wide association study for agronomic traits in soybean using SNP markers and SNP-based haplotype analysis. PLoS ONE **12**: 1–22. doi:10.1371/journal.pone.0171105.

Cui, Z., Thomas, E., Carter, J., and Burton, J. 2000. Genetic diversity patterns in North American public soybean cultivars based on coefficient of parentage. Crop Sci. **40**: 1780–1793. doi:10.2135/cropsci1996.0011183X003600030038x.

Diers, B.W., Specht, J., Rainey, K.M., Cregan, P., Song, Q. Ramasubramanian, V., et al. 2018. Genetic architecture of soybean yield and agronomic traits. G3 Genes|Genomes|Genetics **8**: 3367–3375. doi:10.1534/g3.118.200332.

Ehlers, J.D., and Foster, K.W. 1993. Introgression of agronomic characters from exotic cowpea germplasm into blackeye bean. F. Crop. Res. **35**: 43–50. doi:10.1016/0378-4290(93)90135-A.

Ernpig, L.T., and Fehr, W.R. 1971. Evaluation of methods for generation advance in bulk hybrid soybean populations. Crop Sci. **11**: 51–54.

Ficht, A., Bruce, R., Torkamaneh, D., Grainger, C.M., Eskandari, M., and Rajcan, I. 2022. Genetic analysis of sucrose concentration in soybean seeds using a historical soybean genomic panel. Theor. Appl. Genet. **135**: 1375–1383. doi: https://doi.org/10.1007/s00122-022-04040-z.

Fox, C.M., Cary, T., Nelson, R., and Diers, B.W. 2015. Confirmation of a seed yield QTL in soybean. Crop Sci. **55**: 992–998. doi:10.2135/cropsci2014.10.0688.

Friedman, M., and Brandon, D.L. 2001. Nutritional and health benefits of soy proteins. J. Agric. Food Chem. **49**: 1069–1086. doi:10.1021/jf0009246. PMID: 11312815.

Friedrichs, M.R., Burton, J.W., and Brownie, C. 2016. Heterosis and genetic variance in soybean recombinant inbred line populations. Crop Sci. **56**: 2072–2079. doi:10.2135/cropsci2015.11.0702.

Gaire, R., Ohm, H., Brown-Guedira, G., and Mohammadi, M. 2020. Identification of regions under selection and loci controlling agronomic traits in a soft red winter wheat population. Plant Genome, **13**: 1–18. doi:10.1002/tpg2.20031.

Gizlice, Z., Carter, T.E., and Burton, J.W. 1993. Genetic diversity in North American soybean: I. Multivariate analysis of founding stock and

relation to coefficient of parentage. Crop Sci. **33**: 614. doi:10.2135/cropsci1993.0011183X003300030038x.

Gizlice, Z., Carter, T.E., and Burton, J.W. 1994. Genetic base for North American public soybean cultivars released between 1947 and 1988. Crop Sci. **34**: 1143–1151. doi:10.2135/cropsci1994.0011183X003400050001x.

Guzman, P.S., Diers, B.W., Neece, D.J., St. Martin, S.K., LeRoy, A.R. Grau, C.R., et al. 2007. QTL associated with yield in three backcross-derived populations of soybean. Crop Sci. **47**: 111–122. doi:10.2135/cropsci2006.01.0003.

Hallauer, A.R. 2007. History, contribution, and future of quantitative genetics in plant breeding: lessons from maize. Crop Sci. **47**: S4–S19. doi:10.2135/cropsci2007.04.0002IPBS.

Hegstad, J.M., Nelson, R.L., Renny-Byfield, S., Feng, L., and Chaky, J.M. 2019. Introgression of novel genetic diversity to improve soybean yield. Theor. Appl. Genet. **132**: 2541–2552. doi:10.1007/s00122-019-03369-2. PMID: 31209537.

Hoffmann, L., and Rooney. 2018. Sorghum improvement for yield. *In* Sorghum: a state of the art and future perspectives, Vol. 58. *Edited by* W.L. Ciampitti I. and V. Prasad. Wiley, Hoboken, NJ. pp. 1–16. doi:10.2134/agronmonogr58.2014.0055.

Hyten, D.L., Song, Q., Zhu, Y., Choi, I.Y., Nelson, R.L. Costa, J.M., et al. 2006. Impact of genetic bottlenecks on soybean genome diversity. Proc. Natl. Acad. Sci. U.S.A. **103**: 16666–16671. doi:10.1073/pnas.0604379103. PMID: 17068128.

Hyten, D.L., Cregan, P.B., Specht, J.E., Hwang, E.-Y., Song, Q., Jia, G., and Costa, J. 2014. A genome-wide association study of seed protein and oil content in soybean. BMC Genomics, **15**: 1. doi:10.1186/1471-2164-15-1. PMID: 24382143.

Iquira, E., Gagnon, E., and Belzile, F. 2010. Comparison of genetic diversity between Canadian adapted genotypes and exotic germplasm of soybean. Genome, **53**: 337–345. doi:10.1139/G10-009. PMID: 20616865.

Kabelka, E.A., Diers, B.W., Fehr, W.R., LeRoy, A.R., Baianu, I.C. You, T., et al. 2004. Putative alleles for increased yield from soybean plant introductions. Crop Sci. **44**: 784. doi:10.2135/cropsci2004.7840.

Kilian, B., Dempewolf, H., Guarino, L., Werner, P., Coyne, C., and Warburton, M.L. 2020. Crop science special issue: adapting agriculture to climate change: a walk on the wild side. Crop Sci. **61**: 32–36. doi:10.1002/csc2.20418.

Kim, K.S., Diers, B.W., Hyten, D.L., Rouf Mian, M.A., Shannon, J.G., and Nelson, R.L. 2012. Identification of positive yield QTL alleles from exotic soybean germplasm in two backcross populations. Theor. Appl. Genet. **125**: 1353–1369. doi:10.1007/s00122-012-1944-1. PMID: 22869284.

Kim, M., Hyten, D.L., Niblack, T.L., and Diers, B.W. 2011. Stacking resistance alleles from wild and domestic soybean sources improves soybean cyst nematode resistance. Crop Sci. **51**: 934–943. doi:10.2135/cropsci2010.08.0459.

Kinney, A.J., and Clemente, T.E. 2005. Modifying soybean oil for enhanced performance in biodiesel blends. Fuel Process. Technol. **86**: 1137–1147. doi:10.1016/j.fuproc.2004.11.008.

Kisha, T.J., Diers, B.W., Hoyt, J.M., and Sneller, C.H. 1998. Genetic diversity among soybean plant introductions and North American germplasm. Crop Sci. **38**: 1669–1680. doi:10.2135/cropsci1998.0011183X003800060042x.

Kofsky, J., Zhang, H., and Song, B.H. 2018. The untapped genetic reservoir: the past, current, and future applications of the wild soybean (*Glycine soja*). Front. Plant Sci. **9**: 949. doi:10.3389/fpls.2018.00949. PMID: 30038633.

Lee, S., Van, K., Sung, M., Nelson, R., LaMantia, J., McHale, L.K., and Mian, M.A.R. 2019. Genome-wide association study of seed protein, oil and amino acid contents in soybean from maturity groups I to IV. Theor. Appl. Genet. **132**: 1639–1659. doi:10.1007/s00122-019-03304-5. PMID: 30806741.

Lee, W.K., Kim, N., Kim, J., Moon, J.K., Jeong, N. Choi, I.Y., et al. 2013. Dynamic genetic features of chromosomes revealed by comparison of soybean genetic and sequence-based physical maps. Theor. Appl. Genet. **126**: 1103–1119. doi:10.1007/s00122-012-2039-8. PMID: 23306355.

Liang, H., Xu, L., Yu, Y., Yang, H., Dong, W., and Zhang, H. 2016. Identification of QTLs with main, epistatic and QTL by environment interaction effects for seed shape and hundred-seed weight in soybean across

multiple years. J. Genet. **95**: 475–477. doi:10.1007/s12041-016-0648-8. PMID: 27350695.

Lule, V.K., Garg, S., Pophaly, S.D., Hitesh, and Tomar, S.K. 2015. Potential health benefits of lunasin: a multifaceted soy-derived bioactive peptide. J. Food Sci. **80**: C485–C494. doi:10.1111/1750-3841.12786.

St. Martin, S.K., and Asiam, M. 1986. Performance of progeny of adapted and plant introduction soybean lines. 1. Crop Sci. **26**: 753. doi:10.2135/cropsci1986.0011183X002600040026x.

McCue, P., and Shetty, K. 2004. Health benefits of soy isoflavonoids and strategies for enhancement: a review. Crit. Rev. Food Sci. Nutr. **44**: 361–367. doi:10.1080/10408690490509591. PMID: 15540649.

Mickelbart, M.V., Hasegawa, P.M., and Bailey-Serres, J. 2015. Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. Nat. Rev. Genet. **16**: 237–251. doi:10.1038/nrg3901.

Mittal, G., Jiang, Y., Kakuda, Y., Yeung, D., Arunasalam, K., and Shi, J. 2004. Saponins from edible legumes: chemistry, processing, and health benefits. J. Med. Food **7**: 67–78. doi:10.1089/109662004322984734. PMID: 15117556.

Palomeque, L., Li-Jun, L., Li, W., Hedges, B., Cober, E.R., and Rajcan, I. 2009*a*. QTL in mega-environments: II. Agronomic trait QTL co-localized with seed yield QTL detected in a population derived from a cross of high-yielding adapted × high-yielding exotic soybean lines. Theor. Appl. Genet. **119**: 429–436. doi:10.1007/s00122-009-1048-8.

Palomeque, L., Li-Jun, L., Li, W., Hedges, B., Cober, E.R., and Rajcan, I. 2009*b*. QTL in mega-environments: I. Universal and specific seed yield QTL detected in a population derived from a cross of high-yielding adapted × high-yielding exotic soybean lines. Theor. Appl. Genet. **119**: 417–427. doi:10.1007 /s00122-009-1049-7.

Palomeque, L., Liu, L.J., Li, W., Hedges, B.R., Cober, E.R. Smid, M.P., et al. 2010. Validation of mega-environment universal and specific QTL associated with seed yield and agronomic traits in soybeans. Theor. Appl. Genet. **120**: 997–1003. doi:10.1007/s00122-009-1227-7. PMID: 20012262.

Patil, G., Mian, R., Vuong, T., Pantalone, V., Song, Q. Chen, P., 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. PMID: 28801731.

Patil, G., Vuong, T.D., Kale, S., Valliyodan, B., Deshmukh, R. Zhu, C., et al. 2018. Dissecting genomic hotspots underlying seed protein, oil, and sucrose content in an interspecific mapping population of soybean using high-density linkage mapping. Plant Biotechnol. J. **16**: 1939–1953. doi:10.1111/pbi.12929. PMID: 29618164.

Posadas, L.G., Eskridge, K.M., and Graef, G.L. 2014. Elite performance for grain yield from unadapted exotic soybean germplasm in three cycles of a recurrent selection experiment. Crop Sci. **54**: 2536–2546. doi:10.2135/cropsci2014.01.0090.

Rossi, M.E., Orf, J.H., Liu, L.J., Dong, Z., and Rajcan, I. 2013. Genetic basis of soybean adaptation to North American vs. Asian mega-environments in two independent populations from Canadian × Chinese crosses. Theor. Appl. Genet. **126**: 1809–1823. doi:10.1007/s00122-013-2094-9. PMID: 23595202.

Sah, S. 2018. Functional role of protein kinases and phosphatase in abiotic stress response in plants. Doctoral Dissertation. Mississippi State University.

Sneller, C.H., Nelson, R.L., Carter, T.E., and Cui, Z. 2005. Genetic diversity in crop improvement. J. Crop Improv. **14**: 103–144. doi:10.1300/J411v14n01_06.

Stasko, A.K. 2018. Functional Gene Analysis of Resistance QTL towards Phytophthora sojae on Soybean Chromosome 19.

Teng, W., Feng, L., Li, W., Wu, D., Zhao, X., ,Han, Y., and Li, W. 2017. Dissection of the genetic architecture for soybean seed weight across multiple environments. Crop Pasture Sci. **68**: 358–365. doi:10.1071/CP16462.

Thivierge, M.N., Chantigny, M.H., Bélanger, G., Seguin, P., Bertrand, A., and Vanasse, A. 2015. Response to nitrogen of sweet pearl millet and sweet sorghum grown for ethanol in eastern Canada. BioEnergy Res. **8**: 807–820. doi:10.1007/s12155-014-9558-x.

Torres, L.G., Caixeta, D.G., Rezende, W.M., Schuster, A., Azevedo, C.F., Silva, F.F., and DeLima, R.O. 2019. Genotypic variation and relationships among traits for root morphology in a panel of tropical maize inbred lines under contrasting nitrogen levels. Euphytica, **215**: 51. doi:10.1007/s10681-019-2373-x.

Wang, C., Hu, S., Gardner, C., and Lübberstedt, T. 2017. Emerging avenues for utilization of exotic germplasm. Trends Plant Sci. **22**: 624–637. doi:10.1016/j.tplants.2017.04.002. PMID: 28476651.

Whiting, R.M., Torabi, S., Lukens, L., and Eskandari, M. 2020. Genomic regions associated with important seed quality traits in food-grade soybeans. BMC Plant Biol. **20**: 1–14. doi:10.1186/s12870-020-02681-0. PMID: 31898482.

Willis, S. 2003. The use of soybean meal and full fat soybean meal by the animal feed industry. *In* 12th Australian Soybean Conference. pp. 1–8.

Zatybekov, A., Abugalieva, S., Didorenko, S., Gerasimova, Y., Sidorik, I., Anuarbek, S., and Turuspekov, Y. 2017. GWAS of agronomic traits in soybean collection included in breeding pool in Kazakhstan. BMC Plant Biol. **17**: 179. doi:10.1186/s12870-017-1125-0. PMID: 29143671.

Zhang, K., Liu, S., Li, W., Liu, S., Li, X. Fang, Y., et al. 2018. Identification of QTNs controlling seed protein content in soybean using multi-locus genome-wide association studies. Front. Plant Sci. **871**: 1–14. doi:10.3389/fpls.2018.01690.

Zhe, Y., Lauer, J.G., Borges, R., and de Leon, N. 2010. Effects of genotype × environment interaction on agronomic traits in soybean. Crop Sci. **50**: 696–702. doi:10.2135/cropsci2008.12.0742.