

## **Correlations between soybean seed quality traits using a genome-wide association study panel grown in Canadian and Ukrainian mega-environments**

Authors: Hong, Huilin, Yoosefzadeh-Najafabadi, Mohsen, and Rajcan, Istvan

Source: Canadian Journal of Plant Science, 102(5) : 1040-1052

Published By: Canadian Science Publishing

URL: <https://doi.org/10.1139/cjps-2022-0046>

---

BioOne Complete ([complete.BioOne.org](https://complete.BioOne.org)) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at [www.bioone.org/terms-of-use](https://www.bioone.org/terms-of-use).

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

---

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

# Correlations between soybean seed quality traits using a genome-wide association study panel grown in Canadian and Ukrainian mega-environments

Huilin Hong, Mohsen Yoosefzadeh-Najafabadi, and Istvan Rajcan 

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

Corresponding author: Istvan Rajcan (email: [irajcan@uoguelph.ca](mailto:irajcan@uoguelph.ca))

## Abstract

Improvement of soybean [*Glycine max* (L.) Merr.] seed quality traits in addition to agronomic traits requires a detailed understanding of correlations between these traits. The objective of this study was to determine the correlations between seed compositions in soybeans grown in Canadian and Ukrainian mega-environments (MEs). The correlations between seed quality traits and agronomic traits were also studied. A genome-wide association study panel consisting of 184 soybean accessions was used for the study. The panel was grown in three Ontario field locations and two Ukrainian locations for 2 years, from 2018 to 2019. A total of 18 traits were measured and analyzed. The Pearson's correlation coefficients ( $r$ ) were calculated, and the genotype-by-trait biplots were generated to analyze the linear correlations between the traits. The well-documented negative correlations between protein and oil, as well as oil and the amino acids Lys, Cys, Met, and Thr, were confirmed. In addition, a positive correlation was observed between stearic acid and palmitic acid, while linolenic acid and oleic acid concentrations were negatively correlated. Sucrose was positively correlated with linolenic acid and raffinose and negatively with protein and the four amino acids. Most of the agronomic traits had positive correlations with each other, while there was no strong linear association detected between agronomic traits and the seed quality traits in either ME. The results of this study suggest that improvement of yield and other agronomic traits through breeding may be possible in both Canada and Ukraine without affecting the important seed quality traits.

**Key words:** agronomic traits, mega-environment, Pearson's correlation coefficient, seed quality traits, soybean breeding

## Résumé

Améliorer les paramètres qualitatifs de la graine du soja [*Glycine max* (L.) Merr.] en plus des paramètres agronomiques exige qu'on comprenne en profondeur les corrélations qui existent entre eux. Les auteurs souhaitaient établir les corrélations entre la composition de la graine du soja cultivé dans le méga-environnement canadien ou ukrainien. Parallèlement, ils ont étudié les corrélations entre les paramètres qualitatifs de la graine et les paramètres agronomiques. Le groupe de l'étude d'association à la grandeur du génome (GWAS) comprenait 184 obtentions. Ces dernières ont été cultivées pendant deux ans (2018 et 2019) à trois endroits, en Ontario, et à deux autres, en Ukraine. Les auteurs ont quantifié et analysé dix-huit caractères. Ensuite, ils ont calculé le coefficient de corrélation de Pearson ( $r$ ) et produit des diagrammes de double projection par génotype/caractère pour analyser les corrélations linéaires entre les caractères. Cette étude confirme la corrélation négative, déjà bien documentée, qui existe entre la protéine et l'huile ainsi qu'entre l'huile et les acides aminés Lys, Cys, Met et Thr. Par ailleurs, les auteurs ont noté une corrélation positive entre l'acide stéarique et l'acide palmitique, de même qu'une corrélation négative entre la concentration d'acide linoléique et celle d'acide oléique. Le sucrose est positivement corrélé à l'acide linoléique et au raffinose, et négativement avec la protéine et les quatre acides aminés précités. La plupart des caractères agronomiques sont positivement corrélés les uns avec les autres, mais on n'a décelé aucune association linéaire robuste entre les paramètres agronomiques et les paramètres qualitatifs de la graine dans les deux méga environnements. Les résultats de l'étude laissent croire qu'on pourrait améliorer le rendement et d'autres caractères agronomiques par l'hybridation au Canada comme en Ukraine, sans que les importants caractères qualitatifs de la graine en pâtissent. [Traduit par la Rédaction]

**Mots-clés :** paramètres agronomiques, méga environnement, coefficient de corrélation de Pearson, paramètres qualitatifs de la graine, hybridation du soja

## Introduction

Soybean [*Glycine max* (L.) Merr.] is one of the most important protein and oilseed crops that is widely grown worldwide (Fang et al. 2017). Increasing attention has been paid to the improvement of soybean seed quality traits in addition to the agronomic traits, such as yield, as seed quality traits directly affect the nutritional quality of soybean, especially for the food-grade cultivars (Xie et al. 2012; Gong et al. 2018; Lee et al. 2019). However, most major seed quality traits and essential agronomic traits in soybean are controlled by quantitative trait loci (QTL) with complicated interactions between alleles (Panthee et al. 2006b; Medic et al. 2014; Vaughn et al. 2014; Wang et al. 2014; Zhang et al. 2018; Lee et al. 2019). Some of the traits are linked with each other in a favorable way, but in many cases, it is very common to see undesirable associations between some traits, including the consistent negative correlation between protein and oil concentrations (Lee et al. 2019). Therefore, establishing a thorough understanding of the correlations between different traits may allow breeders to avoid or manage undesirable associations and utilize the desirable ones in their breeding populations. Moreover, combining the correlations with genotypic studies may help further in selecting favorable genes and breaking the undesirable linkages with the assistance of marker-assisted selection (Karikari et al. 2019; Yoosefzadeh-Najafabadi et al. 2022).

Among diverse seed quality traits, amino acids, fatty acids, and sugar-related traits have gained more attention from breeding programs. Amino acids and fatty acids determine the quality of soybean protein and oil, respectively (Lee et al. 2019). The sugar-related traits are critical for the soy-food flavor in addition to the health benefits (Wang et al. 2014; Dhungana et al. 2017; Lee et al. 2019). In contrast to well-documented correlations between agronomic traits, the complex correlations between some important seed quality traits on the one hand and agronomic traits on the other are not as well studied. The potential mechanisms involved in the correlations, when unknown, may lead to barriers for cultivar improvement (Bachlava et al. 2008; Kumar et al. 2010; Zhang et al. 2018; La et al. 2019; Lee et al. 2019). Four essential amino acids, including lysine (Lys), cysteine (Cys), methionine (Met), and threonine (Thr), five major fatty acids (linoleic, oleic, linolenic, palmitic, and stearic acids), and the three sugar-related traits (sucrose, stachyose, and raffinose), together with the major agronomic traits: yield, days to maturity (DTM), lodging, and height were selected to be the focus for this correlation study as they highly affect the quality of commercial soybean cultivars (Palomeque et al. 2010; Lee et al. 2019). The strong negative association between protein and oil is well established in the literature (Hernández-Sebastià et al. 2005; Zhang et al. 2018; La et al. 2019; Lee et al. 2019). Nevertheless, there have been limited studies testing the correlations between seed quality traits per se, and various conclusions were presented based on different analysis and assessment methods (Li et al. 2018). In general, protein concentration has been reported as being positively correlated with amino acid concentrations when calculated based on the total seed dry weight (Zhang et al. 2018). However, negative correlations were observed between protein and Cys, as

well as Met with Cys, Thr, and Lys in several studies (Panthee et al. 2006b; Medic et al. 2014; Vaughn et al. 2014). In terms of fatty acid traits, correlations, as reported, have been relatively consistent, which included the strong positive correlation between palmitic and stearic acids, in addition to the negative associations between oleic acid and linoleic and linoleic acid concentrations (Bachlava et al. 2008; Abdelghany et al. 2020). The strength of correlations for the sugar-related traits varied in different studies (Kumar et al. 2010; Zeng et al. 2014; Bueno et al. 2018; Jiang et al. 2018; La et al. 2019). Sucrose concentration was reported as negatively correlated to the protein concentration by Lee et al. (2019) and Dhungana et al. (2017). No significant correlation was reported for sucrose with stachyose and raffinose (Wang et al. 2014).

Lack of detailed understanding regarding correlations among seed quality traits and their correlations with agronomic traits poses challenges for breeding programs to effectively improve desirable seed traits without affecting others (Zhang et al. 2018). Zhang et al. (2018) indicated weak and nonsignificant correlations between protein or oil and DTM. A negative association between yield and protein was also reported, although it was not as strong as the correlation between protein and oil (Vaughn et al. 2014). In a study conducted by Bachlava et al. (2009), yield was determined to be negatively correlated with oleate concentration but positively correlated with linoleate and linolenate concentrations. There is much less information available in the literature on the correlations between seed quality traits and agronomic traits. Therefore, the objective of this study was to examine and identify correlations between major seed quality traits as well as agronomic traits across the two mega-environments (MEs) in Canada and Ukraine. By understanding the complex trait correlations, the genomic information can be combined with the phenotypic correlations to directly select preferable traits or indirectly select associated traits with the aid of molecular marker technologies to facilitate soybean cultivar development.

## Materials and methods

### Plant materials and experimental design

The soybean diversity panel consisting of 200 accessions within maturity group 0 (MG 0) was used in this study. The majority of the panel was made up of 127 University of Guelph accessions, which were not sister lines, accounting for 63% of the total. The rest of the panel included 35 accessions from the Agriculture and Agri-Food Canada, 19 accessions from Le Centre de recherche sur les grains inc. in Québec, Canada, 10 Northern US cultivars, and nine diverse ancestral soybean cultivars. Sixteen accessions were removed from the original panel consisting of 200 accessions due to the poor emergence in the field or seed mixtures in the source in 2018. These modifications resulted in some changes to the total accessions used in further analyses bringing it to the final number of 184. All the accessions were planted in three locations in Ontario and two locations in Ukraine, respectively. The three locations in Ontario were: Elora Research Station (ERS), Woodstock Research Sta-

**Table 1.** Planting and harvest dates for Elora, Woodstock, and St. Paul's in Ontario, and Kodyma, Lyubar, and Kotovsk in Ukraine in 2018 and 2019.

Year	Location	Planting date	Harvest date
2018	Elora	29th May	20th October
	Woodstock	2nd June	23rd November
	St. Paul's	23rd May	25th October
	Kodyma	6th May	21st September
	Lyubar	4th May	24th September
2019	Elora	11th June	23rd October
	Woodstock	14th June	26th November
	St. Paul's	9th June	17th November
	Kotovsk	12th May	28th September

tion (WRS), and St. Paul's (STP), ON, for both 2018 and 2019. The two Ukrainian locations were Lyubar and Kodyma for 2018, and Lyubar and Kotovsk for 2019. Two replications were planted per location using the randomized complete block design (RCBD). In the three Ontario locations, the plots were 5 m x 1.65 m (8.25 m<sup>2</sup>) with four rows spaced at 35 cm between rows, and 500 seeds were planted per plot. In Ukraine, the plot sizes were 5.425 and 15.736 m<sup>2</sup> for 2018 and 2019, respectively, to accommodate the local equipment settings. In both countries, the whole plot was harvested to ensure enough seeds for data collection in the seed lab upon harvest. The planting and harvest dates are summarized in Table 1.

## Phenotypic data collection

For Ontario locations, the measurements of agronomic traits, including yield, height, lodging, and DTM, were conducted during the growing seasons and after the harvest seasons. The yield measurements were recorded per plot at each location and then converted to kg ha<sup>-1</sup> with the adjustment to 13% moisture. Plant height (cm) was measured as the average distance between the soil surface and the tip of the main stem at maturity. The DTM was defined as the number of days after planting until 95% of the pods in the plots reached the development stage R8 (Fehr et al. 1971). The lodging was visually assessed with a score that ranged from 1 (no lodging) to 5 (completely prostrate) (Fehr et al. 1971). The seed samples were measured for the 14 seed quality traits (protein, oil, Lys, Cys, Met, Thr, linoleic acid, oleic acid, linolenic acid, palmitic acid, stearic acid, sucrose, stachyose, and raffinose) using a Perten DA 7250 near-infrared reflectance (NIR) machine. The Ukrainian agronomic trait (yield, kg ha<sup>-1</sup> at 13% moisture) and the seed quality traits (protein and oil) were measured and collected using the same model of Perten DA 7250 NIR machine by the Ukrainian cooperators. The data of other seed quality traits were not provided by the Ukrainian cooperators due to resource limitations.

## Statistical analyses and GT biplots

Analysis of variance (ANOVA) of seed quality and agronomic traits and the radial smoothing process was conducted within each location using the PROC GLIMMIX procedure in SAS version 9.4 (SAS Institute Inc., Cary, NC, USA) for the

RCBD. The distribution of the residuals was determined and visualized based on the results from PROC UNIVARIATE procedure in SAS version 9.4. The normality of residual distribution was examined using the PROC SGPLOT procedure (SAS Institute Inc.). The LSMEANS used for the correlation analyses were obtained for each seed quality trait and agronomic trait for single and combined locations using the PROC GLIMMIX procedure. Pearson's correlation coefficients were calculated between seed quality traits as well as with agronomic traits using the PROC CORR procedure in SAS version 9.4. The correlation analyses were conducted on single-year LSMEANS as well as combined-year LSMEANS. The traits were considered strongly correlated if the absolute values of the correlation coefficients were equal to or larger than 0.70 ( $|r| \geq 0.70$ ), and they were considered moderately correlated if the absolute values of  $r$  were equal or larger than 0.40, but smaller than 0.70 ( $0.40 \leq |r| < 0.70$ ), using a significant threshold of 0.05. The seed composition trait and the agronomic trait values after spatial correction were organized into a four-way dataset and entered into the GGE Biplot software version 8.1 to generate the genotype-by-trait (GT) biplots with an SD-scaled method (Yan and Rajcan 2002). The missing values were not computed.

## Results

### Correlations between seed quality traits and agronomic traits under Canadian mega-environment and Ukrainian mega-environment

In general, similar correlations were consistently observed in the separated years and combined-year data as presented in Tables 2–4. Nonsignificant correlations were left out from the tables to avoid overloading the tables. In all three datasets, strong positive correlations were consistently observed between protein and the four amino acids, and also within the four amino acids, with  $r$  values ranging from 0.82 to 0.99 (Tables 2–4). Stearic acid was positively associated with palmitic acid, showing an  $r$  value greater than 0.50 (Tables 2–4). Moderate positive correlations also existed between sucrose and linolenic acid, besides sucrose and raffinose (Tables 2–4). Some other positive correlations were presented as well but were not significant in all years and the combined ANOVA, which included stearic acid and oleic acid (Tables 3 and 4), along with stachyose and raffinose (Table 3). In contrast, oil concentration was negatively correlated to protein and the four amino acids (Tables 2–4). Linoleic acid was negatively correlated with oleic acid and stearic acid (Tables 2–4). Consistent moderate-to-strong negative correlations were found between sucrose and the four amino acids as well as the protein concentration (Tables 2–4). In 2019, the stachyose and oil concentrations were indicated to negatively affect each other (Table 3). In terms of the agronomic traits, DTM consistently showed moderate positive associations with yield and plant height (Tables 2–4). Lodging and height were positively correlated (Tables 2–4). In addition, yield and height (Tables 2 and 4), as well as DTM and lodging (Tables 2 and 4), were positively correlated with each other,

**Table 2.** Pearson's correlation coefficient values among 18 seed quality and agronomic traits combining the data from three locations (Elora, Woodstock, and St. Paul's) in Canada in 2018.

Traits	Oil	Lys	Cys	Met	Thr	Linoleic acid	Oleic acid	Linolenic acid	Palmitic acid	Stearic acid	Sucrose	Stachyose	Raffinose	Yield	Height	DTM
Protein	$r = -0.58^{***}$	$r = 0.99^{***}$	$r = 0.83^{***}$	$r = 0.91^{***}$	$r = 0.98^{***}$	$r = -0.15^*$	$r = 0.22^{**}$	$r = -0.20^{**}$			$r = -0.58^{***}$		$r = -0.24^{***}$	$r = 0.18^*$	$r = 0.16^*$	$r = 0.18^*$
Oil		$r = -0.60^{***}$	$r = -0.49^{***}$	$r = -0.54^{***}$	$r = -0.54^{***}$	$r = 0.29^{***}$		$r = -0.35^{***}$	$r = -0.17^*$	$r = -0.21^{**}$		$r = -0.25^{***}$	$r = -0.34^{***}$	$r = -0.18^*$		$r = -0.26^{***}$
Lys			$r = 0.83^{***}$	$r = 0.91^{***}$	$r = 0.98^{***}$		$r = 0.17^*$	$r = -0.17^*$			$r = -0.55^{***}$		$r = -0.25^{***}$	$r = 0.20^{**}$	$r = 0.17^*$	$r = 0.21^{**}$
Cys				$r = 0.91^{***}$	$r = 0.82^{***}$			$r = -0.16^*$			$r = -0.59^{***}$	$r = -0.17^*$	$r = -0.21^{**}$			
Met					$r = 0.89^{***}$	$r = -0.17^*$	$r = 0.24^{**}$	$r = -0.24^{**}$			$r = -0.51^{***}$		$r = -0.22^{**}$		$r = 0.21^{**}$	$r = 0.16^*$
Thr							$r = 0.17^*$	$r = -0.16^*$			$r = -0.62^{***}$		$r = -0.32^{***}$	$r = 0.18^*$	$r = 0.15^*$	$r = 0.16^*$
Linoleic acid							$r = -0.89^{***}$			$r = -0.44^{***}$		$r = -0.18^*$	$r = -0.32^{***}$			
Oleic acid										$r = 0.30^{***}$		$r = 0.30^{***}$	$r = 0.18^*$			
Linolenic acid							$r = -0.43^{***}$				$r = 0.45^{***}$	$r = -0.20^{**}$	$r = 0.18^*$		$r = -0.16^*$	
Palmitic acid								$r = 0.24^{**}$		$r = 0.50^{***}$						$r = -0.15^*$
Stearic acid											$r = 0.20^{**}$	$r = 0.16^*$	$r = 0.32^{***}$			
Sucrose												$r = 0.39^{***}$	$r = 0.47^{***}$	$r = 0.16^*$		$r = 0.24^{***}$
Stachyose													$r = 0.28^{***}$	$r = 0.27^{***}$	$r = 0.20^{**}$	$r = 0.35^{***}$
Raffinose														$r = -0.22^{**}$	$r = -0.15^*$	$r = -0.17^*$
Yield															$r = 0.42^{***}$	$r = 0.55^{***}$
Lodging							$r = 0.19^*$	$r = -0.17^*$		$r = 0.17^*$					$r = 0.67^{***}$	$r = 0.52^{***}$
DTM															$r = 0.67^{***}$	

**Note:** Only significant correlations are presented; the significant threshold was  $P < 0.05$ .

\* $P < 0.05$ .

\*\* $P < 0.01$ .

\*\*\* $P < 0.001$ .



**Table 3.** Pearson's correlation coefficient values among 18 seed quality and agronomic traits combining the data from three locations (Elora, Woodstock, and St. Paul's) in Canada in 2019.

Traits	Oil	Lys	Cys	Met	Thr	Linoleic acid	Oleic acid	Linolenic acid	Palmitic acid	Stearic acid	Sucrose	Stachyose	Raffinose	Yield	Height	DTM
Protein	$r = -0.55^{***}$	$r = 0.99^{***}$	$r = 0.87^{***}$	$r = 0.91^{***}$	$r = 0.98^{***}$		$r = 0.31^{***}$		$r = 0.26^{***}$	$r = 0.15^*$	$r = -0.63^{***}$	$r = 0.27^{***}$	$r = -0.22^{**}$			
Oil		$r = -0.56^{***}$	$r = -0.48^{***}$	$r = -0.55^{***}$	$r = -0.50^{***}$	$r = 0.26^{***}$	$r = 0.20^{**}$	$r = -0.50^{***}$	$r = -0.27^{***}$			$r = -0.45^{***}$	$r = -0.37^{***}$			$r = -0.21^{**}$
Lys			$r = 0.88^{***}$	$r = 0.93^{***}$	$r = 0.98^{***}$		$r = 0.25^{***}$		$r = 0.23^{**}$		$r = -0.61^{***}$	$r = 0.27^{***}$	$r = -0.23^{**}$			
Cys				$r = 0.92^{***}$	$r = 0.86^{***}$		$r = 0.20^{**}$	$r = -0.17^*$	$r = 0.21^{**}$	$r = 0.18^{**}$	$r = -0.60^{***}$		$r = -0.25^{***}$	$r = -0.15^*$		
Met					$r = 0.88^{***}$		$r = 0.26^{***}$		$r = 0.28^{***}$	$r = 0.21^{**}$	$r = -0.50^{***}$	$r = 0.23^{**}$	$r = -0.22^{**}$			
Thr							$r = 0.27^{***}$		$r = 0.22^{**}$		$r = -0.70^{***}$	$r = 0.18^*$	$r = -0.31^{***}$			
Linoleic acid						$r = -0.91^{***}$	$r = -0.21^{**}$		$r = -0.31^{***}$	$r = -0.49^{***}$	$r = -0.17^*$	$r = -0.23^{**}$	$r = -0.36^{***}$			
Oleic acid									$r = 0.37^{***}$	$r = 0.50^{***}$		$r = 0.26^{***}$	$r = 0.16^*$			
Linolenic acid											$r = 0.51^{***}$	$r = 0.15^*$	$r = 0.34^{***}$			
Palmitic acid										$r = 0.76^{***}$				$r = 0.19^*$		
Stearic acid													$r = 0.17^*$			
Sucrose												$r = 0.26^{***}$	$r = 0.57^{***}$	$r = 0.22^{**}$		$r = 0.33^{***}$
Stachyose													$r = 0.40^{***}$			$r = 0.19^*$
Raffinose																
Yield																
Lodging						$r = -0.17^*$		$r = 0.21^{**}$			$r = 0.16^*$		$r = 0.22^{**}$	$r = -0.24^{**}$	$r = 0.36^{***}$	$r = 0.59^{***}$
DTM															$r = 0.46^{***}$	$r = 0.18^*$
															$r = 0.44^{***}$	

**Note:** Only significant correlations are presented; the significant threshold was  $P < 0.05$ .

\* $P < 0.05$ .

\*\* $P < 0.01$ .

\*\*\* $P < 0.001$ .

**Table 4.** Pearson's correlation coefficient values among 18 seed quality and agronomic traits combining the 2018 and 2019 data from three locations (Elora, Woodstock, and St. Paul's) in Canada.

Traits	Oil	Lys	Cys	Met	Thr	Linoleic acid	Oleic acid	Linolenic acid	Palmitic acid	Stearic acid	Sucrose	Stachyose	Raffinose	Yield	Height	DTM
Protein	$r = -0.57^{***}$	$r = 0.99^{***}$	$r = 0.87^{***}$	$r = 0.93^{***}$	$r = 0.98^{***}$	$r = -0.16^*$	$r = 0.28^{***}$		$r = 0.20^{**}$		$r = -0.61^{***}$	$r = 0.18^*$	$r = -0.24^{***}$			
Oil		$r = -0.58^{***}$	$r = -0.49^{***}$	$r = -0.55^{***}$	$r = -0.51^{***}$	$r = 0.32^{***}$	$r = -0.20^{**}$	$r = -0.46^{***}$	$r = -0.26^{***}$	$r = -0.20^{**}$		$r = -0.37^{***}$	$r = -0.38^{***}$			$r = -0.24^{**}$
Lys			$r = 0.87^{***}$	$r = 0.94^{***}$	$r = 0.98^{***}$		$r = 0.24^{**}$		$r = 0.18^*$		$r = -0.59^{***}$	$r = 0.17^*$	$r = -0.25^{***}$			
Cys				$r = 0.93^{***}$	$r = 0.86^{***}$		$r = 0.18^*$	$r = -0.16^*$	$r = 0.15^*$		$r = -0.60^{***}$		$r = -0.23^{**}$			
Met					$r = 0.90^{***}$		$r = 0.27^{***}$	$r = -0.17^*$	$r = 0.19^{**}$	$r = 0.17^*$	$r = -0.51^{***}$	$r = 0.15^*$	$r = -0.23^{**}$			
Thr							$r = 0.25^{***}$		$r = 0.18^*$		$r = -0.67^{***}$		$r = -0.32^{***}$			
Linoleic acid							$r = -0.92^{***}$		$r = -0.19^*$	$r = -0.49^{***}$	$r = -0.15^*$	$r = -0.25^{***}$	$r = -0.36^{***}$			
Oleic acid								$r = -0.26^{***}$		$r = 0.40^{***}$		$r = 0.31^{***}$	$r = 0.17^*$	$r = 0.17^*$		
Linolenic acid									$r = 0.16^*$		$r = 0.48^{***}$		$r = 0.30^{***}$			
Palmitic acid										$r = 0.66^{***}$				$r = 0.15^*$		
Stearic acid											$r = 0.16^*$	$r = 0.19^{**}$	$r = 0.31^{***}$			
Sucrose												$r = 0.34^{***}$	$r = 0.55^{***}$	$r = 0.21^{**}$		$r = 0.29^{***}$
Stachyose													$r = 0.36^{***}$	$r = 0.17^*$		$r = 0.30^{***}$
Raffinose														$r = -0.16^*$		
Yield															$r = 0.41^{***}$	$r = 0.62^{***}$
Lodging											$r = 0.17^*$		$r = 0.18^*$		$r = 0.61^{***}$	$r = 0.45^{***}$
DTM															$r = 0.62^{***}$	

**Note:** Only significant correlations are presented; the significant threshold was  $P < 0.05$ .

\* $P < 0.05$

\*\* $P < 0.01$

\*\*\* $P < 0.001$

respectively, but not in 2019. There was no significant strong correlation found between the agronomic traits and the seed quality traits in this genome-wide association study (GWAS) panel across Ontario locations (Tables 2–4).

The combined year and location GT biplot (Fig. 1) for Ontario visually presented the correlations between all the seed compositions and agronomic traits. The biplot explained 50.6% of the variation, with 31.8% explained by PC1, and 18.8% by PC2 (Fig. 1). The significant and strong correlations observed in the biplot were consistent with the correlation coefficients calculated earlier (Fig. 1 and Tables 2–4).

For Ukraine, the data from Kodyma in 2018 and Lyubar in 2019 were dropped from the combined and correlation analyses as the data were incomplete and invalid according to the ANOVA results. The Pearson's correlation coefficients of the traits for the separate years and the combined-year data are presented in Table 5. The negative correlation between protein and oil ( $r = -0.67$  for 2018, 2019, and combined-year data) was consistently shown for the GWAS panel as expected (Table 5). Combining the two-year data, protein and yield showed significant and negative correlation ( $r = -0.15$ ), while oil and yield showed a significant positive correlation ( $r = 0.19$ ) (Table 5). Moreover, oil was positively associated with the yield in 2019 in Kotovsk ( $r = 0.20$ ) (Table 5). However, these three significant correlations were weak in value (Table 5). The combined GT biplot (Fig. 2) accounted for 89.1% of the variation, with PC1 = 58.3% and PC2 = 30.8%. The negative association between protein and oil was confirmed in the biplot as well (Fig. 2 and Table 5).

## Correlations between seed quality traits and agronomic traits across Canadian and Ukrainian mega-environments

The correlations across the two countries were studied among limited number of traits due to the lack of available data from Ukraine as only protein and oil data were collected by cooperators in Ukraine. The Pearson's correlation coefficients are listed in Table 6. Combining the data across the two MEs, protein and oil consistently interacted with each other in a moderately negative way, with  $r$  value ranging from  $-0.62$  to  $-0.58$  (Table 6). A significant negative correlation between protein and yield was observed in the 2-year combined analysis, and a positive correlation between oil and yield was found in 2019 (Table 6). However, the correlations between yield and protein or oil were weak and inconsistent (Table 6). The trait correlations derived from combined locations and years (Fig. 3) were consistent with Pearson's correlation results (Table 6). A total of 85.7% variation was explained in the biplot, with 52.5% by PC1 and 33.2% by PC2 (Fig. 3).

## Discussion

The negative correlation between protein and oil was consistently observed in both countries representing MEs. This correlation could be explained from both physiological and genetic perspectives (Hernández-Sebastià et al. 2005; Zhang et al. 2018; La et al. 2019; Lee et al. 2019). The competition for the carbon skeletons between protein and oil synthesis path-

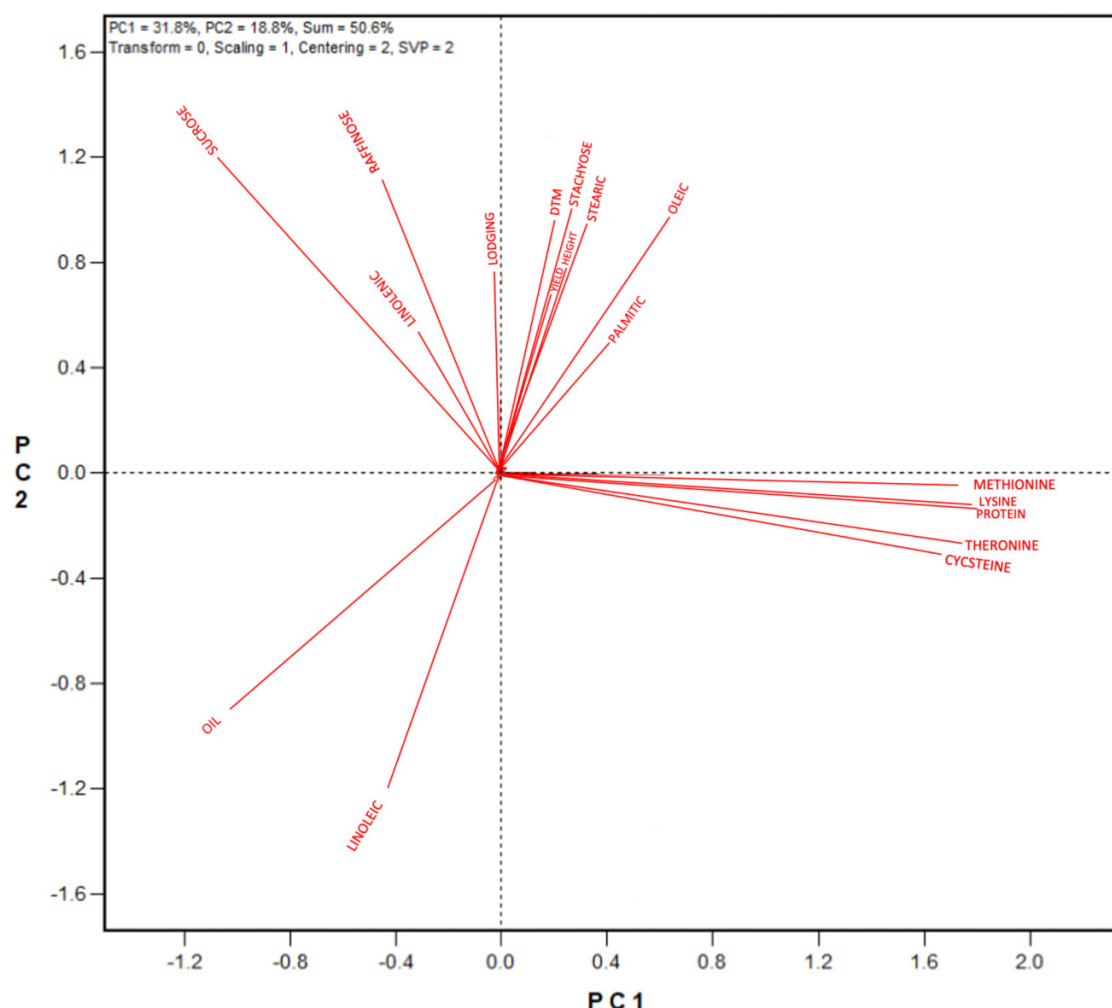
ways could be the potential contributor for the inverse correlation (Hernández-Sebastià et al. 2005). In terms of genetic control, multiple studies confirmed the pleiotropy that exists in QTL that govern protein and oil concentrations (Zhang et al. 2018; La et al. 2019; Lee et al. 2019). There were 23 significant single nucleotide polymorphisms (SNPs) identified in common for both protein and oil, which showed negative effects located on chromosomes (Chr) 15 and 20 (Lee et al. 2019). This was also indicated in a study conducted by Zhang et al. (2018), in which a QTL tagged by the marker *ss715622170* on Chr 15 exhibited significant associations with both protein and oil in addition to other 18 seed quality traits (Zhang et al. 2018).

Correlations between all targeted traits were determined from the data obtained in Ontario locations. The strong positive correlations between protein and the four amino acids, as well as the four amino acids with each other, were in accordance with the literature when calculated based on the dry seed weight basis (Assefa et al. 2018; Zhang et al. 2018; La et al. 2019; Ma et al. 2019). There are strong natural correlations between protein and amino acid concentrations, since amino acids are the major components of protein (Zhang et al. 2018). These correlations could result from diverse reasons such as the shared biosynthetic pathways, co-located QTL, pleiotropic QTL, and gene linkages (Shaul and Galili 1992; Hacham et al. 2007; Kastoori Ramamurthy et al. 2014; Ma et al. 2019). In plants, Lys, Met, and Thr are synthesized through the aspartate (Asp) family biosynthesis pathway by different branches (Hacham et al. 2007; Warrington et al. 2015). In soybean, Cys is the intermediate product of sulfur assimilation of protein synthesis, which also relates to the Met synthesis involving Asp and various enzymes (Wang et al. 2015; Ma et al. 2019). Therefore, these traits may share some common genetic-regulating factors and enzymes belonging to similar biosynthetic pathways (Ma et al. 2019). Wang et al. (2015) found four QTL associated with Cys and Met were also co-localized with the previously identified QTL for protein on Chr 3, 4, 17, and 20. There was a major genomic region on Chr 7 associated with Met, Cys, and protein identified by Ma et al. (2019). Similarly, several other QTL controlling both Cys and Met concentrations were identified on Chr 10 and 20 (Kastoori Ramamurthy et al. 2014; Warrington et al. 2015).

Several studies have identified QTL and biosynthetic pathways that independently associated with some of the traits described above without affecting the others. One QTL tagged by *ss715602750* on Chr 8 was found to be related to the synthesis of Asp-related amino acids but unrelated to the protein concentration (Zhang et al. 2018). Similarly, a very close region on Chr 8 associated with marker *ss715602763* was identified as responsible for Cys, Lys, and Thr but not the total protein concentration (Vaughn et al. 2014). Comparing these results, the *AK-HSDH* gene with described functions in this region was suggested to be the candidate gene regulating the Asp-related amino acid without affecting the total protein concentration (Zhang et al. 2018). This finding was meaningful since some of the correlations between protein and amino acids become undesirable when amino acid concentrations are expressed as the proportion of total crude protein (Zhang et al. 2018; La et al. 2019). Another study con-



**Fig. 1.** Genotype-by-trait biplot for a combined year and location data from Elora, Woodstock, and St. Paul's in 2018 and 2019 for 18 seed quality and agronomic traits.



**Table 5.** Pearson's correlation coefficient values between protein, oil, and yield for 2018, 2019, and combined year and location data in Ukraine.

Environment	Trait	Oil	Yield
Lyubar 2018	Protein	$r = -0.67^{***}$	NS
Kotovsk 2019		$r = -0.67^{***}$	NS
Combined		$r = -0.67^{***}$	$r = -0.15^*$
Lyubar 2018	Oil		NS
Kotovsk 2019			$r = 0.20^{**}$
Combined			$r = 0.19^*$

Note: NS, not significant; the significant threshold was  $P < 0.05$ .

\* $P < 0.05$ .

\*\* $P < 0.01$ .

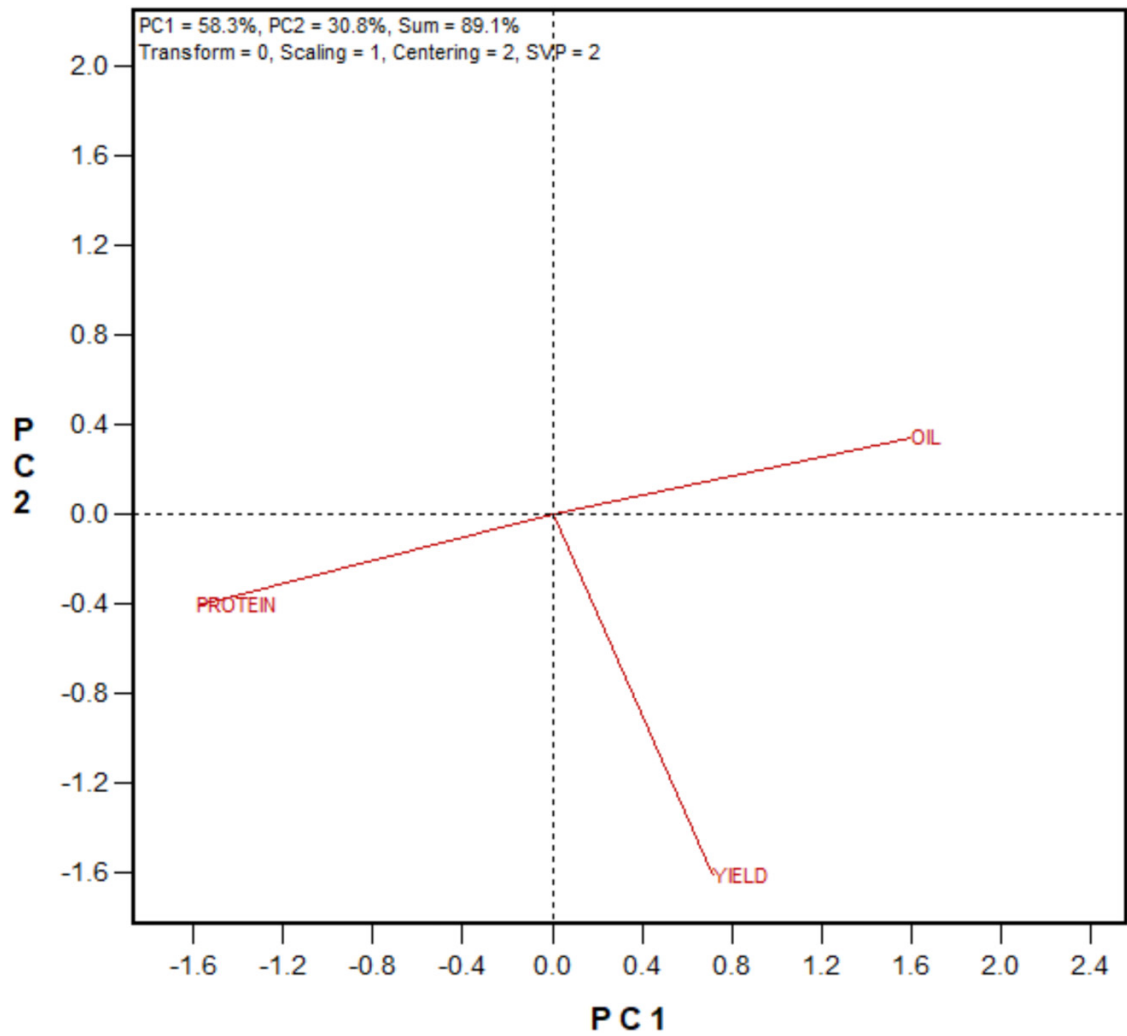
\*\*\* $P < 0.001$ .

ducted on tobacco plants indicated a potential solution to enhance both Lys and Met without reducing the Thr concentration in legume and cereal crops (Hacham et al. 2007). In brief, the increased Lys can alter the expression level of S-adenosylmethionine synthase, resulting in upregulation of

Met concentration without reducing Thr in the transgenic tobacco plants expressing both dihydridipicolinate synthase and Arabidopsis cystathionine  $\gamma$ -synthase (Hacham et al. 2007). These findings provided background information and potential solutions to break the undesirable correlations and utilize the beneficial ones.

The correlations between fatty acids were consistent with previous reports (Maestri et al. 1998; Bachlava et al. 2008; Li et al. 2015; La et al. 2019; Abdelghany et al. 2020). The positive correlations between stearic and palmitic acids, in addition to the negative correlations between linoleic acid with oleic and stearic acids, were well established in different studies (Maestri et al. 1998; Cardinal and Burton 2007; Bachlava et al. 2008; Li et al. 2015; Abdelghany et al. 2020). Cardinal and Burton (2007) suggested the *GmFATB1a* gene might also possess activities toward stearyl acyl carrier protein substrates besides encoding the palmitate thioesterase, which was a potential explanation for the positive correlation between palmitic and stearic acids. Moreover, Zhang et al. (2018) identified two QTL responsible for both palmitic and stearic acids on Chr 5 and 14, which were associated with markers ss715592503 and ss715618427, respectively (Zhang et al. 2018). The dele-

**Fig. 2.** Genotype-by-trait biplot for a combined year and location data from Lyubar and Kotovsk in Ukraine in 2018 and 2019 for protein, oil, and yield.



**Table 6.** Pearson’s correlation coefficient values between protein, oil, and yield for 2018, 2019, and combined year and location data across Canada and Ukraine.

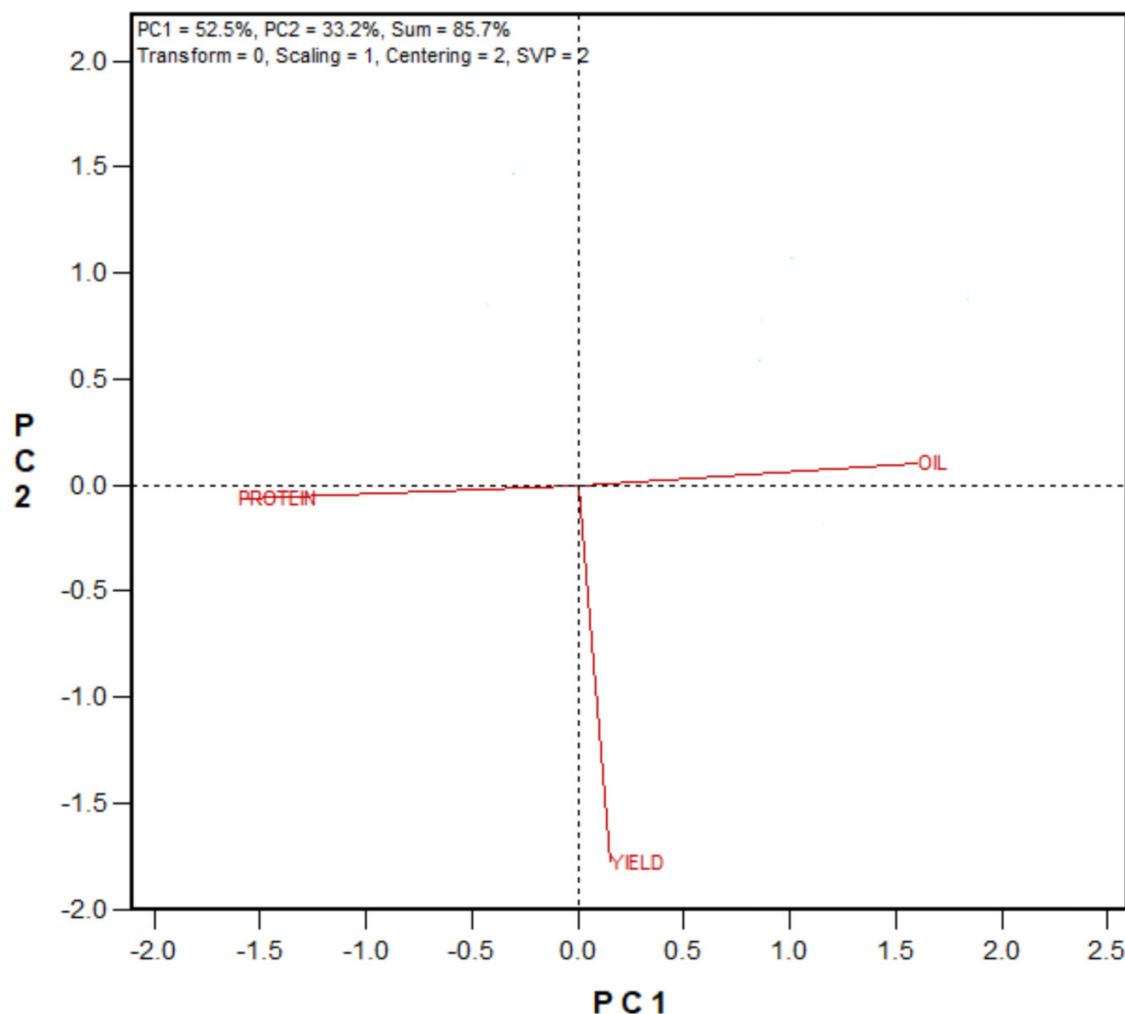
Year	Trait	Oil	Yield
2018	Protein	$r = -0.60^{***}$	NS
2019		$r = -0.62^{***}$	NS
Combined		$r = -0.58^{***}$	$r = -0.17^*$
2018	Oil		NS
2019			$r = 0.15^*$
Combined			NS

**Note:** NS, not significant; the significant threshold was  $P < 0.05$ .  
\* $P < 0.05$ .  
\*\*\* $P < 0.001$ .

tion of the candidate gene *SACPD-C* tagged by *ss715618427* could lead to the elevation of stearic acid concentration in soybean (Gillman et al. 2014). The negative associations between linoleic acid on one hand and stearic acid, oleic acid on the other could be attributed to the fatty acid desaturation pathway (Bachlava et al. 2008; Li et al. 2015). During this

conversion, stearic acid is the precursor of oleic acid, and then linoleic acid is further synthesized from oleic acid by the microsomal  $\omega$ -6 desaturase enzymes coded by the *Fad2-1* and *Fad2-2* genes (Rubel et al. 1972; Bachlava et al. 2008, 2009; Li et al. 2015).  
In addition to the biochemical and genetic factors described above, various QTL with pleiotropic effects were described by previous studies regarding the tight associations between the fatty acids (Hyten et al. 2004; Panthee et al. 2006a; Bachlava et al. 2008; Li et al. 2015; Zhou et al. 2019). Panthee et al. (2006a) identified two markers (Satt185 and Satt263) on Chr 15 consistently associated with both oleic and linoleic acids, which was likely related to the modifier QTL responsible for certain enzymes in the fatty acid desaturation pathway. Four SNPs identified by Li et al. (2015) presented significant associations with both oleic and linoleic acids but with inverse effects (Li et al. 2015). There was an interval on Chr 19 harboring significant QTL for palmitic acid, which was also significantly associated with linoleic, oleic, and linolenic acids (Hyten et al. 2004). A total of seven influencing SNPs for oleic and linoleic acids were located on Chr 1, 7, 8, 9, and

**Fig. 3.** Genotype-by-trait biplot for a combined year and location data from Elora, Woodstock, and St. Paul's in Canada, and Lyubar and Kotovsk in Ukraine in 2018 and 2019 for protein, oil, and yield.



13 (Zhou et al. 2019). The inconsistency observed in different year–location combination for the correlation between stearic acid and oleic acid could be the result of the strong interaction of these traits with the environmental factors such as heat and drought conditions (Hou et al. 2006). These correlations between fatty acids should be considered and properly utilized depending on the specific breeding purposes, such as for human consumption and oil processing, since the emphasis on specific fatty acid may change based on the product in mind (Li et al. 2015; Zhou et al. 2019).

The strong negative correlation between sucrose and protein was described in detail by different research groups (Dhungana et al. 2017; Patil et al. 2018; La et al. 2019). As the corresponding traits for protein and oil, it was not surprising to expect correlations existing between sucrose and amino acids as well as linolenic acid (La et al. 2019). The production of sucrose, protein, amino acids, and oil resulted from integrated pathways including nitrogen assimilation, carbohydrate production, and carbon assimilation, which may provide a potential explanation regarding the correlations (Paul and Foyer 2001; Li et al. 2012). The syntheses of amino acids

and protein require carbon skeletons supported by sucrose, which leads to the negative correlations between them (Paul and Foyer 2001). In addition, common genetic control may exist between protein and sucrose, as the identified protein QTL Seed protein 7-g10 was located close to a significant sucrose SNP at the 34 025 091 bp location on Chr 12 (Zhang et al. 2017). Very limited number of studies reported correlations between sucrose and linolenic acid as well as the underlying mechanism, and the positive correlation identified in this study was not consistent with previous studies (La et al. 2019; Zhao et al. 2019). Moreover, the previously reported results were not consistent with each other, and most of the correlations were nonsignificant (Zhang et al. 2017; Zhao et al. 2019). Consequently, further studies are required to confirm these identified correlations, which might be highly affected by environmental factors (Dhungana et al. 2017). The significant positive correlation between sucrose and raffinose at different levels was previously reported (Wang et al. 2014; Bueno et al. 2018; Jiang et al. 2018; La et al. 2019). This correlation was considered to be the consequence of the raffinose oligosaccharide metabolic pathway, in which the raf-

finose synthesis depends on the sucrose substrate (Hannah et al. 2006; Bueno et al. 2018). From the genetic perspective, the marker Satt359 on Chr 11 was indicated to be common for both sucrose and raffinose, which further supported the correlation between them (Wang et al. 2014). However, the sucrose concentration seemed to exhibit strong interactions with environments, but the raffinose concentration tended to be more genotype dependent, which possibly led to fluctuations in correlations between these sugar-related compositions (Kumar et al. 2010).

The positive correlations between DTM and yield, DTM and plant height, and lodging and height were documented previously (Cicek et al. 2006; Bachlava et al. 2008; Palomeque et al. 2010; Rossi et al. 2013). Egli et al. (1981) suggested that the increase of the DTM also allowed a longer seed-filling period, which dramatically contributed to the yield for grain crops, including soybean. The development of genetic technologies assisted recent studies in revealing the underlying genetic connections between these traits, which resulted in the correlations (Palomeque et al. 2010; Kim et al. 2012; Rossi et al. 2013; Fang et al. 2017). Kim et al. (2012) mapped several contributing QTL for DTM, height, and lodging located very close to the yield QTL on Chr 4, 14, and 18, which supported the opinion that increased yield was associated with later maturity and taller plants. A QTL linked with Satt162 was responsible for both plant height and lodging but was not in all the environments studied (Palomeque et al. 2010). Same authors also identified yield QTL tagged by Satt100, Satt277, and Sat\_126 co-localized with several agronomic traits, including plant height and DTM (Palomeque et al. 2010; Rossi et al. 2013). Although inconsistency existed in the correlations and QTL for yield and other related agronomic traits, most of the previous studies agreed that major QTL were shared between these traits, showing pleiotropic or additive gene effects (Cober and Morrison 2010; Kim et al. 2012; Rossi et al. 2013).

Inconsistent patterns were observed in terms of correlations between some seed quality traits and agronomic traits, such as sugar-related compositions and amino acids or fatty acids with agronomic traits (Bachlava et al. 2008; Kumar et al. 2010; Zhang et al. 2018; La et al. 2019; Lee et al. 2019). For instance, the nonsignificant negative correlation between DTM and oil concentration was reported by Zhang et al. (2018), which was different from the previous result by Bachlava et al. (2008). Several significant correlations between seed quality traits and agronomic traits were observed in this study, such as that between yield with protein and oil, but none of them was strong. The potential influencing factors regarding the inconsistency of these correlations included the difference existing in various populations used, changeable environments, and strong interactions between genotypes and environments (Kumar et al. 2010; Dhungana et al. 2017; Jiang et al. 2018; La et al. 2019).

In conclusion, the accumulation of diverse seed quality traits was correlated with each other and showed consistent trends in both Canada and Ukraine, and between the two MEs through the experimental years as discussed in this study. Consistent strong and significant correlations were observed between the agronomic traits based on 2-year Ontario

location data, but there was no strong significant correlation identified between the seed quality traits and agronomic traits in either MEs. Most of the findings in this study were in agreement with the literature. Identified undesirable correlations have provided difficult challenges for breeding programs aimed at improving desirable traits without sacrificing others (Zhang et al. 2018; La et al. 2019). These are potential challenges for breeding cultivars for Ukrainian market as well. Fortunately, an increasing number of trait-specific loci, which only affect one trait with almost no effect on others, were found for protein, oil, and several amino acids (Eskandari et al. 2013; Zhang et al. 2018; Lee et al. 2019). In addition to the potential modifications on genetic factors, manipulating the growing conditions by selecting proper locations may provide another possibility to enhance the desirable traits, for instance, sucrose (Kumar et al. 2010; Zeng et al. 2014). These findings may facilitate the proper utilization of different correlations between traits in future breeding programs targeting Ukraine.

## Acknowledgements

The authors thanks current and former staff of Dr. Istvan Rajcan's soybean research team at University of Guelph, the cooperation from Ukrainian research team and Université Laval in terms of field trial management, data collection and analytical supports. The authors also appreciate the funding supports from NSERC, Cangro Genetics Inc. and Huron Commodities Inc.

## Article information

### History dates

Received: 1 March 2022

Accepted: 10 June 2022

Accepted manuscript online: 12 July 2022

Version of record online: 2 September 2022

### Copyright

© 2022 The Author(s). Permission for reuse (free in most cases) can be obtained from [copyright.com](https://creativecommons.org/licenses/by/4.0/).

### Data availability

The data have been deposited on OneDrive at the University of Guelph. It may be available upon request to the corresponding author.

## Author information

### Author ORCIDs

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

### Author contributions

HH: formal analysis, investigation, methodology, resources, validation, writing—original draft, and writing—review and editing; MYN: data curation, methodology, resources, validation, and writing—review and editing; IR: conceptualization, data curation, funding acquisition, investigation, methodol-



ogy, project administration, resources, supervision, validation, and writing—review and editing.

## Competing interests

The authors declare there are no competing interests.

## References

- Abdelghany, A.M., Zhang, S., Azam, M., Shaibu, A.S., Feng, Y. Li, Y., et al. 2020. Profiling of seed fatty acid composition in 1025 Chinese soybean accessions from diverse ecoregions. *Crop J.* **8**: 635–644. doi:[10.1016/j.cj.2019.11.002](https://doi.org/10.1016/j.cj.2019.11.002).
- Assefa, Y., Bajjalieh, N., Archontoulis, S., Casteel, S., Davidson, D. Kovács, P., et al. 2018. Spatial characterization of soybean yield and quality (amino acids, oil, and protein) for United States. *Sci. Rep.* **8**: 1–11. doi:[10.1038/s41598-018-32895-0](https://doi.org/10.1038/s41598-018-32895-0). PMID: 29311619.
- Bachlava, E., Burton, J.W., Brownie, C., Wang, S., Auclair, J., and Cardinal, A.J. 2008. Heritability of oleic acid content in soybean seed oil and its genetic correlation with fatty acid and agronomic traits. *Crop Sci.* **48**: 1764–1772. doi:[10.2135/cropsci2008.01.0049](https://doi.org/10.2135/cropsci2008.01.0049).
- Bachlava, E., Dewey, R.E., Burton, J.W., and Cardinal, A.J. 2009. Mapping candidate genes for oleate biosynthesis and their association with unsaturated fatty acid seed content in soybean. *Mol. Breed.* **23**: 337–347. doi:[10.1007/s11032-008-9246-7](https://doi.org/10.1007/s11032-008-9246-7).
- Bueno, R.D., Borges, L.L., God, P.I.G., Piovesan, N.D., Teixeira, A.I., Cruz, C.D., and Barros, E.G. 2018. Quantification of anti-nutritional factors and their correlations with protein and oil in soybeans. *An. Acad. Bras. Ciênc.* **90**: 205–217. doi:[10.1590/0001-3765201820140465](https://doi.org/10.1590/0001-3765201820140465).
- Cardinal, A.J., and Burton, J.W. 2007. Correlations between palmitate content and agronomic traits in soybean populations segregating for the *fap1*, *fapnc*, and *fan* alleles. *Crop Sci.* **47**: 1804–1812. doi:[10.2135/cropsci2006.09.0577](https://doi.org/10.2135/cropsci2006.09.0577).
- Cicek, M.S., Chen, P., Saghai Maroof, M., and Buss, G.R. 2006. Interrelationships among agronomic and seed quality traits in an interspecific soybean recombinant inbred population. *Crop Sci.* **46**: 1253–1259. doi:[10.2135/cropsci2005.06-0162](https://doi.org/10.2135/cropsci2005.06-0162).
- Cober, E.R., and Morrison, M.J. 2010. Regulation of seed yield and agronomic characters by photoperiod sensitivity and growth habit genes in soybean. *Theor. Appl. Genet.* **120**: 1005–1012. doi:[10.1007/s00122-009-1228-6](https://doi.org/10.1007/s00122-009-1228-6). PMID: 20012856.
- Dhungana, S.K., Kulkarni, K.P., Kim, M., Ha, B.-K., Kang, S. Song, J.T., et al. 2017. Environmental stability and correlation of soybean seed starch with protein and oil contents. *Plant Breed. Biotechnol.* **5**: 293–303. doi:[10.9787/PBB.2017.5.4.293](https://doi.org/10.9787/PBB.2017.5.4.293).
- Egli, D., Fraser, J., Leggett, J., and Poneleit, C. 1981. Control of seed growth in soya beans [*Glycine max* (L.) Merrill]. *Ann. Bot.* **48**: 171–176. doi:[10.1093/oxfordjournals.aob.a086110](https://doi.org/10.1093/oxfordjournals.aob.a086110).
- Eskandari, M., Cober, E.R., and Rajcan, I. 2013. Genetic control of soybean seed oil: II. QTL and genes that increase oil concentration without decreasing protein or with increased seed yield. *Theor. Appl. Genet.* **126**: 1677–1687. doi:[10.1007/s00122-013-2083-z](https://doi.org/10.1007/s00122-013-2083-z). PMID: 23536049.
- Fang, C., Ma, Y., Wu, S., Liu, Z., Wang, Z. Yang, R., et al. 2017. Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.* **18**: 1–14. doi:[10.1186/s13059-017-1289-9](https://doi.org/10.1186/s13059-017-1289-9). PMID: 28077169.
- Fehr, W., Caviness, C., Burmood, D., and Pennington, J. 1971. Stage of development descriptions for soybeans, *Glycine max* (L.) Merrill. *Crop Sci.* **11**: 929–931. doi:[10.2135/cropsci1971.0011183X001100060051X](https://doi.org/10.2135/cropsci1971.0011183X001100060051X).
- Gillman, J.D., Stacey, M.G., Cui, Y., Berg, H.R., and Stacey, G. 2014. Deletions of the SACP-D locus elevate seed stearic acid levels but also result in fatty acid and morphological alterations in nitrogen fixing nodules. *BMC Plant Biol.* **14**: 1–19. doi:[10.1186/1471-2229-14-143](https://doi.org/10.1186/1471-2229-14-143). PMID: 24387633.
- Gong, Q.-C., Yu, H.-X., Mao, X.-R., Qi, H.-D., Yan, S. Xiang, W., et al. 2018. Meta-analysis of soybean amino acid QTLs and candidate gene mining. *J. Integr. Agric.* **17**: 1074–1084. doi:[10.1016/S2095-3119\(17\)61783-0](https://doi.org/10.1016/S2095-3119(17)61783-0).
- Hacham, Y., Song, L., Schuster, G., and Amir, R. 2007. Lysine enhances methionine content by modulating the expression of S-adenosylmethionine synthase. *Plant J.* **51**: 850–861. doi:[10.1111/j.1365-3113X.2007.03184.x](https://doi.org/10.1111/j.1365-3113X.2007.03184.x). PMID: 17617175.
- Hannah, M.A., Zuther, E., Buchel, K., and Heyer, A.G. 2006. Transport and metabolism of raffinose family oligosaccharides in transgenic potato. *J. Exp. Bot.* **57**: 3801–3811. doi:[10.1093/jxb/eri152](https://doi.org/10.1093/jxb/eri152). PMID: 17050641.
- Hernández-Sebastià, C., Marsolais, F., Saravitz, C., Israel, D., Dewey, R.E., and Huber, S.C. 2005. Free amino acid profiles suggest a possible role for asparagine in the control of storage-product accumulation in developing seeds of low- and high-protein soybean lines. *J. Exp. Bot.* **56**: 1951–1963. doi:[10.1093/jxb/eri191](https://doi.org/10.1093/jxb/eri191). PMID: 15911557.
- Hou, G., Ablett, G.R., Pauls, K.P., and Rajcan, I. 2006. Environmental effects on fatty acid levels in soybean seed oil. *J. Am. Oil Chem. Soc.* **83**: 759–763. doi:[10.1007/s11746-006-5011-4](https://doi.org/10.1007/s11746-006-5011-4).
- Hyten, D.L., Pantalone, V.R., Saxton, A.M., Schmidt, M.E., and Sams, C.E. 2004. Molecular mapping and identification of soybean fatty acid modifier quantitative trait loci. *J. Am. Oil Chem. Soc.* **81**: 1115–1118. doi:[10.1007/s11746-004-1027-z](https://doi.org/10.1007/s11746-004-1027-z).
- Jiang, G.-L., Chen, P., Zhang, J., Florez-Palacios, L., Zeng, A. Wang, X., et al. 2018. Genetic analysis of sugar composition and its relationship with protein, oil, and fiber in soybean. *Crop Sci.* **58**: 2413–2421. doi:[10.2135/cropsci2018.03.0173](https://doi.org/10.2135/cropsci2018.03.0173).
- Karikari, B., Li, S., Bhat, J.A., Cao, Y., Kong, J. Yang, J., et al. 2019. Genome-wide detection of major and epistatic effect QTLs for seed protein and oil content in soybean under multiple environments using high-density bin map. *Int. J. Mol. Sci.* **20**: 979. doi:[10.3390/ijms20040979](https://doi.org/10.3390/ijms20040979).
- Kastoori Ramamurthy, R., Jedlicka, J., Graef, G.L., and Waters, B.M. 2014. Identification of new QTLs for seed mineral, cysteine, and methionine concentrations in soybean [*Glycine max* (L.) Merr.]. *Mol. Breed.* **34**: 431–445. doi:[10.1007/s11032-014-0045-z](https://doi.org/10.1007/s11032-014-0045-z).
- Kim, K.-S., Diers, B., Hyten, D., Mian, R., Shannon, J., and Nelson, R. 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](https://doi.org/10.1007/s00122-012-1944-1). PMID: 22869284.
- Kumar, V., Rani, A., Goyal, L., Dixit, A.K., Manjaya, J., Dev, J., and Swamy, M. 2010. Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by genotype and growing location. *J. Agric. Food Chem.* **58**: 5081–5085. doi:[10.1021/jf903141s](https://doi.org/10.1021/jf903141s). PMID: 20353171.
- La, T., Large, E., Taliercio, E., Song, Q., Gillman, J.D. Xu, D., et al. 2019. Characterization of select wild soybean accessions in the USDA germplasm collection for seed composition and agronomic traits. *Crop Sci.* **59**: 233–251. doi:[10.2135/cropsci2017.08.0514](https://doi.org/10.2135/cropsci2017.08.0514).
- Lee, S., Van, K., Sung, M., Nelson, R., LaMantia, J., McHale, L.K., and Mian, M. 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](https://doi.org/10.1007/s00122-019-03304-5). PMID: 30806741.
- Li, X., Tian, R., Kamala, S., Du, H., Li, W., Kong, Y., and Zhang, C. 2018. Identification and verification of pleiotropic QTL controlling multiple amino acid contents in soybean seed. *Euphytica*, **214**: 1–14. doi:[10.1007/s10681-018-2170-y](https://doi.org/10.1007/s10681-018-2170-y).
- Li, Y.-H., Reif, J.C., Ma, Y.-S., Hong, H.-L., Liu, Z.-X., Chang, R.-Z., and Qiu, L.-J. 2015. Targeted association mapping demonstrating the complex molecular genetics of fatty acid formation in soybean. *BMC Genomics*, **16**: 1–13. doi:[10.1186/s12864-015-2049-4](https://doi.org/10.1186/s12864-015-2049-4). PMID: 25553907.
- Li, Y.-S., Du, M., Zhang, Q.-Y., Wang, G.-H., Hashemi, M., and Liu, X.-B. 2012. Greater differences exist in seed protein, oil, total soluble sugar and sucrose content of vegetable soybean genotypes [*Glycine max* (L.) Merrill] in Northeast China. *Aust. J. Crop Sci.* **6**: 1681–1686.
- Ma, Y., Ma, W., Hu, D., Zhang, X., Yuan, W. He, X., et al. 2019. QTL mapping for protein and sulfur-containing amino acid contents using a high-density bin-map in soybean (*Glycine max* L. Merr.). *J. Agric. Food Chem.* **67**: 12313–12321. doi:[10.1021/acs.jafc.9b04497](https://doi.org/10.1021/acs.jafc.9b04497). PMID: 31618030.
- Maestri, D.M., Guzmán, G.A., and Giorda, L.M. 1998. Correlation between seed size, protein and oil contents, and fatty acid composition in soybean genotypes. *Grasas Aceites*, **49**: 450–453. doi:[10.3989/gya.1998.v49i5-6.757](https://doi.org/10.3989/gya.1998.v49i5-6.757).
- Medic, J., Atkinson, C., and Hurburgh, C.R. 2014. Current knowledge in soybean composition. *J. Am. Oil Chem. Soc.* **91**: 363–384. doi:[10.1007/s11746-013-2407-9](https://doi.org/10.1007/s11746-013-2407-9).
- 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](https://doi.org/10.1007/s00122-009-1227-7). PMID: 20012262.



- Panthee, D., Pantalone, V., and Saxton, A. 2006a. Modifier QTL for fatty acid composition in soybean oil. *Euphytica*, **152**: 67–73. doi:[10.1007/s10681-006-9179-3](https://doi.org/10.1007/s10681-006-9179-3).
- Panthee, D., Pantalone, V., Sams, C., Saxton, A., West, D., Orf, J.H., and Killam, A. 2006b. Quantitative trait loci controlling sulfur containing amino acids, methionine and cysteine, in soybean seeds. *Theor. Appl. Genet.* **112**: 546–553. doi:[10.1007/s00122-005-0161-6](https://doi.org/10.1007/s00122-005-0161-6).
- 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](https://doi.org/10.1111/pbi.12929). PMID: [29618164](https://pubmed.ncbi.nlm.nih.gov/29618164/).
- Paul, M.J., and Foyer, C.H. 2001. Sink regulation of photosynthesis. *J. Exp. Bot.* **52**: 1383–1400. doi:[10.1093/jexbot/52.360.1383](https://doi.org/10.1093/jexbot/52.360.1383). PMID: [11457898](https://pubmed.ncbi.nlm.nih.gov/11457898/).
- 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](https://doi.org/10.1007/s00122-013-2094-9). PMID: [23595202](https://pubmed.ncbi.nlm.nih.gov/23595202/).
- Rubel, A., Rinne, R., and Canvin, D. 1972. Protein, oil, and fatty acid in developing soybean seeds. *Crop Sci.* **12**: 739–741. doi:[10.2135/cropsci1972.0011183X001200060006x](https://doi.org/10.2135/cropsci1972.0011183X001200060006x).
- Shaul, O., and Galili, G. 1992. Increased lysine synthesis in tobacco plants that express high levels of bacterial dihydrodipicolinate synthase in their chloroplasts. *Plant J.* **2**: 203–209. doi:[10.1111/j.1365-313X.1992.00203.x](https://doi.org/10.1111/j.1365-313X.1992.00203.x).
- Vaughn, J.N., Nelson, R.L., Song, Q., Cregan, P.B., and Li, Z. 2014. The genetic architecture of seed composition in soybean is refined by genome-wide association scans across multiple populations. *G3: Genes Genomes Genetics*, **4**: 2283–2294. doi:[10.1534/g3.114.013433](https://doi.org/10.1534/g3.114.013433).
- Wang, X., Jiang, G.-L., Song, Q., Cregan, P.B., Scott, R.A. Zhang, J., et al. 2015. Quantitative trait locus analysis of seed sulfur-containing amino acids in two recombinant inbred line populations of soybean. *Euphytica*, **201**: 293–305. doi:[10.1007/s10681-014-1223-0](https://doi.org/10.1007/s10681-014-1223-0).
- Wang, Y., Chen, P., and Zhang, B. 2014. Quantitative trait loci analysis of soluble sugar contents in soybean. *Plant Breed.* **133**: 493–498. doi:[10.1111/pbr.12178](https://doi.org/10.1111/pbr.12178).
- Warrington, C.V., Abdel-Haleem, H., Hyten, D., Cregan, P., Orf, J. Killam, A., et al. 2015. QTL for seed protein and amino acids in the Benning × Danbaekkong soybean population. *Theor. Appl. Genet.* **128**: 839–850. doi:[10.1007/s00122-015-2474-4](https://doi.org/10.1007/s00122-015-2474-4). PMID: [25673144](https://pubmed.ncbi.nlm.nih.gov/25673144/).
- Xie, D., Han, Y., Zeng, Y., Chang, W., Teng, W., and Li, W. 2012. SSR- and SNP-related QTL underlying linolenic acid and other fatty acid contents in soybean seeds across multiple environments. *Mol. Breed.* **30**: 169–179. doi:[10.1007/s11032-011-9607-5](https://doi.org/10.1007/s11032-011-9607-5).
- Yan, W., and Rajcan, I. 2002. Biplot analysis of test sites and trait relations of soybean in Ontario. *Crop Sci.* **42**: 11–20. doi:[10.2135/cropsci2002.1100](https://doi.org/10.2135/cropsci2002.1100). PMID: [11756248](https://pubmed.ncbi.nlm.nih.gov/11756248/).
- Yoosefzadeh-Najafabadi, M., Eskandari, M., Torabi, S., Torkamaneh, D., Tulpan, D., and Rajcan, I. 2022. Machine-learning-based genome-wide association studies for uncovering QTL underlying soybean yield and its components. *Int. J. Mol. Sci.* **23**: 5538. doi:[10.3390/ijms23105538](https://doi.org/10.3390/ijms23105538). PMID: [35628351](https://pubmed.ncbi.nlm.nih.gov/35628351/).
- Zeng, A., Chen, P., Shi, A., Wang, D., Zhang, B. Orazaly, M., et al. 2014. Identification of quantitative trait loci for sucrose content in soybean seed. *Crop Sci.* **54**: 554–564. doi:[10.2135/cropsci2013.01.0036](https://doi.org/10.2135/cropsci2013.01.0036).
- Zhang, D., Lü, H., Chu, S., Zhang, H., Zhang, H. Yang, Y., et al. 2017. The genetic architecture of water-soluble protein content and its genetic relationship to total protein content in soybean. *Sci. Rep.* **7**: 1–13. PMID: [28127051](https://pubmed.ncbi.nlm.nih.gov/28127051/).
- Zhang, J., Wang, X., Lu, Y., Bhusal, S.J., Song, Q. Cregan, P.B., et al. 2018. Genome-wide scan for seed composition provides insights into soybean quality improvement and the impacts of domestication and breeding. *Mol. Plant*, **11**: 460–472. doi:[10.1016/j.molp.2017.12.016](https://doi.org/10.1016/j.molp.2017.12.016). PMID: [29305230](https://pubmed.ncbi.nlm.nih.gov/29305230/).
- Zhao, X., Jiang, H., Feng, L., Qu, Y., Teng, W. Qiu, L., et al. 2019. Genome-wide association and transcriptional studies reveal novel genes for unsaturated fatty acid synthesis in a panel of soybean accessions. *BMC Genomics*, **20**: 1–16. doi:[10.1186/s12864-019-5449-z](https://doi.org/10.1186/s12864-019-5449-z). PMID: [30606130](https://pubmed.ncbi.nlm.nih.gov/30606130/).
- Zhou, Z., Lakhssassi, N., Cullen, M.A., El Baz, A., Vuong, T.D., Nguyen, H.T., and Meksem, K. 2019. Assessment of phenotypic variations and correlation among seed composition traits in mutagenized soybean populations. *Genes*, **10**: 975. doi:[10.3390/genes10120975](https://doi.org/10.3390/genes10120975).