



OPEN

Historical trend on seed amino acid concentration does not follow protein changes in soybeans

Andre Froes de Borja Reis¹✉, Santiago Tamagno^{1,4}, Luiz H. Moro Rosso¹, Osler A. Ortez², Seth Naeve³ & Ignacio A. Ciampitti¹✉

Soybean [*Glycine max* (L.) Merr.] is the most important oilseed crop for animal industry due to its high protein concentration and high relative abundance of essential and non-essential amino acids (AAs). However, the selection for high-yielding genotypes has reduced seed protein concentration over time, and little is known about its impact on AAs. The aim of this research was to determine the genetic shifts of seed composition for 18 AAs in 13 soybean genotypes released between 1980 and 2014. Additionally, we tested the effect of nitrogen (N) fertilization on protein and AAs trends. Soybean genotypes were grown in field conditions during two seasons under a control (0 N) and a N-fertilized treatment receiving 670 kg N ha⁻¹. Seed yield increased 50% and protein decreased 1.2% comparing the oldest and newest genotypes. The application of N fertilizer did not significantly affect protein and AAs concentrations. Leucine, proline, cysteine, and tryptophan concentrations were not influenced by genotype. The other AAs concentrations showed linear rates of decrease over time ranging from -0.021 to -0.001 g kg⁻¹ year⁻¹. The shifts of 11 AAs (some essentials such as lysine, tryptophan, and threonine) displayed a relative-to-protein increasing concentration. These results provide a quantitative assessment of the trade-off between yield improvement and seed AAs concentrations and will enable future genetic yield gain without overlooking seed nutritional value.

Soybean [*Glycine max* (L.) Merr.] is a major oilseed and protein crop which is responsible for roughly 70% of the world's plant-based meal¹. Soybean meal is the by-product of oil extraction and provides a high-quality protein for animal feed and other uses². The quality of the protein is defined by the relative constitution of amino acids (AAs) and the profile of essential and non-essential AAs^{3,4}. From a production viewpoint, improvements in agronomic management and breeding have led to increases in soybean yields⁵ in parallel with the global demand for this crop. A main challenge in soybean breeding has been to improve yields while maintaining market standard for protein concentration (e.g., soybean meal levels) despite the consistent trade-off between seed yield and protein concentration^{6,7}.

Average protein concentration in current commercially available soybean genotypes ranges between 33 and 39% (on 130 g kg⁻¹ moisture content)⁸. Commonly, 18 AAs (essential and non-essential) are reported as soybeans protein constituents⁹. Asparagine and glutamine are the most abundant, corresponding to approximately 12% and 18%, respectively, of total AAs; however, they are considered non-essential for humans and monogastric animals¹⁰. The essential AAs constitute a relatively smaller proportion of seeds including lysine, threonine, tryptophan, isoleucine, leucine, histidine, phenylalanine, valine, and the sulfur amino acids cysteine and methionine^{11,12}.

The relationship between seed protein and AA concentration varies according to each AA. For instance, lysine, methionine, cysteine, tryptophan, and threonine are negatively correlated with seed protein concentration, whereas arginine and glutamic acid increase with seed protein concentration^{9,13}. In China, significant negative trends for glutamic acid, histidine, and arginine were reported in soybean genotypes released between 1923 and 2007¹⁴. However, the study did not contemplate the effect of N on seed composition and the changes of AAs relative to the modifications in seed protein concentration.

At the plant level, soybean seed composition is the result of complex genotype and environment interactions^{15,16}. During the seed-filling period, protein synthesis occurs in the developing seeds on the basis of carbon and N compounds provided by the mother plant. Seed N demand represents about 75% of the total plant

¹Department of Agronomy, Kansas State University, Manhattan, KS, USA. ²Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE, USA. ³Department of Agronomy and Plant Genetics, University of Minnesota, Saint Paul, MN, USA. ⁴Present address: Department of Plant Sciences, University of California, Davis, CA, USA. ✉email: andrefbr@ksu.edu; ciampitti@ksu.edu

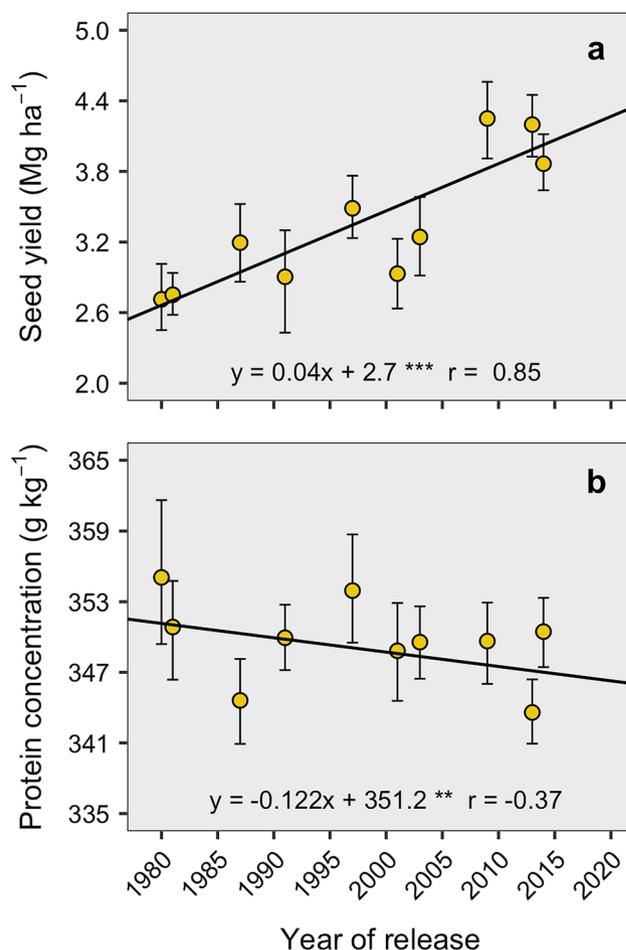


Figure 1. Relationship between seed yield (13 g kg⁻¹ of moisture content) (a), protein concentration (b), and year of release of 13 genotypes released from 1980 to 2014 period. Each point represents the mean of both N treatments and genotypes (if more than one per year). Solid black lines denote the best fitted linear model. Asterisks indicate significance of the coefficient: *** $P < 0.001$; ** $P < 0.01$. The r value represents the Pearson correlation.

N uptake during the season¹⁷, being mostly remobilized from vegetative organs¹⁸ and concurrently assimilated from soil mineral N and biological N fixation. Hence, N accumulated prior to seed formation will be a predominant source for protein synthesis in seeds¹⁹. The effect of N fertilization on seed protein concentration has been studied, but response seems to be erratic¹⁶ and historical trends in protein dilution could not be reversed even in heavily N-fertilized environments²⁰. However, the degree to which individual AAs respond to N application across historical genotypes remains unknown.

Therefore, considering a historical set of soybean genotypes released between 1980 and 2014 in the United States, the aims of this research were as follows: (i) determine the genetic gain of 18 AAs concentrations in soybean seeds, (ii) evaluate the response in seed AA profile to N fertilization, and (iii) compare the AAs genetic gain by clustering those AAs presenting similar shifts (in both absolute and relative terms) relative to seed weight and protein. This information contributes to our ability to understand the magnitude and potential determinants of seed composition changes, therefore enabling future investigations for a more effective selection of seed nutritional value in soybeans.

Results

Seed yield and protein genetic gain in historical genotypes. The absolute genetic gain was given by the relationship between the crop trait and the genotype year of release. The mean crop yield genetic gain was 0.04 Mg ha⁻¹ year⁻¹ (Fig. 1a) regardless of the N fertilization treatment (Table 1). Seed yield ranged between 2.7 and 4.1 Mg ha⁻¹ with an estimated yield increase of approximately 50% from 1980 to 2014. Protein absolute genetic gain decreased at a rate of 0.122 g kg⁻¹ year⁻¹ (Fig. 1b). Similar to yield, seed protein concentration was not affected either by the N fertilization nor its interaction with year of release (Table 1). The average protein concentration was 349 g kg⁻¹, with an overall reduction of 1.18% when considering the entire evaluation period, and a total decrease of -4.15 g kg⁻¹ (Fig. 1b).

Crop Trait	N fertilization (N)	Genotype year of release (YR)	N x YR
Yield	0.33 ^{ns}	<0.001***	0.31 ^{ns}
Protein	0.26 ^{ns}	<0.01**	0.83 ^{ns}
Glutamic acid	0.39 ^{ns}	<0.01**	0.66 ^{ns}
Aspartic acid	0.28 ^{ns}	<0.01**	0.72 ^{ns}
Leucine	0.11 ^{ns}	0.03*	0.87 ^{ns}
Arginine	0.40 ^{ns}	<0.01*	0.67 ^{ns}
Lysine	0.22 ^{ns}	<0.01**	0.86 ^{ns}
Phenylalanine	0.13 ^{ns}	<0.01*	0.78 ^{ns}
Proline	0.12 ^{ns}	0.11 ^{ns}	0.95 ^{ns}
Isoleucine	0.24 ^{ns}	0.02*	0.99 ^{ns}
Valine	0.24 ^{ns}	0.02*	0.98 ^{ns}
Serine	0.27 ^{ns}	<0.01**	0.97 ^{ns}
Glycine	0.15 ^{ns}	0.01*	0.72 ^{ns}
Alanine	0.12 ^{ns}	0.02*	0.84 ^{ns}
Threonine	0.13 ^{ns}	0.02*	0.86 ^{ns}
Tyrosine	0.27 ^{ns}	0.02*	0.65 ^{ns}
Histidine	0.17 ^{ns}	<0.01**	0.98 ^{ns}
Cysteine	0.53 ^{ns}	0.17 ^{ns}	0.88 ^{ns}
Methionine	0.46 ^{ns}	0.02*	0.52 ^{ns}
Tryptophan	0.70 ^{ns}	0.24 ^{ns}	0.35 ^{ns}

Table 1. Analysis of variance *F*-test probabilities for nitrogen fertilization and genotype year of release affecting soybean traits.

Amino acid genetic gain over time. The majority of the AAs presented a significant decrease in absolute genetic gain (Table 1, Fig. 2), with the exception of leucine, proline, cysteine, and tryptophan (Fig. 2c, i, p, r). Similarly to yield and protein results, the N fertilization did not affect the AA trends over time. Glutamic acid displayed a rate of $-0.021 \text{ g kg}^{-1} \text{ year}^{-1}$, with a decrease of 1.22% over the 1980 to 2014 timeframe (Fig. 2a). The concentration of aspartic acid decreased by 1.07% at a rate of $-0.012 \text{ g kg}^{-1} \text{ year}^{-1}$ (Fig. 2b). The other non-essential AAs (arginine, alanine, serine, glycine, and tyrosine) followed the same overall decreasing trend. Arginine concentration decreased by 1.43% in the selected time period, with a rate of $-0.01 \text{ g kg}^{-1} \text{ year}^{-1}$ (Fig. 2d), whereas alanine concentration decreased 0.8% with a rate of $-0.003 \text{ g kg}^{-1} \text{ year}^{-1}$ (Fig. 2j). Serine and glycine concentrations decreased at a rate of $-0.005 \text{ g kg}^{-1} \text{ year}^{-1}$ and $0.004 \text{ g kg}^{-1} \text{ year}^{-1}$, respectively, (Fig. 2k, l) with an overall reduction of 1.21% and 0.88%, respectively, over the evaluated period of time.

In the essential AAs group, lysine concentration decreased by 0.92% for the evaluated period at a rate of $-0.006 \text{ g kg}^{-1} \text{ year}^{-1}$ (Fig. 2e). Phenylalanine decreased at a rate of $-0.005 \text{ g kg}^{-1} \text{ year}^{-1}$ with concentration ranging from 17.54 to 17.37 g kg^{-1} (Fig. 2f). Valine decreased by 0.81% with a rate of $-0.004 \text{ g kg}^{-1} \text{ year}^{-1}$ (Fig. 2g). Isoleucine, threonine, and histidine showed rates of -0.004 , -0.003 , and $-0.003 \text{ g kg}^{-1} \text{ year}^{-1}$ with an average decrease of 0.85%, 1.03%, and 1.15%, respectively (Fig. 2h, m, o). Both leucine and tryptophan concentrations remained steady across genotypes with averages of 25.5 and 3.8 g kg^{-1} , respectively (Fig. 2c, r). For the sulfur amino acids, only methionine concentration was influenced by genotype's year of release. Genetic gain for methionine was $-0.001 \text{ g kg}^{-1} \text{ year}^{-1}$ in seed concentration, with the modern genotype attaining 0.71% less methionine than the oldest genotype evaluated in this study (Fig. 2q). Cysteine was constant in soybean seeds across the years of release with a mean concentration of 5.6 g kg^{-1} (Fig. 2p).

Shifts of seed composition. In order to compare the different slope magnitudes between protein and all AAs, each individual slope of absolute genetic gain was divided by the last fitted value (Fig. 3), thereby resulting in a relative gain. The estimated relative genetic gain of protein was $-0.035\% \text{ year}^{-1}$ (Fig. 3a). The confidence intervals of leucine, proline, cysteine and tryptophan overlapped the 0% genetic gain threshold, and therefore, were considered non-significant (empty points). The remaining AAs all decreased in concentration over time (black solid points). The magnitude of gain is provided here from the largest negative gain to the smallest in the following order: arginine (-0.044%), glutamic acid (-0.037%), serine (-0.035%), aspartic acid (-0.032%), histidine (-0.032%), methionine (-0.028%), phenylalanine (-0.026%), lysine (-0.025%), glycine (-0.025%), isoleucine (-0.025%), valine (-0.024%), tyrosine (-0.021%), alanine (-0.021%), and threonine (-0.020%) (Fig. 3a).

Not all AAs presented a negative trend of the same magnitude of seed protein as portrayed by the relative-to-protein genetic gain (Fig. 3b). This relationship is presented by the individual AA relative concentration to protein (%) and genotype year of release. Some AAs presented a less than proportionate reduction relative-to-protein and thus, those AAs were clustered as less negatively affected by the overall decrease in protein (black solid points). Within this cluster, we found the following AAs: phenylalanine, lysine, glycine, isoleucine, valine, leucine, tyrosine, alanine, threonine, proline, and tryptophan. In a second cluster, we found AAs presenting a

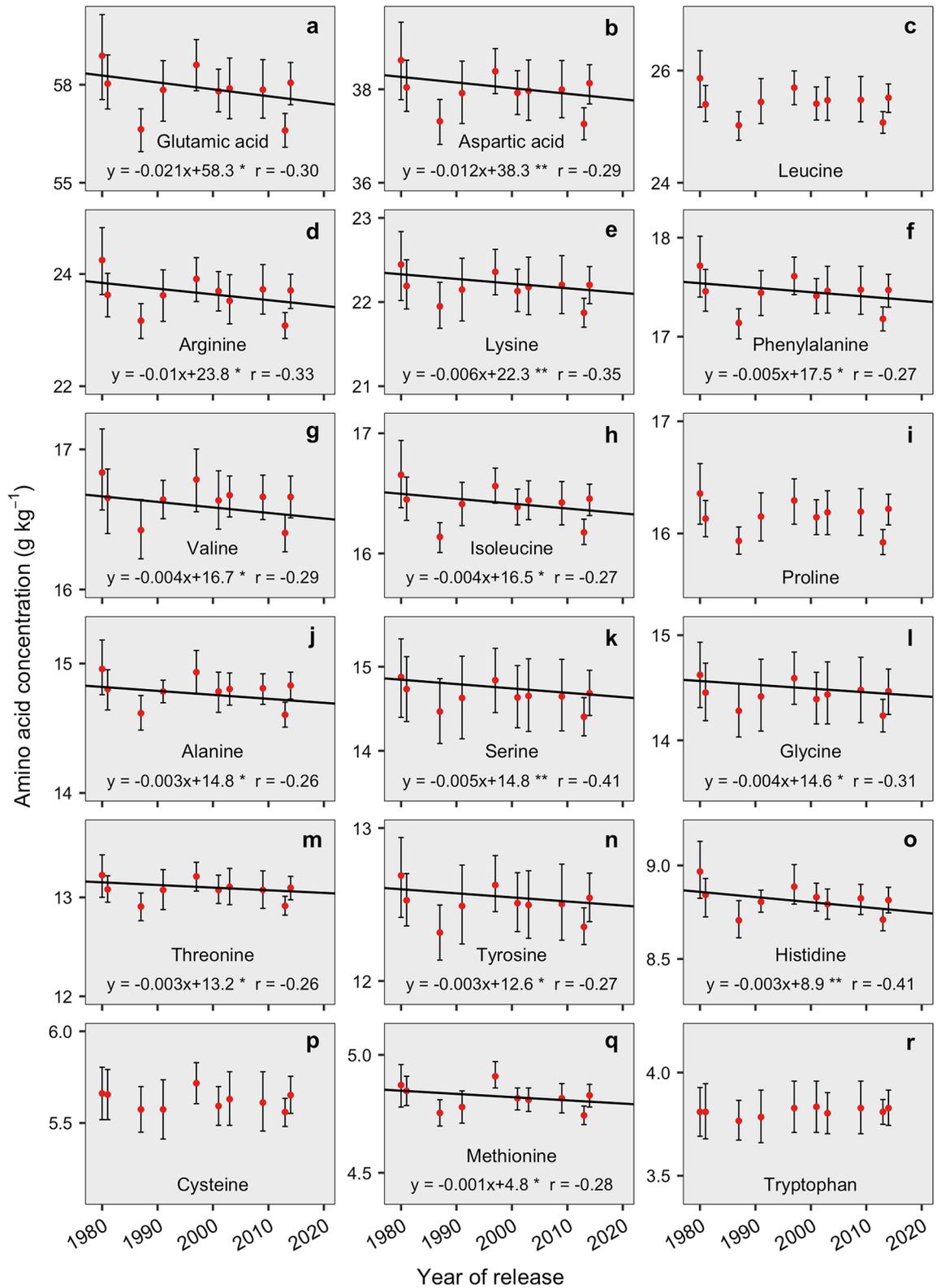


Figure 2. Relationship between glutamic acid (a), aspartic acid (b), leucine (c), arginine (d), lysine (e), phenylalanine (f), valine (g), isoleucine (h), proline (i), alanine (j), serine (k), glycine (l), threonine (m), tyrosine (n), histidine (o), cysteine (p), methionine (q), tryptophan (r), and year of release of 13 soybean genotypes released from 1980 to 2014. Each point represents the average between the two N treatments and genotypes (if more than one per year). Solid black lines denote the AA genetic gain fitted model. Asterisks indicate significance of the coefficient: *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$. The r value represents the Pearson correlation.

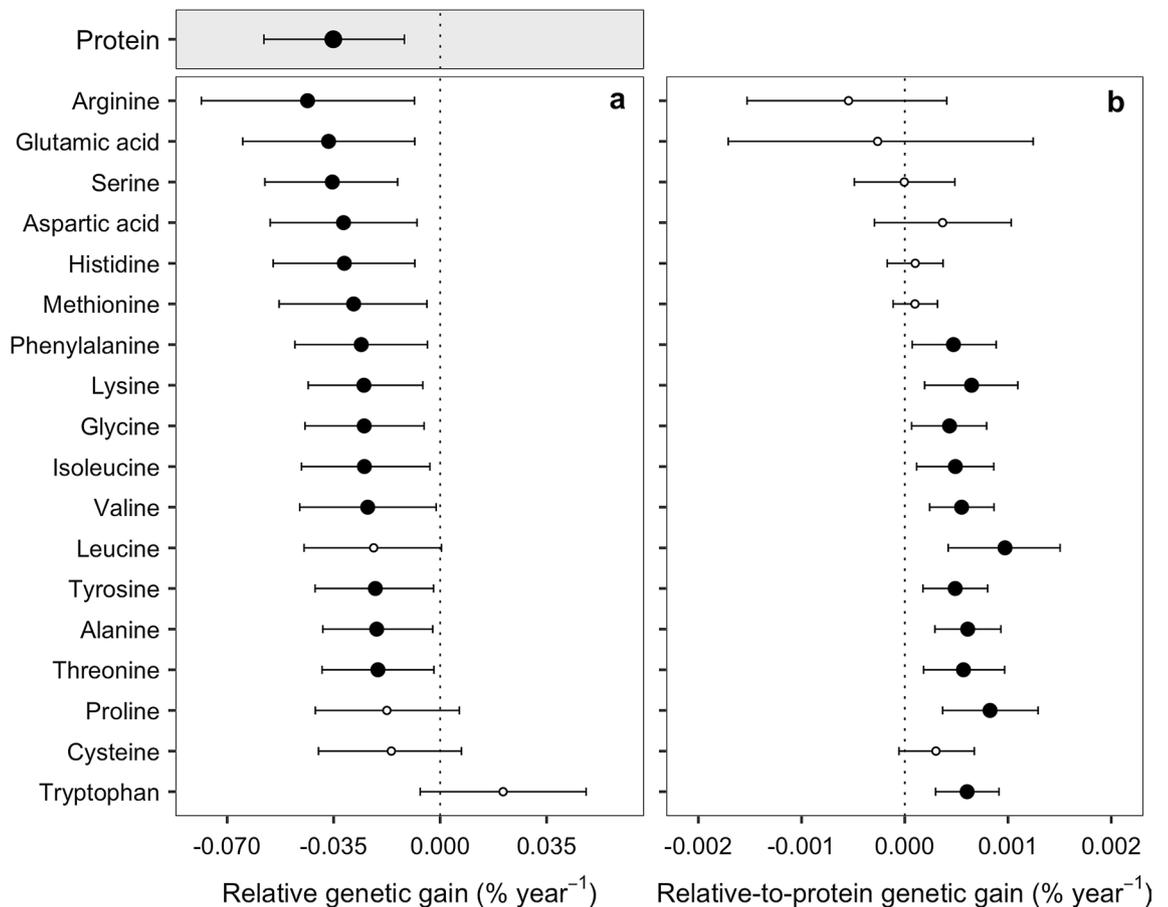


Figure 3. Relative genetic gain of protein and amino acid concentrations (a) and the relative-to-protein genetic gain of amino acid concentrations (b), calculated from 13 soybean genotypes released from 1980 to 2014 period. Points represent the medians of bootstrapped distribution with their respective 95% confidence intervals (small horizontal lines). Black solid points represent the AAs with significant relative gains, whereas the empty points are non-significant (95% CI including zero).

decreasing relative-to-protein concentration but not significantly different from zero. They are arginine, glutamic acid, serine, aspartic acid, histidine, methionine, and cysteine (empty points) (Fig. 3b).

Relationship between protein and amino acid concentration in historical genotypes. The estimated correlation between an AAs relative concentration (%) and protein concentration (g kg^{-1}) was significant ($P < 0.05$) in 10 out of 18 AA in seeds from historical soybeans genotypes (Fig. 4). Glutamic acid and arginine were the only AAs to show a positive slope (Fig. 4a, d). The rate of increase was 0.013 and 0.010% for glutamic acid and arginine, respectively. The remaining 8 AAs with a significant relationship to protein concentration demonstrated a negative slope. This was the case for lysine (-0.005%), valine (-0.004%), proline (-0.004%), alanine (-0.003%), glycine (-0.003%), threonine (-0.003%), tyrosine (-0.002%), and tryptophan (-0.003%). Aspartic acid, leucine, phenylalanine, isoleucine, serine, histidine, cysteine, and methionine had no significant relationship to protein concentration in the historical genotypes.

Discussion

Our results highlight the historical trends (1980–2014) in seed AAs concentrations. Yield improvement and protein reduction were within the range reported in the literature^{5,21–25}. In addition, as previously documented for soybeans in maturity group III²⁰, the rate of protein reduction over years was unchanged by the application of N fertilizer (670 kg ha^{-1}). Fourteen of the 18 AAs analyzed were present in lower concentrations in more recently released genotypes (Figs. 2, 3a). The shifts of the most abundant AAs in soybean (glutamic acid and aspartic acid) were in the same range as the protein reduction rate (Fig. 3a). Alternatively, the concentrations of the essential AAs lysine and threonine increased relative to protein which may represent an impact on the nutritional quality of soybean meal⁴. Although a majority of the AAs decreased in absolute values, 11 AAs increased in concentrations relative-to-protein (Fig. 3b), including leucine, isoleucine, histidine, phenylalanine, and valine, which are essential AAs for animal nutrition^{12,26}. Therefore, breeding efforts to develop high protein genotypes should consider the underlying impact on AAs and the potential impact on the nutritional value of the seeds^{6,27}.

Addition of N did not affect protein or AA shifts over time (Table 1). A significant effect of N fertilization on protein is more likely in environments with poor N supply such as greenhouse conditions¹⁶ or low activity

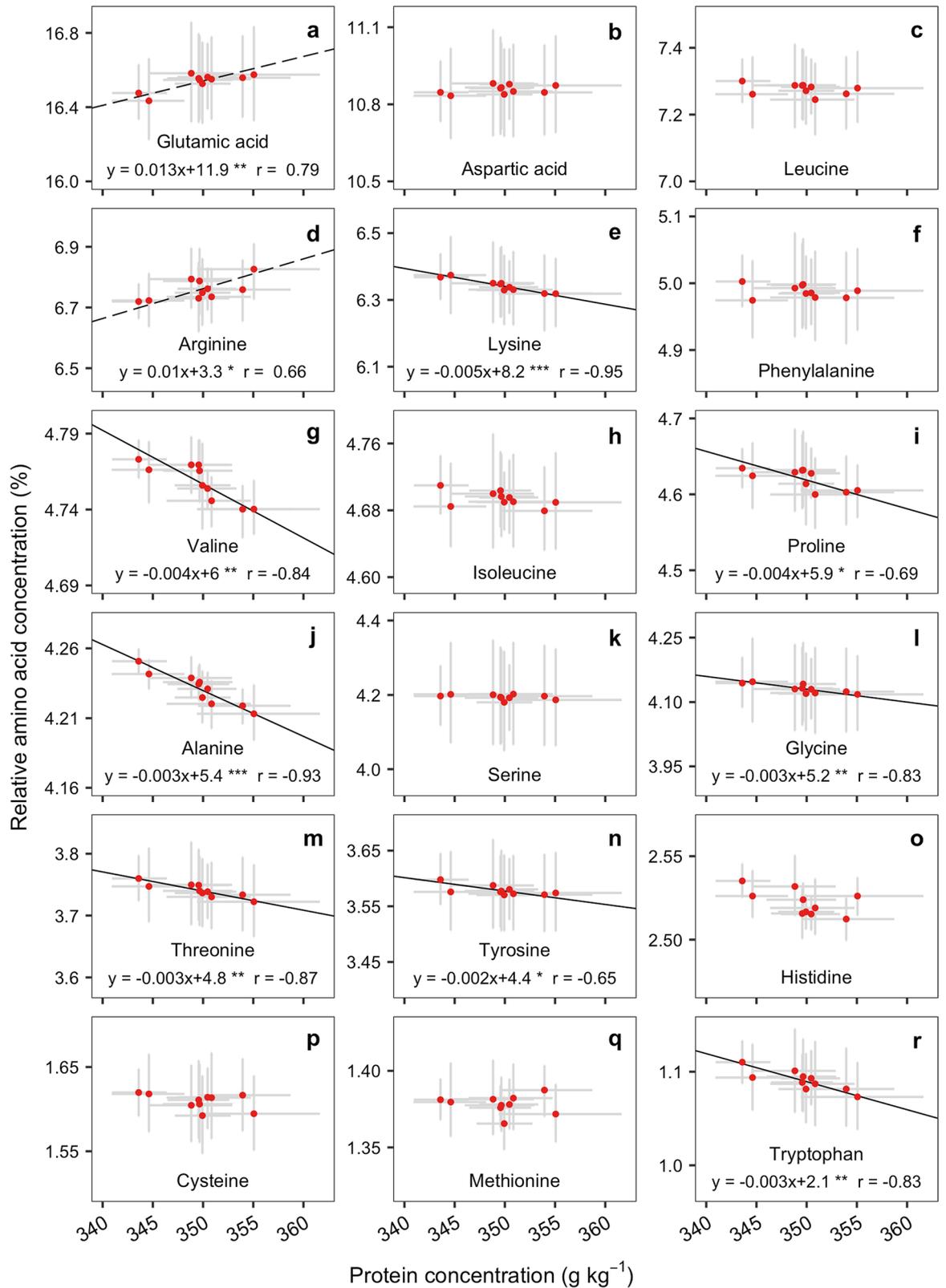


Figure 4. Relationship between glutamic acid (a), aspartic acid (b), leucine (c), arginine (d), lysine (e), phenylalanine (f), valine (g), isoleucine (h), proline (i), alanine (j), serine (k), glycine (l), threonine (m), tyrosine (n), histidine (o), cysteine (p), methionine (q), tryptophan (r), and protein concentration in seed of genotypes released from 1980 to 2014. Each point represents the average amino acid concentration between the two N treatments and genotypes (if more than one per year). Solid black line indicates the fitted model when the slope is negative, whereas dashed line indicates models with positive slope ($P < 0.05$). Asterisks indicate significance of the coefficient: *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$. The r value represents the Pearson correlation.

of biological nitrogen fixation²⁸. In field studies, N biological fixation resulting from seed inoculation or indigenous soil rhizobia infection, may provide sufficient N to support high crop performance^{29–31}. For the tested yield levels, our results indicate the inability of N fertilization to reverse the decline on soybean seed protein and AAs concentrations. A similar outcome was presented by Wilson et al. (2013), documenting a protein decline of $-0.25 \text{ g kg}^{-1} \text{ year}^{-1}$ for soybean genotypes released between 1923 and 2008 under contrasting N rates (zero vs. 560 kg N ha^{-1}). Regarding the AA profile, controlled condition studies support a positive relationship between supra-optimal N and essential³² or storage AAs³³. For non-leguminous crops, field studies have validated the concept of N application enhancing AAs concentrations^{34–36}. For soybeans, however, only a few studies were carried out in the field with N fertilization, and the results showed an increase in AAs concentrations only under low N availability²⁸ or associated with sulfur AAs reduction³⁷. These findings were not observed in our results, therefore highlighting a lack of effect of non-limiting N supply for offsetting protein and AAs depression in historical soybean genotypes.

Additionally, protein concentration was shown to be a better predictor of changes in AAs over time, but only for a select few such as glutamic acid, arginine, lysine, valine, proline, alanine, glycine, threonine, tyrosine, and tryptophan (Fig. 4). Using protein as a predictor of the AA changes over time was previously reported⁹, but considering only modern genotypes rather than a historical set as presented in this current study. Our findings show similar relationships for some AA changes such as glutamic acid and arginine relative-to-protein genetic gain (Fig. 3b), but other AAs did not exactly follow the trend of protein, e.g., aspartic acid, leucine, phenylalanine, isoleucine, serine, histidine, cysteine, and, methionine (Fig. 4). To date, there are no reported predictive models describing the entire AA profile as a function of protein concentration (as a reference seed composition fraction). Therefore, establishing foundational prediction models for AAs will assist breeders and ultimately growers in delivering soybean genotypes focusing on specific market demands.

Conclusions

This research explored the shifts in protein and AAs due to the genetic improvement of soybean genotypes from 1980 to 2014. These shifts were not driven by an increased N supply via inorganic N, as the N fertilization treatment did not change any trends for AAs concentrations. Similar negative rates, in absolute concentrations, were observed for some AAs such as arginine and glutamic acids but not for the rest of the AA profile relative-to-protein. Therefore, the concept of utilizing seed protein concentration genetic gain as an indicator of potential changes in AAs is not a valid rationale. Emerging areas of research focusing the genetic control of amino acids synthesis and its interaction to the environment will provide the foundation for improving seed traits either maintaining or improving the nutritional value of soybean.

Methods

Field conditions and experimental design. Two field experiments were conducted at the Kansas River Valley research station in Rossville, Kansas, United States ($39^{\circ}07' \text{ N}$; $95^{\circ}55' \text{ W}$) during the 2016 and 2017 growing seasons. The local weather is Dfa continental humid with hot and wet summers³⁸. Temperature in the 2016 and 2017 seasons averaged 22 and 23 °C, respectively. The seasonal precipitation was 450 mm in 2016 and 523 mm in 2017. The experimental area was kept primarily under rainfed conditions, although 345 mm and 221 mm (2016 and 2017, respectively) of water was supplemented as needed to avoid potential drought stress. The soil type was a Fluventic Hapludoll with the following chemical attributes at a 0–0.15 m depth: pH 6.9 (2016) and 7.3 (2017); organic matter (%): 2.2 (2016) and 1.3 (2017); N-nitrate (mg kg^{-1}): 3.0 (2016) and 2.7 (2017); Cation exchange capacity ($\text{cmol}_c \text{ dm}^{-3}$): 11 (2016) and 5.8 (2017); phosphorous Mehlich (mg kg^{-1}): 21 (2016) and 13 (2017); potassium (mg kg^{-1}): 153 (2016) and 90 (2017); calcium (mg kg^{-1}): 2074 (2016) and 951 (2017); magnesium (mg kg^{-1}): 202 (2016) and 95 (2017). The area is permanently maintained under a maize-soybean rotation.

The experimental design was a randomized complete block in split-plot arrangement with four replications in both seasons. The main-plot consisted of the N factor with two levels and the sub-plot was the genotype factor with 13 levels. The N treatments were N fertilization at a rate of 670 kg N ha^{-1} and the control without N (0 N). The N fertilizer was equally split at sowing, R1, and R3 phenological stages³⁹ as a side dressed application of liquid urea ammonium nitrate (N P K, 28-0-0). Seed inoculation was performed shortly before sowing with the application of 3×10^9 colony units of *Bradyrhizobium japonicum* per 1 kg of seeds.

The subplot consisted of 13 genotypes released between 1980 and 2014 with maturity groups (MG) ranging between 3.0 and 4.0. Genotypes, associated MG and release dates are as follow: P3981 (1980—MG 3.0), Williams 82 (1981—MG 3.0), 9391 (1987—MG 4.0), 9392 (1991—MG 3.8), P93B82 (1997—MG 3.8), 93B67 (2001—MG 3.9), 93M90 (2003—MG 3.0), P93Y92 (2009—MG 3.9), 94Y23 (2013—MG 4.0), P35T58R (2013—MG 3.0), P39T67R (2013—MG 4.0), P31T11R (2014—MG 3.1), and P34T43R (2014—MG 3.4). Planting dates were May 12 in 2016 and May 18 in 2017. The plot size was 10 m long by four 0.76 m rows. The experimental area was kept free of weeds, pests, and diseases.

Seed yield, protein and amino acids determination. At harvest maturity (R8), the two center rows in each plot were harvested with a plot combine, and the seed yield was adjusted to 130 g kg^{-1} water content basis. Approximately one kilogram of seed was sampled from each plot to measure seed composition. After seeds were dried to constant weight, the samples were ground to 0.1 mm final particle size. Protein and AAs concentrations were estimated with near-infrared spectroscopy (NIR) using the Perten DA7200 Feed Analyzer (Perten Instruments, Stockholm, Sweden). Briefly, the raw ground material was scanned between 1000 and 2500 nm wavelength and the reflectance normalized to a reference ceramic plate. The readings were subject to error removal due to an eventually uneven cup filling or sample size heterogeneity. The calibration between normalized reflec-

tance and AA concentration was cross-validated using standard samples analyzed by wet chromatography following the AOAC 982.30 method⁴⁰. The calibration curves were tested by root-mean-square error (RMSE). This method estimates the protein and 18 AAs concentrations (g kg^{-1}) corrected to water content. However, this method does not distinguish between asparagine and aspartate, or between glutamine and glutamate. Therefore, glutamic acid and aspartic acid forms were reported as the sum of their respective components.

The absolute genetic gain ($\text{g kg}^{-1} \text{ year}^{-1}$) was estimated by the regression of yield, protein, or each individual AA to genotype year of release. The relative genetic gain ($\% \text{ year}^{-1}$) was calculated to allow the comparison between amino acids and protein concentrations. Thus, the slope of the absolute genetic gain for each AA was divided by the most recent estimated concentration⁴¹ (Eq. 1).

$$\text{Relative genetic gain} (\% \text{ year}^{-1}) = 100 \times \left(\frac{\text{Absolute genetic gain} (\text{g kg}^{-1} \text{ year}^{-1})}{\text{Most recent estimated concentration} (\text{g kg}^{-1})} \right) \quad (1)$$

The relative-to-protein genetic gain ($\% \text{ year}^{-1}$) was determined by the relationship between the relative to protein concentration ratio (Eq. 2) with genotype year of release.

$$\text{Relative protein ratio} (\%) = 100 \times \frac{\text{Seed AA concentration} (\text{g kg}^{-1})}{\text{Protein AA concentration} (\text{g kg}^{-1})} \quad (2)$$

Finally, to investigate the correlation between AAs and protein regardless the year of release, AA concentrations relative-to-protein (Eq. 2) were tested against protein concentration.

Data analysis. We first tested the effect of N treatment by fitting two linear mixed models for each variable. The first model included the year of release, N treatment, and the interaction as fixed effect factors, and the second model included only the year of release as the fixed effect. As N treatment was not significant, the model with the lowest score for Akaike Information Criterion (AIC) was selected. The random factors included year, block nested in year, N treatment nested in the interaction of block with year, and genotype nested in the interaction of N treatment, block, and year. Models were fitted using the package “lme4”⁴² within the R software⁴³. Assumptions of normality and homogeneity of the residuals were checked and no transformation was required.

The Resampling with Replacement Bootstrap was used to estimate the slope coefficient and the empirical distribution of model estimators⁴⁴. A total of 5000 iterations were performed. All the distributions were summarized by the median, and the 2.5 and 97.5% percentiles were used as the boundaries of the 95% confidence intervals (CI), allowing statistical inference on the parameters⁴⁵. The Pearson correlation coefficient (r) was estimated from the variable median estimation for each year of release (means distribution). The relative-to-protein genetic gain and relative genetic gain were empirically clustered using confidence intervals different from zero to separate increasing, neutral, or decreasing trends over the years. The standardized major axis (sma) regression was fitted⁴⁶ to test the relationship between AAs concentrations relative to protein and protein concentration. Data visualization (Figs. 1, 2, 3, 4) was performed using the package “ggplot2”⁴⁷ within the R software⁴³.

Received: 8 May 2020; Accepted: 29 September 2020

Published online: 19 October 2020

References

1. USDA. *Oilseeds: World Markets and Trade. Global Market Analysis* <https://apps.fas.usda.gov/psdonline/circulars/oilseeds.pdf> (2020).
2. Bernard, J. K. Oilseed and oilseed meal. In *Encyclopedia of Dairy Sciences* (eds Fuguay, J. W. et al.) 349–355 (Academic Press, New York, 2011).
3. Cromwell, D. G. Soybean Meal—An Exceptional Protein Source (2008).
4. Parsons, C. M. Digestible amino acids for poultry and swine. *Anim. Feed Sci. Technol.* **59**, 147–153 (1996).
5. de Felipe, M., Gerde, J. A. & Rotundo, J. L. Soybean genetic gain in maturity groups III to V in Argentina from 1980 to 2015. *Crop Sci.* **56**, 3066 (2016).
6. Patil, G. et al. Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. *Theor. Appl. Genet.* **130**, 1975–1991 (2017).
7. Chung, J. et al. The seed protein, oil, and yield QTL on soybean linkage group I. *Crop Sci.* **43**, 1053–1067 (2003).
8. Assefa, Y. et al. Assessing variation in us soybean seed composition (protein and oil). *Front. Plant Sci.* **10**, 298 (2019).
9. Pfarr, M. D., Kazula, M. J., Miller-Garvin, J. E. & Naeve, S. L. Amino acid balance is affected by protein concentration in soybean. *Crop Sci.* **58**, 2050 (2018).
10. Kumar, V. et al. Differential distribution of amino acids in plants. *Amino Acids* **49**, 821–869 (2017).
11. Utsumi, S., Matsumura, Y. & Mori, T. Structure–function relationships of soy proteins. In *Food Proteins and Their Applications* (eds Damodaran, S. & Paraf, A.) 257–291 (Marcel Dekker Inc., New York, 1997).
12. Thakur, M. & Hurburgh, C. R. Quality of us soybean meal compared to the quality of soybean meal from other origins. *J. Am. Oil Chem. Soc.* **84**, 835–843 (2007).
13. Medic, J., Atkinson, C. & Hurburgh, C. R. Current knowledge in soybean composition. *J. Am. Oil Chem. Soc.* **91**, 363–384 (2014).
14. Wu, T. et al. Temporal–spatial characterization of seed proteins and oil in widely grown soybean cultivars across a century of breeding in China. *Crop Sci.* **57**, 748–759 (2017).
15. Wilson, R. F. Seed composition. In *Soybeans: Improvement, Production and Users. Agronomy Monograph 16621–677* (ASA, CSSA, SSSA, 2004).
16. Rotundo, J. L. & Westgate, M. E. Meta-analysis of environmental effects on soybean seed composition. *F. Crop. Res.* **110**, 147–156 (2009).
17. Tamagno, S. et al. Nutrient partitioning and stoichiometry in soybean: a synthesis-analysis. *F. Crop. Res.* **200**, 18–27 (2017).
18. Ortez, O. A., Tamagno, S., Salvagiotti, F., Prasad, P. V. V. & Ciampitti, I. A. Soybean nitrogen sources and demand during the seed-filling period. *Agron. J.* **111**, 1779–1787 (2019).

19. Warembourg, F. R. & Fernandez, M. P. Distribution and remobilization of symbiotically fixed nitrogen in soybean (*Glycine max*). *Physiol. Plant.* **65**, 281–286 (1985).
20. Wilson, E. W. *et al.* Genetic gain \times management interactions in soybean: ii. Nitrogen utilization. *Crop Sci.* **54**, 340–348 (2014).
21. Mahmoud, A. A. *et al.* Effect of six decades of selective breeding on soybean protein composition and quality: a biochemical and molecular analysis. *J. Agric. Food Chem.* **54**, 3916–3922 (2006).
22. Ortez, O. A. *et al.* Exploring nitrogen limitation for historical and modern soybean genotypes. *Agron. J.* **110**, 2080–2090 (2018).
23. Morrison, M. J., Voldeng, H. D. & Cober, E. R. Agronomic changes from 58 years of genetic improvement of short-season soybean cultivars in Canada. *Agron. J.* **92**, 780–784 (2000).
24. Rincker, K. *et al.* Genetic improvement of U.S. Soybean in maturity groups II, III, and IV. *Crop Sci.* **54**, 1419–1432 (2014).
25. Rowntree, S. C. *et al.* Physiological and phenological responses of historical soybean cultivar releases to earlier planting. *Crop Sci.* **54**, 804–816 (2014).
26. Kiang, Y. T., Gorman, M. B. & Chiang, Y. C. Genetic and linkage analysis of a leucine aminopeptidase in wild and cultivated soybean. *Crop Sci.* **25**, 319 (1985).
27. Herman, E. M. Soybean seed proteome rebalancing. *Front. Plant Sci.* **5**, 1–8 (2014).
28. Ham, G. E., Liener, I. E., Evans, S. D., Frazier, R. D. & Nelson, W. W. Yield and composition of soybean seed as affected by n and s fertilization. *Agron. J.* **67**, 293–297 (1975).
29. Hungria, M. *et al.* Nitrogen nutrition of soybean in Brazil: Contributions of biological N₂ fixation and N fertilizer to grain yield. *Can. J. Plant Sci.* **86**, 927–939 (2006).
30. Kaschuk, G., Nogueira, M. A., de Luca, M. J. & Hungria, M. Response of determinate and indeterminate soybean cultivars to basal and topdressing N fertilization compared to sole inoculation with *Bradyrhizobium*. *Field Crop Res.* **195**, 21–27 (2016).
31. Salvagiotti, F. *et al.* Nitrogen uptake, fixation and response to fertilizer N in soybeans: a review. *Field Crop Res.* **108**, 1–13 (2008).
32. Nakasathien, S., Israel, D. W., Wilson, R. F. & Kwanyuen, P. Regulation of seed protein concentration in soybean by supra-optimal nitrogen supply. *Crop Sci.* **40**, 1277–1284 (2000).
33. Allen, D. K. & Young, J. D. Carbon and nitrogen provisions alter the metabolic flux in developing soybean embryos. *Plant Physiol.* **161**, 1458–1475 (2013).
34. Thanapornpoonpong, S., Vearasilp, S., Pawelzik, E. & Gorinstein, S. Influence of various nitrogen applications on protein and amino acid profiles of amaranth and quinoa. *J. Agric. Food Chem.* **56**, 11464–11470 (2008).
35. Zhang, P. *et al.* Effect of irrigation and nitrogen application on grain amino acid composition and protein quality in winter wheat. *PLoS ONE* **12**, e0178494 (2017).
36. Eppendorfer, W. H., Bille, S. W. & Patipanawattana, S. Protein quality and amino acid-protein relationships of maize, sorghum and rice grain as influenced by nitrogen, phosphorus, potassium and soil moisture stress. *J. Sci. Food Agric.* **36**, 453–462 (1985).
37. Krishnan, H. B., Bennett, J. O., Kim, W. S., Krishnan, A. H. & Mawhinney, T. P. Nitrogen lowers the sulfur amino acid content of soybean (*Glycine max* [L.] Merr.) by regulating the accumulation of Bowman-Birk protease inhibitor. *J. Agric. Food Chem.* **53**, 6347–6354 (2005).
38. Ackerman, E. A. The Koppen classification of climates in North America. *Geogr. Rev.* **31**, 105 (1941).
39. Fehr, W. R. & Caviness, C. E. Stages of soybean development. vol. *Special Re* 1–12 (1977).
40. AOAC. *Official methods of analysis* (2016).
41. Fischer, R. A., Byerlee, D. & Edmeades, G. *Crop yields and global food security*. *ACIAR Monograph No. 158* (2014).
42. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Software; Vol 1, Issue 1* (2015).
43. R Core Team. *R: A Language and Environment for Statistical Computing*. *R Foundation for Statistical Computing, Vienna, Austria*, <https://www.R-project.org/> (2019).
44. Härdle, W., Horowitz, J. & Kreiss, J.-P. Bootstrap methods for time series. *Int. Stat. Rev.* **71**, 435–459 (2003).
45. Altman, D. G. & Bland, J. M. Statistics notes: how to obtain the P value from a confidence interval. *BMJ* **343**, 1–2 (2011).
46. Warton, D. I., Duursma, R. A., Falster, D. S. & Taskinen, S. smatr 3- an R package for estimation and inference about allometric lines. *Methods Ecol. Evol.* **3**, 257–259 (2012).
47. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, New York, 2016).

Acknowledgements

The authors would like to thank students and interns from the KSU Crops Team. A special gratitude to Rachel Veenstra and Jill Miller Garvin for the English review. This research was funded by United States Soybeans Board, project no. 2020-152-0104. This is contribution no. 21-081-J from Kansas Agricultural Extension Station.

Author contributions

I.A.C., conceptualized the objectives, research questions, and developed the study design. I.A.C., O.A.O., S.T., S.N. provided leadership and/or execution of data collection. L.M.R., A.F.B.R., S.T., I.A.C., developed the code for data analysis and visualization. A.F.B.R., S.T., L.M.R., drafted the manuscript. I.A.C., S.T., S.N., O.A.O., A.F.B.R., made critical revision to the manuscript and contributed to the final version.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.F.B.R. or I.A.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020