ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization (2018) **16**(2); 94–102
doi:10.1017/S1479262116000472

Evaluation of seed components of wild soybean (*Glycine soja*) collected in Japan using near-infrared reflectance spectroscopy

Chi-Do Wee[1], Masatsugu Hashiguchi[2], Genki Ishigaki[2], Melody Muguerza[2], Chika Oba[2], Jun Abe[3], Kyuya Harada[4] and Ryo Akashi[1,2]*

[1]*Interdisciplinary Graduate School of Agriculture and Engineering, University of Miyazaki, Miyazaki 889-2192, Japan,* [2]*Faculty of Agriculture, University of Miyazaki, Miyazaki 889-2192, Japan,* [3]*Graduate School of Agriculture, Hokkaido University, Sapporo 060-8589, Japan and* [4]*Graduate School of Engineering, Osaka University, Suita, Osaka 565-0871, Japan*

Received 25 August 2016; Accepted 16 December 2016 – First published online 23 January 2017

Abstract

Seed composition, including the protein, lipid and sucrose contents of 334 accessions of wild soybean (*Glycine soja*) collected in Japan, was evaluated using near-infrared reflectance spectroscopy (NIRS) technology. The distribution of protein, lipid and sucrose contents and correlations among these three classes of seed components were determined. Protein, lipid and sucrose levels ranged in accessions from 48.6 to 57.0, 9.0 to 14.3 and 1.24 to 3.53%, respectively. Average levels of protein, lipid and sucrose in the accessions were 54, 11 and 2.5%, respectively. High negative correlations were observed between the protein and lipid contents, and the protein and sucrose contents. Mean levels of the three constituents were compared among collection sites classified by climatic conditions. The total protein content of accessions from regions with a high annual mean temperature was high. The protein content of accessions from the II-1 region was higher than those from the III-3 region, and the sucrose content from the II-1 region was lower than those from regions III-2 and IV-3. The lipid content of plants from the II-1 region was lower than those from other regions, and the accessions in region II had a higher protein content and lower sucrose and lipid contents than the other regions. These results provide diverse and wide-ranged protein, lipid and sucrose contents information of Japanese wild soybean resources according to climatic region; thus, providing a foundation for the future development and selection of new soybean varieties with desired traits in global environmental changes.

Keywords: wild soybean, protein, lipid, sucrose, NIRS, climatic regions

Introduction

Soybean (*Glycine max* (L.) Merr.) is the most important grain legume crop in the world. Cultivated soybean seeds contain ~40% protein and ~20% lipid, and its crop economic value is dependent on the concentration of protein and oil (Brumm and Hurburgh, 1990). The relationship between the protein, oil and sugar components of mature soybean seed has been the subject of numerous investigations (Hurburgh *et al.*, 1990; Hurburgh, 1994; Piper and Boote, 1999; Yaklich *et al.*, 2002; Brumm and Hurburgh, 2006; Dardanelli *et al.*, 2006). Endemic over a wide range of East Asia, wild soybean (*Glycine soja* Sieb. and Zucc.) is the closest relative of cultivated soybeans (*G. max*) but both have prominent morphological and physiological differences, known as the domestication syndrome (Broich and Palmer, 1980; Hyten *et al.*, 2006). However, the significant loss of diversity among high-yielding adapted lines ultimately inhibits future genetic gains in productivity,

*Corresponding author. E-mail: rakashi@cc.miyazaki-u.ac.jp

broadens susceptibility to new pests and diseases, and acts as a threat to food security (Carter *et al*., 2004; Hyten *et al*., 2006). The narrowed genetic base in cultivated soybean limits improvements in yield and quality (Hajjar and Hodgkin, 2007; Chung and Singh, 2008). In contrast to modern soybean cultivars, wild soybeans are genetically diverse and have many valuable alleles (Lee *et al*., 2008).

Recent advances in sequencing technologies has also highlighted the unique genomic structures of both cultivated and wild soybeans and has provided an opportunity to use *G. soja* to broaden the genetic base of cultivated soybean (Kim *et al*., 2012; Joshi *et al*., 2013). Furthermore, assessing genomic differences for important traits will provide insights into the process of speciation and domestication and will deepen our understanding of the origin of genes involved in complex traits (Kim *et al*., 2010; Joshi *et al*., 2013; Li *et al*., 2013; Zhao *et al*., 2015; Zhou *et al*., 2015; Han *et al*., 2016).

Isolation and characterization of seed storage proteins from wild soybeans are essential to physiological studies of protein accumulation during development and their utilization during germination. In addition, sucrose contributes to good taste and desirable texture of soyfoods, and the sucrose content determines the sweetness of raw vegetable soybean seed (Kumar *et al*., 2011). Furthermore, sucrose content is higher when soybeans are grown at cooler locations, whereas the stachyose content is genotype-dependent across different growing environments (Kumar *et al*., 2010). Another factor that has made breeding for the sucrose content in soybean seed difficult is the cost for quantifying this disaccharide (Maughan *et al*., 2000). Conventional analytical chemistry techniques for seed constituents are expensive, time consuming and destructive. Fourier transform near-infrared reflectance spectroscopy (NIRS) technology was introduced over the last few decades for the wide-scale, inexpensive chemical analysis of food and crop seed composition (Wilcox and Cavins, 1995; Buning and Diller, 2000).

Here, we present the distribution and characteristics of three major classes of seed components, protein, lipid and sucrose, based on NIRS measurements of 334 wild soybean accessions from regions representing a wide range of climatic conditions in Japan.

Materials and methods

Plant materials and collection sites

A total of 334 Japanese wild soybean (*G. soja*) accessions were used in this study. All accessions were collected across several climatic zones from Niikappu, Hokkaido (42°25′06″N) to the southernmost site, Minamikyushu, Kagoshima (31°16′11″N). The collection sites of the wild accessions covered a broad geographic range in Japan and are listed in Table S1.

All accessions were classified into the macro (Roman numeral) and meso (Arabic number) subdivisions of climatic regions as described by Yoshino (1980). For simplicity, the collection sites were classified into geographic areas according to the climatic classification in Japan (Figure 1). These accessions were obtained from 'Legumebase', an online database (http://www.legumebase.brc.miyazaki-u.ac.jp/) supported by the National BioResource Project in Japan. For the comparison, a *G. max* cultivar, 'Fukuyutaka', was evaluated for seed protein, lipid and sucrose contents using identical methods as those used for the wild accessions.

Quantitative analysis of protein content

Seed protein in the wild accessions was measured quantitatively using the Biuret method. Seed coats were removed from 20 to 30 seeds per accession, the naked seeds were ground using a TissueLyser II (QIAGEN, Hilden, Germany), and finally the particles were filtered through a 0.5 mm sieve. Soybean powder was dried in a 100°C dryer for 1 h and left to cool in a desiccator at room temperature for 1 h. Soybean powder (50 mg) was placed in a flask to which 0.5 ml of chloroform (CHC_{13}), Biuret reagent (10 mM KOH, 5 mM $KNaC_4H_4O_6$ $4H_2O$) and 20 ml of 1.6 mM Cu_2SO_4 $5H_2O$ were added. The flask was shaken for 15 min and held for 1 h without shaking. The flask contents were transferred to a 50 ml tube and centrifuged at 3300 × *g* for 15 min. The absorbance of the supernatant liquid was measured at 550 nm using a Nano Drop 1000 spectrophotometer (Thermo Fisher Scientific, Inc., MA, USA).

Quantitative analysis of lipid content

Seed lipids were quantitatively measured from the wild accessions using a CHC_{13} extraction method (Southgate, 1971) with a few modifications. Soybean seeds (100 mg) were ground into a fine powder and transferred into a 2-ml Eppendorf tube containing an iron bead. After adding 1.5 ml of hexane, the mixture was extracted by shaking in a multi-beads shocker (Yasui Kikai Corporation, Osaka, Japan) at 800 × *g* for 60 s, followed by centrifugation at 5100 × *g* for 15 min. The aqueous phase was collected and transferred to a 13 × 100 mm glass test tube, and the remaining pellet was re-suspended in 1 ml of hexane. The above extraction steps were repeated three times. For purification, the lipid-containing solution was washed with 3 ml of H_2O, mixed gently by inversion, and centrifuged at 800 × *g* for 15 min. The upper phase was removed, transferred to a clean glass test tube using a Pasteur pipette, and filtered

https://doi.org/10.1017/S1479262116000472 Published online by Cambridge University Press

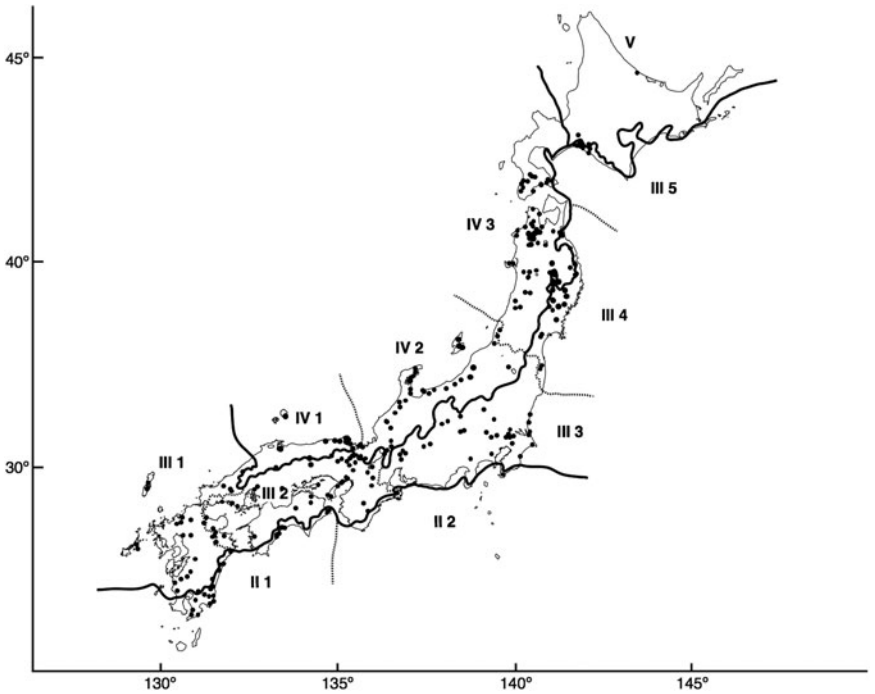

Fig. 1. The collection sites of *Glycine soja* accessions used in this study. Axes indicate latitude and longitude. Macro (Roman numerals) and meso (Arabic numbers) scale subdivisions indicate the climatic regions of Japan according to Yoshino (1980).

through a 0.2 µm pore size syringe filter. Next, the hexane solution was evaporated at 70°C in a Dry Thermo Unit TAH-2 G (TAITEC Corporation, Saitama, Japan) for 4–6 h, followed by drying in a SpeedVac-SC100 centrifugal evaporator (SAVANT INSTRUMENTS, INC, NY, USA). Finally, the lipid was dissolved in 0.5 ml of CHC_{13} and transferred to a weighed fresh glass test tube by means of Pasteur pipette. This final step was repeated three times to thoroughly rinse the inner part of the test tubes. After evaporating the CHC_{13} at 70°C in a Dry Thermo Unit for 2 h, the test tube was weighed again. The difference in weight was considered to be the lipid weight, and the lipid content was expressed as a percentage of the 100 mg ground-seed weight.

Lipid content (%) = $(W_1 - W_0) \times 100$

W_1, weight of the tube containing the extracted lipid; and W_0, weight of the empty tube.

Quantitative analysis of sucrose content

Quantitative measurements of seed sucrose content in the wild accessions were obtained using a high-performance liquid chromatograph (HPLC) (SHIMADZU Corporation, Kyoto, Japan) method after a preliminary fat extraction. Seed coat were removed and ground using a TissueLyser II (QIAGEN, Hilden, Germany), and finally the particles were filtered through a 0.5 mm sieve. Soybean powder (0.15 mg) was mixed with 1.5 ml of distilled water in a 2-ml Eppendorf tube. The mixture was shaken horizontally at 150 rpm for 15 min using a Uni Thermo shaker NTS-1300 (TOKYO RIKAKIKAI CO., LTD., Tokyo, Japan) at room temperature, and centrifuged at 15,300 × *g*) for 10 min at room temperature. The supernatant liquid (0.5 ml) was transferred to a 1.5 ml tube, 0.5 ml of 95% acetonitrile was added at room temperature, and the tube was centrifuged again. The supernatant liquid (0.2 ml) was transferred into a new 1.5 ml test tube and dried at 95°C in a dry bath incubator for 45 min. Finally, the residue was dissolved in 0.2 ml of 65% acetonitrile and filtered through a 0.2 µm syringe filter prior to HPLC analysis.

Sucrose was identified and quantified using a TSK-GEL amide column (80 × 5µm, column temperature: 50°C) (TOSOH corporation, Tokyo, Japan) in a CBM-20A HPLC system equipped with a LC10AD pump, a degasser DGU-14A, an auto sampler SIL-10AD, a system controller CBM-20A, a column thermostat CTO-14A and a refractive index detector RID-10A (SHIMADZU corporation, Kyoto, Japan). Sucrose was eluted with 75% (v/v) acetonitrile and 25% (v/v) distilled water at a flow rate of 0.5 ml/min under isocratic conditions.

Seed composition analyses for protein, lipid and sucrose using NIRS

Fourier NIRS (NIRFlex N-500, BUCHI Labortechnik AG, Flawil, Switzerland) was used to develop calibration equations for simultaneously measuring the protein, lipid and

https://doi.org/10.1017/S1479262116000472 Published online by Cambridge University Press

sucrose contents of seeds. Seed coats of 334 wild accessions were removed, approximately 50–100 naked seeds per accession were pulverized using a TissueLyser II (QIAGEN, Hilden, Germany), and finally the particles were filtered through a 0.5 mm sieve. The interval for the spectrometric analyses was 4 cm^{-1} within a wavenumber range of 4,000–10,000 cm^{-1}; the near-infrared range is 800–2,500 nm (4,000–12,500 cm^{-1}). The preconditioning process, including smoothing, differentiation and normalization of spectrometric analyses, was carried out to revise environmental factors and spectrometer confusion using NIRCal (BUCHI Labortechnik AG, Flawil, Switzerland). The linear combination of variables for absorbance changes in near-infrared wavelength range was translated to explanatory variables, and a partial least-squares (PLS) regression was conducted using the chemical analyses values as target variables.

The equation for the PLS regression algorithm was

$$y_i = \sum_{k=1}^{r} c_k \cdot t_{ik} + e_i,$$

where y_i is the target variable (chemical analysis value) of the sample i; c_k is a regression coefficient for wavelength k; t_{ik} is an explanatory variable for sample i and wavelength k; and e_i is a potential variable for the residual error of sample i. Moreover, the accuracy of calibration was estimated by the Evaluation index (EI) testing method (Mizuno *et al.*, 1988).

$$\mathrm{EI} = \frac{2 \times \mathrm{SDP}}{\mathrm{Range}} \times 100\ (\%),$$

where SDP means the standard deviation of the prediction error; and Range means the range of prediction for sample values.

Results

Set of calibration curves and accuracy tests for the NIRS results

We measured the protein, lipid and sucrose contents of seeds from wild and cultivated soybeans using NIRS, and designed a calibration curve set from the results obtained by chemical analysis. A total of 50 accessions were examined for the calibration set; the correlation coefficients (r^2) and standard error of calibration (SEC) values for the protein, lipid and sucrose contents were 0.917 and 1.999, 0.899 and 1.473, 0.758 and 0.771, respectively (Table 1). In order to estimate the accuracy of the calibration curve, an EI test was conducted using this calibration curve set. From the NIRS analyses, the range, r^2 and SDP values for protein, lipid and sucrose contents are presented in Table 1. In accordance with the accuracy testing method, the EI values for protein, lipid and sucrose were calculated to be 20.0, 21.6 and 22.9%, respectively. All EI values for seed composition as measured by NIRS analysis were within a 12.5–24.9% range, and the criterion based on EI values was classified as a rank B, which has high accuracy and good practical usability (Mizuno *et al.*, 1988). These results indicate that NIRS can be used to measure these three chemical components instead of using chemical analysis methods.

Evaluation and distribution of protein, sucrose and lipid contents in wild soybean seeds

Three hundred and thirty-four accessions of wild soybean collected in Japan were evaluated for their seed protein, lipid and sucrose contents by NIRS-measured frequencies relative to the respective calibration equations. Figure 2 shows the distributions and ranges of the three constituents in the same sample set. The distribution of protein content ranged from 48.6% (B03039) to 57.0% (B02120), and 196 accessions (~59%) of all wild soybean accessions had protein contents of 54%, the greatest frequency (Figure 2A). One, 4, 53 and 80 wild accessions had protein contents of approximately 48, 50, 52 and 56%, respectively. As the protein content of the *G. max*. 'Fukuyutaka' was 44.6%, the protein content of all wild accessions was higher than that of the control, suggesting that the protein content of *G. soja* seeds is greater than that of *G. max*.

The lipid content of the wild accessions varied from 9.0% (B02156) to 14.3% (B03039), and 158 accessions had a lipid content of approximately 11%, the highest frequency for this seed component (Figure 2B). Four, 54, 90, 23 and 5 accessions had lipid contents of approximately 9, 10, 12, 13

Table 1. Calibration set and accuracy test of NIRS results

	NIRS calibration set				Accuracy test							
	n	*r*	r^2	SEC	*n*	Range	*r*	r^2	SDP	Bias	EI (%)	Rank[a]
Protein	50	0.958	0.917	1.999	26	37.9–60.6	0.940	0.884	2.273	0.424	20.0	B
Lipid	50	0.944	0.899	1.473	34	8.1–23.7	0.912	0.832	1.684	0.051	21.6	B
Sucrose	50	0.871	0.758	0.771	28	1.38–7.28	0.915	0.884	0.675	0.005	22.9	B

[a]Rank A = under 12.4%, B = 12.5–24.9%, C = 25.0–37.4%, D = 37.5–49.9%, and E = upper 50.0% of EI value (Mizuno *et al.*, 1988).

https://doi.org/10.1017/S1479262116000472 Published online by Cambridge University Press

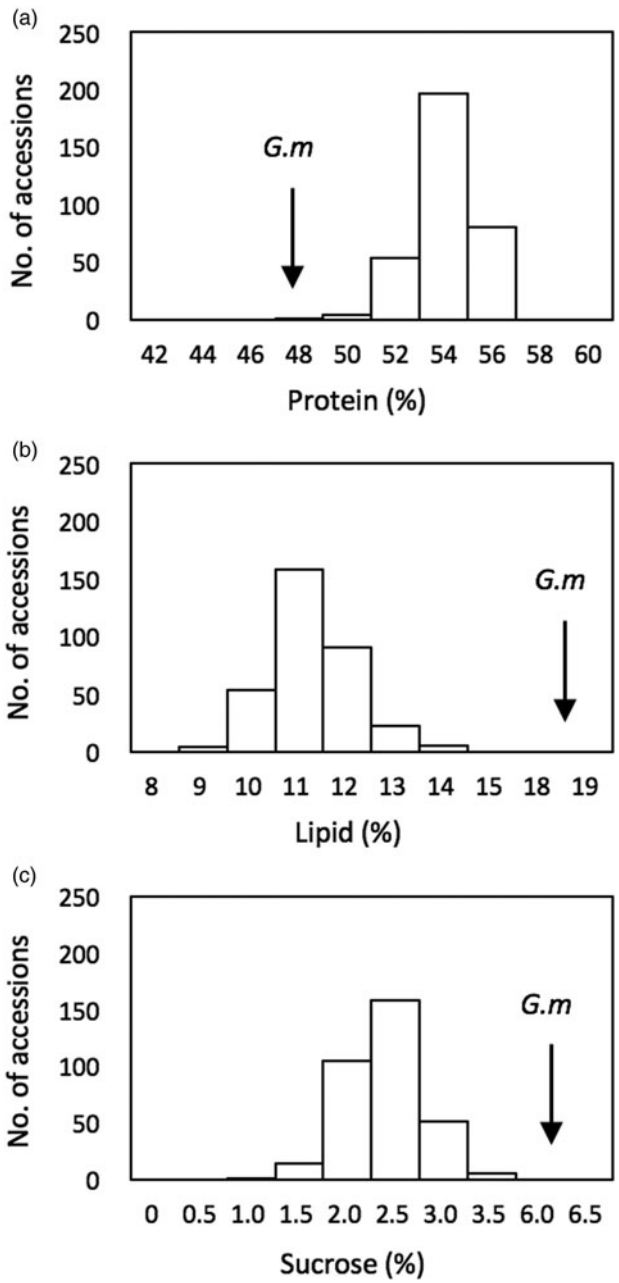

Fig. 2. Distribution of protein, lipid and sucrose contents of *Glycine soja* accessions collected in Japan (*G.m* = *Glycine max* 'Fukuyutaka').

and 14%, respectively. Furthermore, the lipid content of 'Fukuyutaka' was 20.1%, suggesting that the lipid content of *G. soja* tends to be lower compared with *G. max*.

Among the wild accessions, the seed sucrose content varied from 1.24% (B02237) to 3.53% (B02258), and 158 accessions had sucrose contents of approximately 2.5%, the highest frequency for this seed constituent (Figure 2C). One, 14, 105, 51 and 5 accessions had approximately sucrose contents of 1.0, 1.5, 2.0, 3.0 and 3.5%, respectively. The sucrose content of Fukuyutaka with 6.89%, suggests that *G. soja* tends to have a lower sucrose content than *G. max*, a result similar to the pattern for lipid content.

Correlations between the protein, sucrose and lipid contents in wild soybean seeds

Linear regressions between protein, lipid and sucrose contents were calculated using 334 Japanese wild soybean accessions, and Figure 3 shows the results and correlation

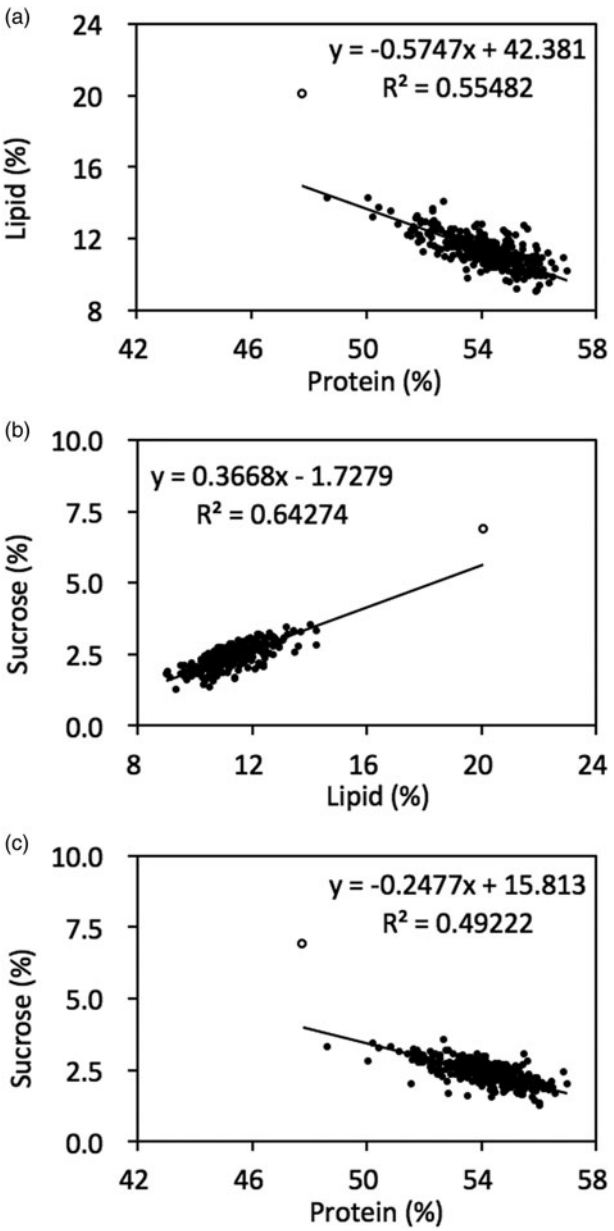

Fig. 3. Correlation between protein, lipid and sucrose contents of *Glycine soja* accessions collected in Japan. Solid circles: *G. soja*, open circles: *Glycine max* 'Fukuyutaka'.

https://doi.org/10.1017/S1479262116000472 Published online by Cambridge University Press

coefficients. High negative correlations were observed between lipid and sucrose contents and protein content, and the r^2 values were 0.5548 and 0.4922, respectively (Figure 3A and C). Moreover, the correlation between the sucrose and lipid contents was highly positive with a 0.6427 r^2 value (Figure 3B). In contrast, the ratio of the lipid and sucrose contents as a function of the protein content in *G. soja* accessions was lower than that of 'Fukuyutaka' (Figure 3A and C). Similarly, the ratio of the sucrose content as a function of the lipid content in *G. soja* accessions was also lower than that of 'Fukuyutaka' (Figure 3B).

Comparison of seed compositions of plants collected from various climatic regions in Japan

Seed protein content means for *G. soja* plants collected from regions II-1, II-2, III-1, III-2, III-3, III-4, III-5, IV-1, IV-2, IV-3 and V-1 were 54.6, 55.2, 53.9, 54.1, 53.7, 54.3, 54.2, 53.7, 54.0, 54.1 and 54.1%, respectively (Figure 4A). The total seed protein content of *G. soja* collected from regions with high annual average temperatures was high, especially the protein content for accessions from the II-1 region were significantly higher than those from region III-3. The sucrose content means in the same respective climatic regions described above were 2.19, 2.02, 2.36, 2.44, 2.42, 2.43, 2.48, 2.42, 2.47, 2.43 and 2.24% (Figure 4C). The seed sucrose content of accessions collected in region II-1 was significantly lower than those from regions III-2 and IV-3. Although it was not significant, the seed lipid content of accessions from region II-1 was lower than those from the other regions (Figure 4B). The *G. soja* accessions collected in region II had higher protein contents and lower sucrose and lipid contents than the other regions.

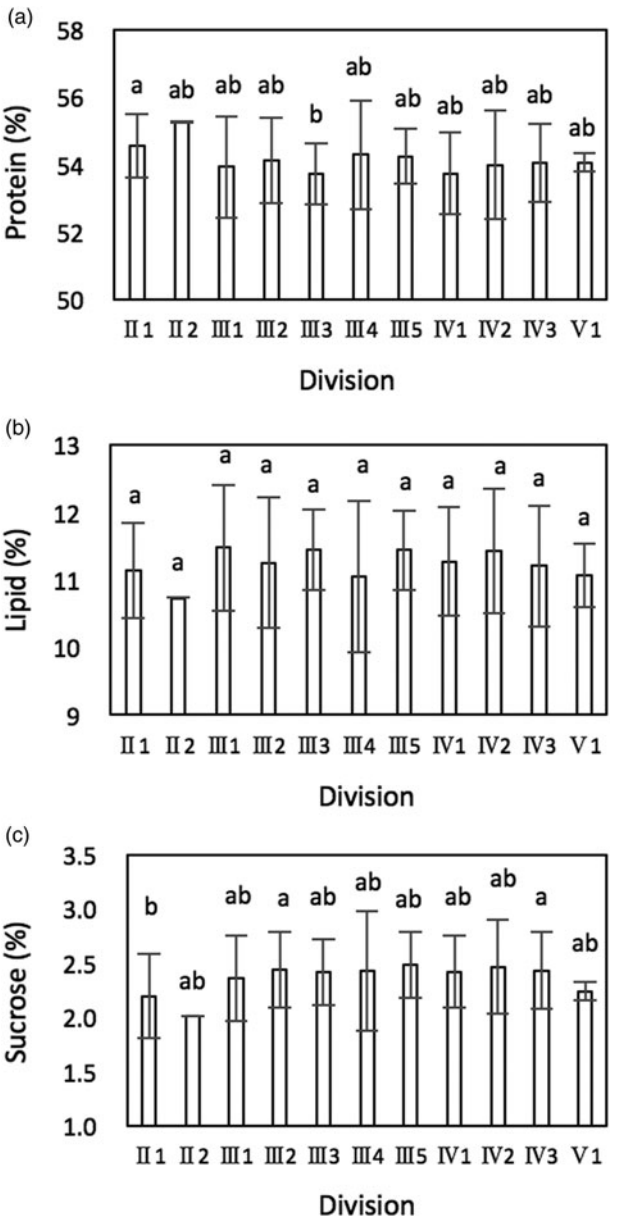

Fig. 4. Comparison of protein, lipid and sucrose contents of *Glycine soja* collected from distinct climatic divisions in Japan. The different letters on the bars show significance at $P < 0.05$ by the Steel-Dwass test.

Discussion

NIRS is time-efficient analytical technology for screening a large number of samples for different chemical constituents and their concentrations (Font and de Haro-Bailon, 2006). Therefore, the NIRS technique has been utilized to determine the protein and oil content of progeny during plant breeding and has been used to measure concentrations of fatty acids, anthocyanins, 11S to 7S protein fractions, isoflavones, lecithin, oligosaccharides and amino acids in soybean (Lee *et al.*, 2011). The availability of NIRS calibrations for screening breeding materials is especially desirable because a single NIRS spectrum can be collected to measure seed protein and oil content. Furthermore, the NIRS-based method does not require additional sample preparation as compared with chromatographic or enzymatic methods for sucrose analysis (Choung, 2010; Sato *et al.*, 2012).

In this study, we investigated the protein, lipid and sucrose contents of wild soybean (*G. soja*) accessions collected in Japan using NIRS. The collection sites for the 334 accessions were selected to represent populations present in diverse regions throughout Japan. We investigated the abundance of the three most important seed components in these accessions to determine the degree of diversity present in seeds collected from different climatic areas.

Although there remains a strong economic incentive to develop cultivars with high protein and oil contents, while maintaining a competitive yield, little progress towards these goals has been made. Effective breeding for seed

https://doi.org/10.1017/S1479262116000472 Published online by Cambridge University Press

composition requires accurate, inexpensive and reliable techniques for measuring seed constituents. Certain areas of breeding have benefited from methods to analyse a single soybean seed or even parts of seeds.

Numerous investigations of the regional variations in seed composition in soybean cultivars have been reported (Hurburgh *et al.*, 1990; Hurburgh, 1994; Piper and Boote, 1999; Yaklich *et al.*, 2002; Brumm and Hurburgh, 2006; Dardanelli *et al.*, 2006). Among these investigations, the protein and lipid contents of soybean cultivars in America were reported to be in the approximate ranges of 33–48 and 17–21%, respectively. According to a study on seed components of wild soybean in China (Dong *et al.*, 2001), 5929 accessions had an average protein content of 45.42%, with significant differences seen among accessions collected at different latitude zones. There were also extreme cases of protein content among the Chinese accessions: the highest protein content was 55.7% and the lowest was 29.0%. The average lipid content of the wild accessions in China was 10.6% with accessions varying in lipid content from 4.4 to 20.2%.

In our study, the seed protein content varied from 48.6 to 57.0% and had an average of approximately 54.0%; the lipid content varied from 9.0 to 14.3% and had an average value of 11.4%. The results of our investigation indicate higher protein and lower lipid contents in the wild accessions than those reported in other studies. This result might be due to the method used to measure the seed components; however, the seed composition characteristics of wild soybean accessions collected in Japan have not previously been investigated.

Many studies have discussed the inverse relationship between seed protein and oil contents (Brim and Burton, 1979; Carver *et al.*, 1986; Dornbos and Mullen, 1992; Bellaloui *et al.*, 2009), but little is known about the relationships among protein, oil and sugars, even though sugars can contribute to more than 10% of the total seed dry weight. A previous study indicates that the total sugar and oil concentrations in soybean seeds are positively associated, and each is negatively correlated with protein concentration (Openshaw and Hadley, 1978). Breeding efforts to modify seed composition have concentrated on changing only the oil and protein percentages while ignoring the sugar concentration. Sucrose was the predominant soluble sugar, accounting for an average of 55.16% (33–90%) of the total sugar. The average sucrose content was 5.35% among all *G. max* accessions screened, with more than threefold differences between the extremes, ranging from 2.98 to 10.55% (Hou *et al.*, 2009). Our results indicate that the sucrose content of Japanese wild soybean varied from 1.24 to 3.53%.

In this paper, we classified the collection sites based on the climatic regions with reference to the standard scale of climates in Japan (Yoshino, 1980). The boundaries are established as follows. The boundary between regions II-1 and II-2 approximately coincides with the 180°C warmth index line, or the 20°C isotherm for annual air temperature. The boundary between regions II and III coincides in most parts with the 0°C isothermal line for the mean daily minimum temperature in January, which is an indicator for frequent frost occurrence. This boundary roughly corresponds to the 16°C isotherm of the annual mean temperature on the Pacific coast of Honshu, Shikoku and Kyushu. The boundary between regions III and IV, which is an important boundary in Japan, reflects the divide between the Pacific climate and the Japan Sea climate. The boundary is drawn in accordance with the 50 cm isoline for mean maximum snow depth that plays a defining role as a delimiting factor in the distribution of vegetation. The boundary between regions IV and V is drawn by taking into account the regions where the monthly mean temperature is 0°C for 4 months or more (Yoshino, 1980).

Although no wild accessions from region I were evaluated in our study, the seed composition of accessions from region II showed interesting results, especially as related to protein and sucrose contents. The average value for protein content in accessions collected in region II was higher than the other regions, and the average value for the sucrose content was low. These results indicate that the wild soybean accessions from the region with the highest annual temperature in Japan were characterized as having high seed protein levels and low lipid and sucrose levels.

The wild soybean is the direct progenitor of the soybean cultivar (Hymowitz, 1970) is distributed in Russia, Korea, Japan and China. Because of the geography of Japan (north (46°N, 146°E) and south (24°N, 122°E)) variation in photoperiod and temperature are great; thus, ecotype divergence may also vary according to the regional adaptation (Kaga *et al.*, 2012). Information of geographic distribution and genetic diversity of wild and cultivated soybeans in Japan, including other countries is important especially in breeding programmes and conservation of soybeans. Previous studies have reported low genetic exchanges between germplasm from northern, central and southern regions in Japan (Zhou *et al.*, 2002). Evaluation of the genetic structure of wild soybean in Japan based on microsatellite markers and found that soybean varieties accounting for 95% of the cultivated area in Japan had much lesser genetic variation as compared with that of the wild soybeans (Kuroda *et al.*, 2006). This trend is supported by other previous studies for wild and cultivated soybeans using germplasm from other countries (Maughan *et al.*, 1995; Xu *et al.*, 2002) using microsatellite variation. Moreover, geographical structure of genetic variation in Japan indicated that wild soybeans in northern Japan are distinct from southern Japan, while wild soybeans in central Japan had variations from both regions (Kuroda *et al.*, 2006). On the other hand, Zhou *et al.* (2002) reported

https://doi.org/10.1017/S1479262116000472 Published online by Cambridge University Press

distinct genetic base of Japanese improved varieties when compared with that of China, USA, Canada. Comparison between Chinese soybean ecotypes from the three large ecoregions with improved Japanese varieties found that Japanese varieties had a lower genetic diversity than Chinese ecotypes but distinct alleles from them and suggested the potential to broaden the genetic base (Guan *et al.*, 2010). With the availability of high-throughput genotyping platforms, SNPs markers based could estimate more precise estimation of genetic relationships. Recently, SNP markers were used to characterize the genetic variation and population structure of 1603 soybean accessions: 832 Japanese landraces, 109 old and 57 recent Japanese varieties, 341 landrace from 16 Asian countries and 264 wild soybean accessions, and it was revealed that the gene diversity of soybean in Japan was slightly lower than that of exotic soybean germplasm; however, there was a clear genetic differentiation among Japanese cultivated soybeans, exotic cultivated soybeans and wild soybeans as indicated by population differentiation and clustering analysis (Kaga *et al.*, 2012).

In this study, our results provide diverse and wide-ranged seed chemical composition information of wild soybean resources according to the climatic region for improving of new soybean varieties in global environmental changes. And the findings from this investigation will be useful for modifying soybean seed composition in a breeding programme using wild soybean accessions. A major objective of this study was to correlate the effects of environmental conditions on wild seed composition in order to predict regions and environmental conditions that consistently produce high-quality soybeans. As a direct result of this project, it will be feasible to use the wild soybean accessions in the future development and selection of new soybean varieties with desired traits.

Supplementary Material

The supplementary material for this article can be found at https://doi.org/10.1017/S1479262116000472

Acknowledgement

This work was supported by the National Bio-Resource Project of the Japanese Ministry of Education, Culture, Sports, Science and Technology (Mext), Japan.

References

Bellaloui N, Smith JR, Ray JD and Gillen AM (2009) Effect of maturity on seed composition in the early soybean production system as measured on near-isogenic soybean lines. *Crop Science* 49: 608–620.

Brim CA and Burton JW (1979) Recurrent selection in soybeans (*Glycine max*). 2. Selection for increased percent protein in seeds. *Crop Science* 19: 494–498.

Broich SL and Palmer RG (1980) A cluster analysis of wild and domesticated soybean phenotypes. *Euphytica* 29: 23–32.

Brumm TJ and Hurburgh CR (1990) Estimating the processed value of soybeans. *Journal of American Oil Chemists' Society* 67: 302–307.

Brumm TJ and Hurburgh CR (2006) Changes in long-term soybean compositional patterns. *Journal of American Oil Chemists' Society* 83: 981–AR981.

Buning PH and Diller M (2000) Rapid analysis of foods using near-infrared spectrometry (NIRS): development, use and new possibilities. *Ernahrungs Umschau*. 47: 15–20.

Carter TE Jr, Nelson RL, Sneller C and Cui Z (2004) Genetic diversity in soybean. In: Boerma HR, Specht JE eds. *Soybean: Improvement, Production, and Uses*, 3rd edn. Madison: American Society for Agronomy, pp. 303–416.

Carver BF, Burton JW, Carter TE and Wilson RF (1986) Response to environmental variation of soybean lines selected for altered unsaturated fatty acid composition. *Crop Science* 26: 1176–1180.

Choung MG (2010) Determination of sucrose content in soybean using near-infrared reflectance spectroscopy. *Journal of the Korean Society for Applied Biological Chemistry* 54: 478–484.

Chung G and Singh RJ (2008) Broadening the genetic base of soybean; a multiciplinary approach. *Plant Science* 27: 295–341.

Dardanelli JL, Balzarini M, Martinez MJ, Cuniberti M, Resnik S, Ramunda SE, Herrero R and Baigorri H (2006) Soybean maturity groups, environments, and their interaction define mega-environments for seed composition in Argentina. *Crop Science* 46: 1939–1947.

Dong YS, Zhuang BC, Zhao LM, Sun H and He MY (2001) The genetic diversity of annual wild soybeans grown in China. *Theoretical and Applied Genetics* 103: 98–103.

Dornbos DL and Mullen RE (1992) Soybean seed protein and oil contents and fatty-acid composition adjustments by drought and temperature. *Journal of the American Oil Chemists' Society* 69: 228–231.

Font RM and de Haro-Bailon A (2006) The use of near-infrared spectroscopy (NIRS) in the study of seed quality components in plant breeding programs. *Industrial Crops and Products* 24: 307–313.

Guan R, Chang R, Li Y, Wang L, Liu Z and Qiu L (2010) Genetic diversity comparison between Chinese and Japanese soybeans (*Glycine max* (L.) Merr.) revealed by nuclear SSRs. *Genetic Resources and Crop Evolution* 57: 229–242.

Hajjar R and Hodgkin T (2007) The use of wild relatives in crop improvement; a survey of developments over the last 20 years. *Euphytica* 156: 1–13.

Han Y, Zhao X, Liu D, Li Y, Lightfoot D, Yang Z, Zhao L, Zhou G, Wang Z, Huang L, Zhang Z, Qiu L, Zheng H and Li W (2016) Domestication footprints anchor genomic regions of agronomic importance in soybeans. *New Phytologist* 209: 871–884.

Hou A, Chen P, Shi A, Zhang B and Wang YJ (2009) Sugar variation in soybean seed assessed with a Rapid Extraction and Quantification Method. *International Journal of Agronomy* 2009: doi:10.1155/2009/484571, 8 p.

Hurburgh CR (1994) Long-term soybean composition patterns and their effect on processing. *Journal of the American Oil Chemists' Society* 71: 1425–1427.

Hurburgh CR, Brumm TJ, Guinn JM and Hartwig RA (1990) Protein and oil patterns in united-states and world soybean markets. *Journal of the American Oil Chemists' Society* 67: 966–973.

Hymowitz T (1970) On the domestication of the soybean. *Economic Botany* 24: 408–421.

Hyten DL, Song QJ, Zhu Y, Choi IY, Nelson RL and Costa JM (2006) Impacts of genetic bottlenecks on soybean genome diversity. *Proceeding of the National Academy of Science* 103: 16666–16671.

Joshi T, Valliyodan B, Wu J, Lee S, Xu D and Nguyen HT (2013) Genomic differences between cultivated soybean, G. max and its wild relative *G. soja*. *BMC Genomics* 14 (Suppl. 1): S5.

Kaga A, Shimizu T, Watanabe S, Tsubokura Y, Katayose Y, Harada K, Vaughan DA and Tomooka N (2012) Evaluation of soybean germplasm conserved in NIAS genebank and development of mini core collections. *Breeding Science* 61: 566–592.

Kim MY, Lee S, Van K, Kim TH, Jeong SC and Choi I (2010) Whole-genome sequencing and intensive analysis of the undomesticated soybean genome. *Proceeding of the National Academy of Sciences of the United States of America* 107: 22032–22037.

Kim KS, Diers BW, Hyten DL, Roufmian MA, Shannon JG and Nelson RL (2012) Identification of positive yield QTL alleles from exotic soybean germplasm in two backcross populations. *Theoretical and Applied Genetics* 125: 1353–1369.

Kumar V, Rani A, Goyal L, Dixit AK, Manjaya JG, Dev J and Swamy M (2010) Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by genotype and growing location. *Journal of Agricultural and Food Chemistry*. 58: 5081–5085.

Kumar V, Rani A, Goyal L, Pratap D, Billore SD and Chauhan GS (2011) Evaluation of vegetable-type soybean for sucrose, taste-related amino acids, and isoflavone contents. *International Journal of Food Properties* 14: 1142–1151.

Kuroda Y, Kaga A, Tomooka N and Vaughan D (2006) Population genetic structure of Japanese wild soybean (*Glycine soja*) based on microsatellite variation. *Molecular Ecology* 15: 959–974.

Lee J, Yu J, Hwang Y, Blake S, So Y and Lee G (2008) Genetic diversity of wild soybean accessions from South Korea and other countries. *Crop Science* 48: 606–616.

Lee JD, Shannon JG and Choung MG (2011) Application of non-destructive measurement to improve soybean quality by near infrared reflectance spectroscopy. *Soybean Applications and Technology* 16: 287–304.

Li Y, Zhao S, Ma J, Li D, Yan L, Li J, Qi X, Guo X, Zhang L, He W, Chang R, Liang Q, Guo Y, Ye C, Wang X, Tao Y, Guan R, Wang J, Liu Y, Jin L, Zhang X, Liu Z, Zhang L, Chen J, Wang K, Nielson R, Li R, Chen P, Li W, Reif J, Purugganan M, Wang J, Zhang M, Wang J and Qiu L (2013) Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. *BMC Genomics* 14: 579.

Maughan PJ, Saghai MA and Buss GR (1995) Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean. *Genome* 38: 715–723.

Maughan PJ, Maroof MA and Buss GR (2000) Identification of quantitative trait loci controlling sucrose content in soybean (*Glycine max*). *Molecular Breeding* 6: 105–111.

Mizuno K, Ishiguri T, Kondo T and Kato T (1988) Prediction of forage compositions and sheep responses by Near Infrared Reflectance spectroscopy. I. Evaluation of accuracy. *Bulletin of the National Grassland Research Institute* 38: 35–47. (Japanese with English Summary).

Openshaw SJ and Hadley HH (1978) Maternal effects on sugar content in soybean seeds. *Crop Science* 18: 581–584.

Piper EL and Boote KJ (1999) Temperature and cultivar effects of soybean seed oil and protein concentration. *Journal of the American Oil Chemists' Society* 76: 1233–1241.

Sato T, Zahlner V, Berghofer E, Losak T and Vollmann J (2012) Near-infrared reflectance calibration for determining sucrose content in soybean breeding using artificial reference samples. *Plant Breeding* 131: 531–534.

Southgate DAT (1971) A procedure for the measurement of fats in foods. *Journal of the Science of Food and Agriculture* 22: 590–591.

Wilcox JR and Cavins JF (1995) Backcrossing high seed protein to a soybean cultivar. *Crop Science* 35: 1036–1041.

Xu DH, Abe J, Gai JY and Shimamoto Y (2002) Diversity of chloroplast DNA SSRs in wild and cultivated soybeans: evidence for multiple origins of cultivated soybean. *Theoretical and Applied Genetics* 105: 645–653.

Yaklich RW, Vinyard B, Camp M and Douglass S (2002) Analysis of protein and oil from soybean northern and southern region uniform tests. *Crop Science* 42: 1504–1515.

Yoshino M (1980) The climatic regions of Japan. *Erdkunde* 34: 81–87.

Zhao S, Zheng F, He W, Wu H, Pan S and Lam H (2015) Impact of nucleotide fixation during soybean domestication and improvement. *BMC Plant Biology* 15: 81.

Zhou XL, Carter TE, Cui ZL, Miyazaki S and Burton JW (2002) Genetic diversity patterns in Japanese soybean cultivars based on coefficient of parentage. *Crop Science* 42: 1331–1342.

Zhou L, Wang S, Jian J, Geng Q, Wen J, Song Q, Wu Z, Li G, Liu Y, Dunwell J, Zhang J, Feng J, Niu Y, Zhang L, Ren W and Zhang Y (2015) Identification of domestication-related loci associated with flowering time and seed size in soybean with the RAD-seq genotyping method. *Scientific Reports* 5: 9350.