Check for updates

Genetic diversity in early maturity Chinese and European elite soybeans: A comparative analysis

Xindong Yao · Jiang-yuan Xu · Zhang-xiong Liu · Martin Pachner ·
Eva Maria Molin[ID] · Leopold Rittler · Volker Hahn[ID] · Willmar Leiser[ID] ·
Yong-zhe Gu · Yu-qing Lu · Li-juan Qiu[ID] · Johann Vollmann[ID]

Abstract Soybean (*Glycine max* [L.] Merr.) is a legume crop important for global food security due to supplying high-quality vegetable protein and oil. China is the origin and rich center of soybean genetic diversity represented by numerous landraces and other genetic resources. During the nineteenth and twentieth centuries, world-wide dissemination of plant introductions laid the foundation of modern soybean production. Due to selection for regional adaptation, present-day elite soybean cultivars do have a comparatively narrow genetic base. However, as genetic variation is the major component of future breeding progress, diversity of early maturity Chinese and European elite soybeans was comparatively analyzed using a high-throughput functional SNP array and an SSR marker set. Results revealed a clear differentiation between Chinese and European elite cultivars, and the level of genetic diversity was similar between the two populations. Both in Chinese and European cultivars, unique SSR alleles were found which indicates that selection for specific adaptation can preserve genetic variation. The European population was markedly structured by maturity groups which was less clear in the Chinese population. Structure analysis indicated that European cultivars are based on two major ancestral lines, whereas Chinese elite soybean cultivars trace back to more ancestral lines pointing to the rich natural soybean diversity of China. The results confirm that long-term selection under divergent environmental and agronomic

Xindong Yao and Jiang-yuan Xu have contributed equally to work.

X. Yao · M. Pachner · J. Vollmann (✉)
Department of Crop Sciences, University of Natural Resources and Life Sciences, Vienna, Konrad-Lorenz-Strasse 24, 3430 Tulln an der Donau, Austria
e-mail: johann.vollmann@boku.ac.at

J. Xu · Z. Liu · Y. Gu · Y. Lu (✉) · L. Qiu (✉)
National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, No. 12 Zhongguancun South Street, Haidian District, Beijing 100081, People's Republic of China
e-mail: luyuqing@caas.cn

L. Qiu
e-mail: qiulijuan@caas.cn

E. M. Molin
Center for Health and Bioresources, AIT Austrian Institute of Technology GmbH, Konrad-Lorenz-Strasse 24, 3430 Tulln an der Donau, Austria

L. Rittler
Donau Soja, Wiesingerstraße 6/14, 1010 Vienna, Austria

V. Hahn · W. Leiser
State Plant Breeding Institute, University of Hohenheim, 70593 Stuttgart, Germany

conditions can produce specific diversity which could potentially be utilized for future enhancing both Chinese and European soybean breeding.

Keywords Soybean · Genetic diversity · Plant introduction · Elite cultivar · Plant breeding · Genetic structure

Introduction

Soybean (*Glycine max* [L.] Merr.) is playing a key role in food security through supplying high-quality vegetable protein and oil. Over decades, world soybean production has steadily increased, and during the previous five years soybean was grown on 122–127 mio ha of arable land with a total annual harvest of 335–360 mio metric tons (FAOSTAT 2022). At present, North and South America are the main soybean producers and exporters, whereas China and Europe are the main soybean importing regions (Hartman et al. 2011). While China is the homeland and diversity center of soybean, world-wide dissemination, specific adaptation over many decades and continuous plant breeding activities have contributed to the evolution of regional populations in different geographic regions of soybean growing (Liu et al. 2020a).

Genetic diversity is the major driver for positive response to selection in breeding programs. During soybean domestication, early stages of landrace development, world-wide dispersal of plant introductions and modern cultivar development, however, genetic diversity has continuously decreased as a result of selection and the emergence of important domestication-related traits (Hyten et al. 2006; Sedivy et al. 2017). While the reduction of diversity has frequently been expressed as a loss of rare alleles, a significant gene loss (*i.e.* presence/absence variation and reduction of genome size) in elite soybean cultivars as compared to old cultivars, landraces or wild accessions has recently been reported from analysis of a soybean pan-genome assembled from over 1100 accessions (Bayer et al. 2022). This suggests that modern plant breeding has even selected for the absence of dispensable genes. As a consequence, monitoring and comparing of genetic diversity appears to be increasingly important to ensure future soybean breeding progress in yield and adaptation to specific environmental conditions.

On the basis of soybean genetic resources and germplasm accessions, a high level of diversity has been found in Chinese soybeans which were structured into seven primary ecotypes representing the different geographic soybean growing regions of China (Wang et al. 2006). In contrast, North American soybeans have a narrow genetic base tracing back to only a small number of ancestral plant introductions and their progeny (Gizlice et al. 1994), whereas a strong structuring of diversity according to growing regions has been found here as well (Gizlice et al. 1996). In Europe, early maturity soybean accessions from the north of Europe are related to chilling tolerant plant introductions similar to Canadian genotypes (Yamaguchi et al. 2018), whereas south European soybeans frequently have a northern United States (US) genetic background (Hrustic and Miladinovic 2011; Tomicic et al. 2015). Genetic diversity in French (Tavaud-Pirra et al. 2009), Polish (Czembor et al. 2021), Central European (Hahn and Würschum 2014) and Central-South European (Žulj Mihaljević et al. 2020) soybean germplasm and breeding materials has been reported in a number of individual studies utilizing different genomic tools and marker systems, but a comprehensive European analysis is still missing.

In contrast to the diversity present in various genetic resource collections, genetic diversity on the level of elite cultivars is of special interest in plant breeding, as it represents the latest breeding developments and the actual agrobiodiversity on farmers' fields. In a comparison between present-day Chinese and US elite soybean cultivars, genetic diversity was larger in Chinese than in the US cultivars investigated; the genetic basis between the two populations was clearly different due to selection for traits related to environmental adaptation and yield components, and signals of selection were described which might be associated with important physiological properties (Liu et al. 2017). Likewise, a comparative analysis of genetic diversity between Chinese and European soybean cultivars would be of great interest for studying the effects of separated long-term selection under divergent environmental and agronomic conditions. And in practical soybean breeding, the potential of reciprocally enriching diversity in Chinese and European breeding programs could be utilized for better meeting future challenges of soybean production. Thus, for the present study two representative

sets of early maturity Chinese and European elite soybeans were created, and a total number of 156 cultivars were genotyped with a 200 K SNP array and in addition with 71 simple sequence repeat (SSR) markers. The research objectives were (1) to compare the genetic diversity patterns present within and between Chinese and European elite soybean cultivars, (2) to study diversity on the level of maturity groups, and (3) to investigate the genetic structure present within cultivars from the two regions.

Materials and methods

Plant materials

A total set of 156 early maturity elite soybean cultivars of either Chinese (CN) or European (EU) origin was assembled for the present investigation (Suppl. Table 1 for descriptive information, Suppl. Table 2 for overview). European soybeans (77 genotypes from 10 countries) were representative of maturity groups 0000–II, and Chinese soybeans (79 genotypes) were classified into maturity groups 000–III (i.e. Northeastern Spring (NEsp) or Northern Spring (Nsp) Chinese soybean ecotype classification). For the purpose of handling and comparing similar numbers of genotypes in sub-sets, all soybean cultivars were assigned to one of four experimental groups depending on their actual maturity group classification (Suppl. Table 2).

DNA extraction

In genomic DNA extraction for SSR analysis, six seeds per genotype were germinated on filter paper. After three days, root tips (1 cm length) of 3–6 seedlings per genotype were pooled in 1.5 ml tubes and dried at room temperature in plastic bags with silica gel for two days. For DNA extraction, dried root tips were finely ground with 2 mm glass beads in a ball mill (Retsch GmbH, Haan, Germany). DNA was extracted using Wizard® DNA extraction kit from Promega (Promega Corp., Madison, WI, USA) according to protocol and stored at -20 °C until further processing. To extract genomic DNA for SNP analysis, 10 fresh leaves per genotype were frozen in liquid N_2 and ground using a mortar. Subsequently, DNA was extracted according to the protocol described by Kisha et al. (1997).

SSR genotyping

A total of 67 unlinked microsatellite loci across all 20 soybean linkage groups were chosen for analysis. In addition, data from the four E gene loci *E1*, *E2*, *E3* and *E4* were added to achieve a wider genome coverage. Thus, 2 to 4 SSR loci per linkage group were available for analysis. Primer sequence information for each individual locus including *E* genes is available either from Žulj Mihaljević et al. (2020), Kurasch et al. (2017) or SoyBase (2022). Further information about all loci utilized is available in Suppl. Table 3 and in the SoyBase database (SoyBase 2022). For PCR reactions, the total volume of the PCR mixture was 10 µl containing 3.3 µl DNA (10 ng/µl), 1×GoTaq® Green Master Mix (Promega), 0.25 pmol forward primer with an M13 tail added to its 5′ end (5′-CCCAGTCACGAC GTTG-3′), 2.5 pmol reverse primer and 2.25 pmol fluorescent labelled M13 tail (FAM, Cy5) synthesized by MWG (Ebensburg, Germany). A 2-step PCR was performed as follows: initial denaturation at 95 °C for 2 min, followed by seven cycles of 45 s at 94 °C, 45 s at 68 °C (with each cycle the annealing temperature decreasing by 2 °C), and of 60 s at 72 °C. Products were subsequently amplified in the second step for 30 cycles at 94 °C for 45 s, 50 °C for 45 s, and 72 °C for 60 s, with a final extension at 72 °C for 5 min.

The PCR amplification products were separated using 12% polyacrylamide gels, 1×TBE buffer in a C.B.S. electrophoresis chamber (C.B.S. Scientific Inc., Del Mar, CA, USA). Electrophoresis conditions were set at constant 400 V and 10 °C for 2 h. Gels were recorded using a Typhoon (GE Healthcare, Uppsala, Sweden) scanner in fluorescent mode. To evaluate the SSR results, alleles were counted manually as "1" for present and "0" for absent for each individual allele. In total 67 SSR markers with 379 alleles (2–11 per locus) were evaluated and a 0/1 table was created. Data from E-gene analysis were also included using all 14 possible alleles of the four E-genes utilized.

SNP genotyping and filtering

A 200 K SNP array was used for further analyzing the 156 soybean cultivars. Genotyping was carried out by the BeadChip with 159,072 SNP markers selected

(A)

(B)

◄**Fig. 1** Neighbor-joining trees illustrating genetic relationships between elite soybean cultivars with tree branch color representing geographic origin, China (CN, in red) vs. Europe (EU, in blue), text background color indicating experimental grouping (representing the maturity groups) and outer ring for country of origin. **a.** Dendrogram of 156 cultivars based on genetic distance in 71 SSR markers. **b.** Dendrogram of 153 cultivars based on genetic distance estimates using the 200 K SNP array

from the No. 1 Zhongdouxin Soybean Breeding Array ZDX1 (Sun et al. 2022) by COMPASS BIOTECHNOLOGY (Beijing Compass Biotechnology Co, Ltd., Beijing, China) through the Illumina iScan platform (Illumina, Inc., San Diego, CA, USA) following a protocol described in Infinium HD Assay Super Manual. The SNP alleles were called using the GenomeStudio genotyping software (V2011.1, Illumina, Inc.). The SNP data set was filtered by PLINK 1.9; variants with missing rates greater than 0.01, samples with missing rates greater than 0.05, and MAF smaller than 0.05 were removed from the data set. Soybean cultivars 100 (Heinong 61), 129 (Heinong 63) and 94 (Mengdou 30) were excluded from further analysis because of poor data quality, and finally 61,316 SNP markers and 153 genotypes passed the quality control for SNP analysis.

Statistical data analysis

Genetic diversity parameters and informativeness of the analyzed SSR loci as well as SSR distances were calculated by Powermaker 3.25 (Liu and Muse 2005) and GenAlEx 6.51 (Peakall and Smouse 2012). Estimation of population genetic parameters for both SSR and SNP marker data as well as AMOVA were carried out by the R package poppr (Kamvar et al. 2015) using R software vers. 4.1.2 (R Core Team 2022). The neighbour joining tree distance matrices were established by APE 5.0 package based on genetic distance and visualized by iTOL 6.4.3 (Letunic and Bork 2021). For SSR marker data, principal component analysis (PCA) was carried out by poppr package, while for the SNP data PCA was carried out using PLINK 1.9 (Chang et al. 2015). Ten components were calculated for each of the two marker types, and the two most significant components were utilized for graphical visualization of genetic diversity using GGPLOT2 (Wickham 2009). The population structure was analyzed by the software tools STRUCTURE 2.3.1 (Pritchard et al. 2000) for SSR

and ADMIXTURE (Alexander et al. 2009) for SNP data with a proposed grouping (K) from 1 to 10. Bar graphs for revealing population structure were drawn by POPHELPER (Francis 2017) package. The linkage disequilibrium (LD) analysis was performed in POPLDDECAY (Zhang et al. 2019) software, LD values (r^2) were calculated for all pairwise SNP combinations located at a maximum of 1000 Kb distance.

Results

Genetic diversity between Chinese and European cultivars

The elite soybean cultivars were clearly divided into two major groups of genotypes corresponding to the two regions of origin, i.e. China and Europe (Fig. 1a and b), and both marker systems revealed comparable major groupings. The clear regional separation between Chinese and European cultivars was confirmed as well by the scattering pattern of genotypes in principal component analysis (PCA), as indicated in Fig. 2a and b for SSR-markers and the SNP-array, respectively. As disclosed by the two dendrograms, clustering within the regions is largely based on maturity classification (text background color in Fig. 1a and b) with the early maturity experimental group (MG 0000, 000, 00) clearly separated from later groups, particularly in the European cultivars and most clearly visible in SNP-analysis (Fig. 1b). In addition to maturity, European cultivars tended to cluster in a country-wise manner: Both in SSR-and SNP-based dendrograms, Italian, Austrian, Serbian or Hungarian cultivars tended to fall into separate clusters thus representing particular breeding programs and genetic similarity within programs. In contrast, Chinese cultivars were partly clustered according to the experimental grouping, but the clustering based on institutions (i.e. individual breeding programs) was less reproducible than for the European region when comparing both marker systems. While both of the neighbor-joining tree dendrograms seemingly indicate European cultivars to be a sub-cluster within a Chinese cluster, this view is not supported by the PCA. Moreover, the Chinese cluster to which the European cultivars seem to be agglomerated to consists of different genotypes in the SSR and SNP based dendrograms indicating that the neighbor tree does

Fig. 2 Scatter plots of principal component analyses (PCA) of Chinese (CN) and European (EU) elite soybean cultivars based on either SSR markers **a** or the 200 K SNP array **b**
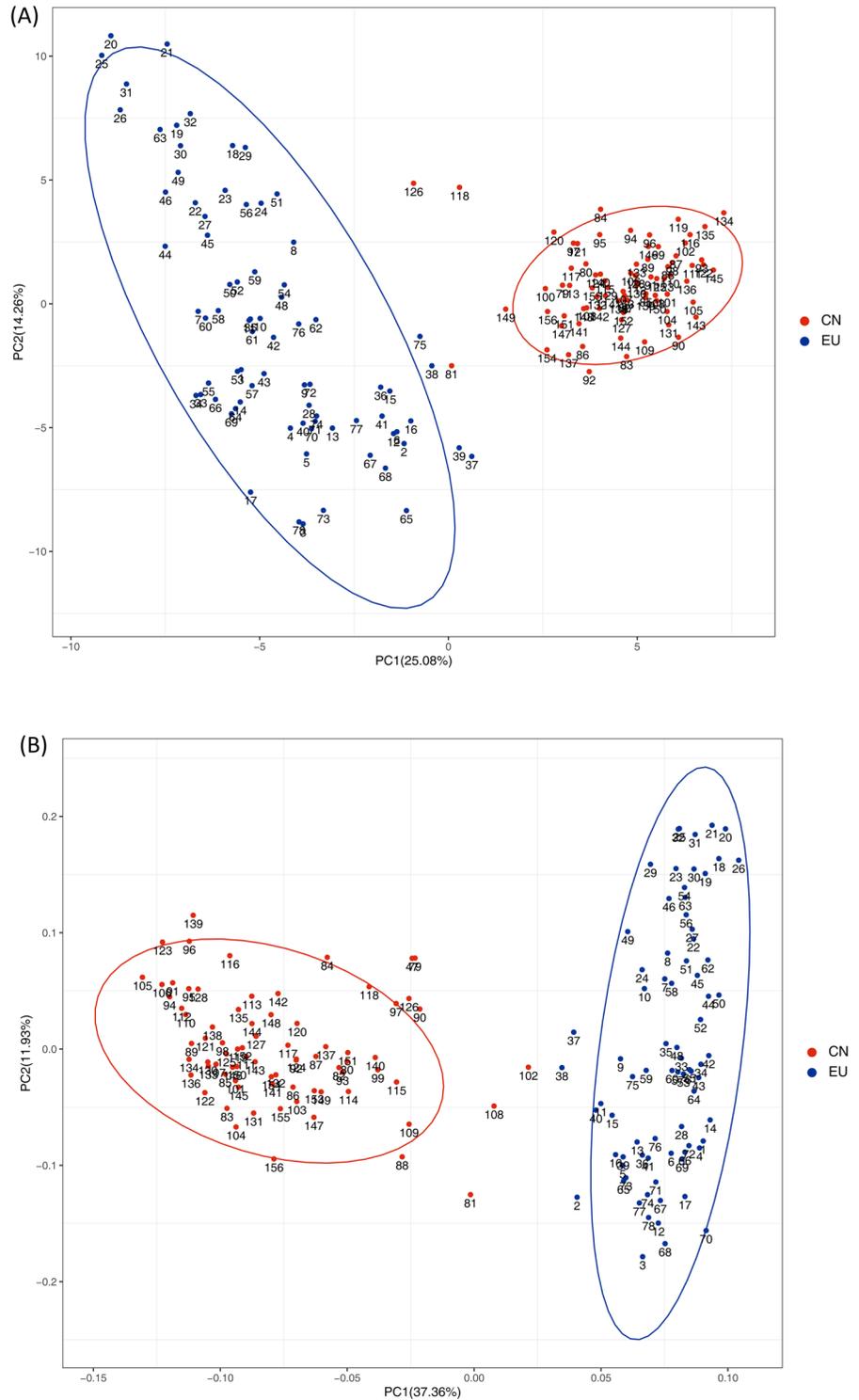
not reveal particular Chinese ancestors for the European cultivars.

The two dendrograms of Fig. 1 suggest a slightly larger diversity within the European than within the Chinese set of cultivars, as indicated by longer tree branches for the European cultivars as compared to the Chinese ones. Similarly, the Chinese cultivars are more densely scattered in PCA analysis (Fig. 2) thus suggesting a comparatively higher diversity to be present in European cultivars. Differences in diversity are also reflected by a comparison of average genetic distances within and between experimental groups across the two regions, as indicated in Suppl. Table 5 for SSR-derived differences. For experimental groups 1 and 4, distances are larger within the European than within the Chinese region, while average differences within Chinese experimental groups are slightly higher for groups 2 and 3. As expectable, average genetic distances between experimental groups are highest between the two regions (Suppl. Table 5). A similar pattern was found with SNP-array distances (detail results not shown). All pairwise genetic distances between cultivars calculated either from SSR markers or the SNP-array were also used to examine the concordance between the two marker systems; a clear correlation between the two types of distance measures is illustrating the agreement between marker systems in estimating the distance between genotypes (Suppl. Figure 1). This agreement is even closer when based on average distances within or between European and Chinese experimental groups (r = 0.965, Suppl. Figure 2).

The diversity pattern between genotypes was further investigated by analysis of molecular variance (AMOVA, Tables 1 and 2). In both marker systems, the percentage of variation was higher within populations than between populations both for geographical origin and for maturity groups which indicates the presence of additional factors relevant for variation.

Allelic diversity

For the 71 SSR loci, 392 different alleles were detected in total among all 156 cultivars. The number of alleles and the number of private (unique) alleles were both higher in the Chinese than in the European population (Table 3). With respect to maturity group, the number of alleles was highest for maturity groups 0 and I, and it was lower for early maturity groups 000 and 00 as well as for the later maturity group II. An individual description of all SSR loci with their respective number of alleles per locus as well as other parameters of genetic diversity and informativeness is provided in Suppl. Table. 3. While the average number of different alleles per SSR locus was 5.5, a maximum of 11 alleles could be identified in one particular locus. In comparing of SNP-derived population genetic estimators between European and Chinese cultivars, indices were similar except for the index of association (Ia) between the two regions (Table 4).

Table 1 Molecular analysis of variance (AMOVA) for geographical origin (CN vs. EU) and soybean maturity groups based on SSR markers

Source	Geographical origin					Maturity group				
	df	SS	MS	Est. var	%	df	SS	MS	Est. var	%
Between pops	1	581.622	581.622	6.932	14	4	389.396	97.349	1.840	4
Within pops	154	6320.648	41.043	41.043	86	149	6416.890	43.066	43.066	96
Total	155	6902.269		47.975	100	153	6806.286		44.906	100

Table 2 AMOVA analysis for geographical origin (CN vs. EU) and soybean maturity groups based on SNP data

Source	Geographical origin					Maturity group				
	df	SS	MS	Sigma	%	df	SS	MS	Sigma	%
Between pops	1	560,572	560,572	3483.5	21.7	6	278,335	46,392	491.7	3.4
Within pops	152	3,665,209	24,113	11,543.5	71.9	147	3,947,426	26,853	12,913.5	89.5
Within indiv	154	158,043	1026	1026.3	6.4	154	158,043	1026	1026.3	7.1
Total	307	4,383,825	14,279	16,053.2	100	307	4,383,825	14,279	14,431.4	100

Table 3 Comparison of population genetic estimators based on SSR markers for geographical origin and maturity group of elite soybean cultivars

Population	Geographical origin		Maturity group*				
	Europe	China	000	00	0	I	II
Number of alleles	327	337	265	276	330	313	269
Alleles with freq. > = 5%	236	241	265	229	251	243	221
Number of private alleles	56	66	11	9	14	17	9
Mean He	0.156	0.155	0.151	0.155	0.160	0.154	0.157

*Maturity groups 0000 and III were removed because only one genotype was assigned to each of these maturity groups

Table 4 Comparison of population genetic estimators between European and Chinese elite soybean cultivars based on SNP data

Population	N	MLG	H	lambda	H_{exp}	Ia	rbarD
Europe	77	77	4.34	0.987	0.289	534	0.0142
China	77	77	4.34	0.987	0.273	711	0.0181
Total	154	154	5.04	0.994	0.324	827	0.0193

N–number of individuals in the population; MLG-number of multilocus genotypes found in the specified population; H–Shannon–Wiener Diversity index; lamda–Simpson's index; H_{exp}–Nei's gene diversity (expected heterozygosity); Ia-value of the Index of Association for each population factor; rbarD-value of the Standardized Index of Association for each population factor

Genetic structure

An analysis of the population structure based on SNP data is presented in Fig. 3. Both for K = 2 and for K = 5 (optimum cross validation error), the separation across the two regions is clearly evident. Chinese cultivar no.s 102 (Heinong 51), 108 (Suinong 24) and 81 (Dongnong 50) were identified to contain a higher proportion of European ancestry; in Fig. 2b, these cultivars were also grouped as PCA outliers of the Chinese cluster located closest to the European cluster. Likewise, the cultivar no.s 40 (NS Kaca) and 11 (Atlanta) were containing the largest proportion of Chinese ancestry among all European cultivars. Moreover, at K = 5, the European sub-population is structured into two ancestral lines (represented by dark blue and green bars), whereas the Chinese sub-population is classified into three ancestral lines (yellow, red and light blue bars). For comparison, an additional structure analysis has been carried out using SSR data (Suppl. Figure 3) for K = 2 and K = 5: Also there, the separation of cultivars according to their region of origin is clearly given at K = 2.

LD decay

The average LD decay across all chromosomes was calculated for the whole cultivar set and separately for the Chinese and European sub-populations. In the European sub-population, LD decay to its half is reached faster (at 110 Kbp distance) than in the Chinese sub-population (135 Kbp).

Discussion

Genetic and allelic diversity

Elite soybeans from the Northeastern spring and Northern spring regions of China and early maturity European elite soybean cultivars are genetically separated from each other as independently revealed by SNP and SSR marker analysis (Figs. 1 and 2). This clearly indicates that the two soybean sets are representing different world soybean populations as proposed by Liu et al. (2020a). In their world soybean classification based on historical dissemination patterns and phylogeographical relationships, soybean populations from the north of China, far east Russia, north of North America and north of Europe (represented by Swedish accessions only) are forming one
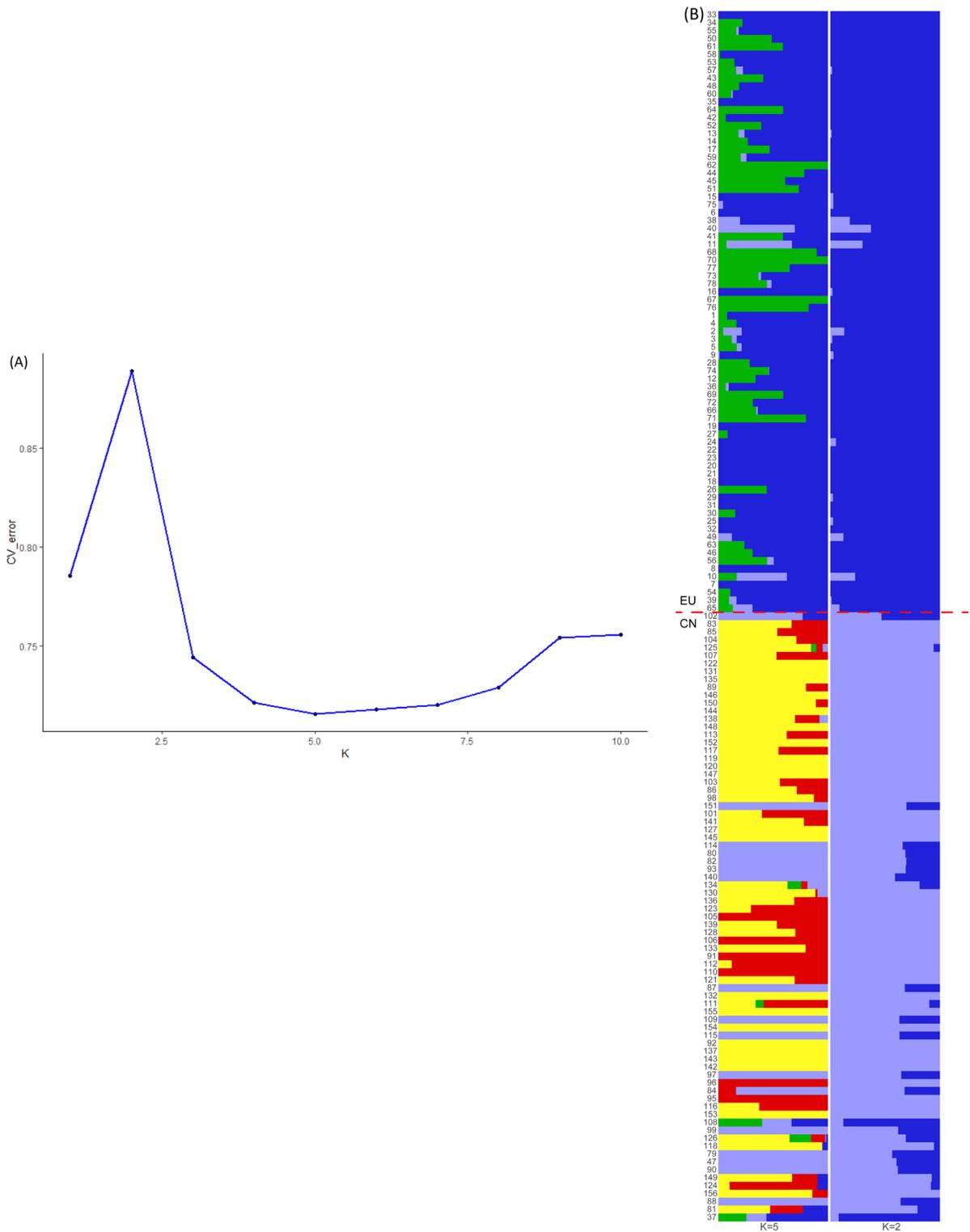
Fig. 3 Analysis of population genetic structure of 153 soybean cultivars based on 200 K SNP array data. **a** Line graph of cross validation (CV) errors of K values for 1–10 with minimum at K = 5. **b** Population structure for K = 2 (top) and K = 5 (bottom) indicating the proportion of membership of each cultivar in two or five hypothetical subpopulations, respectively

cluster according to phenotypic data considering day-length- and temperature-sensitive adaptation traits, whereas genetically these populations are clustering differently thus confirming the present results with respect to north Chinese and European cultivars. Similarly, Chinese and American soybean cultivars were also described as having clearly distinct genetic bases (Liu et al. 2017).

The continuous reduction of genetic diversity from soybean domestication through the generation of landraces, ancestors of selective breeding and the subsequent development of present-day elite cultivars is well documented, and several genetic bottlenecks have been identified which caused the loss of rare alleles or larger structural variants (Hyten et al. 2006; Liu et al. 2020b). However, on the level of elite cultivars covered in the present study, a lower level of diversity also would have been assumed for European cultivars as compared to Chinese ones due to the much shorter growing history and lower extend of breeding as well as a rather narrow genetic base in Europe (Tavaud-Pirra et al. 2009). In contrast to this assumption, the level of overall diversity within each the European and the Chinese population appears to be rather similar (Figs. 1, 2, Tables 3, 4) with average genetic distances suggesting slightly higher diversity in Europe for experimental groups 1 and 4 vs. higher diversity in China for groups 2 and 3, respectively (Suppl. Table 5). However, as Chinese cultivars were represented in lower numbers in experimental groups 1 and 4 (Suppl. Table 2) than European ones, this might also add to lower estimates of diversity within these groups here. The finding of a remarkable European diversity is corroborating earlier research (Hahn and Würschum 2014) which identified significant genetic variation existing in Central European soybean germplasm. Moreover, the present results also demonstrate that genetic diversity can be maintained by breeding and selection which was similarly shown for Canadian soybean diversity (Bruce et al. 2019) or for North American ancestors vs. elite cultivars (Hyten et al. 2006).

Within the populations, soybean cultivars were grouped according to their maturity classification. This was particularly evident for the European population, which was additionally grouped according to their country of origin (Fig. 1b). In a comparable set of European cultivars, genotypes also clearly clustered according to maturity group and country of cultivar origin (Žulj Mihaljević et al. 2020). Similarly, time to maturity, geographical location of a breeding program and specific breeding decisions have been identified as major factors affecting diversity of North American public soybeans (Gizlice et al. 1996). Thus, the country-wise grouping (Fig. 1b) probably indicates different breeding programs as well as regional adaptation due to specific environmental conditions between the different European soybean growing regions. Within the Chinese elite population, however, the clustering of cultivars according to the experimental maturity grouping or according to breeding institutions was less clear. This might be due to much larger and more homogenous soybean production regions in China as compared to Europe, and a classification scheme of cultivars into primary ecotypes such as the Northeastern Spring or Northern Spring soybeans (Wang et al. 2006) rather than more narrow maturity groups. In addition, this is also supported by AMOVA results (Tables 1 and 2) indicating a larger percentage of variation attributable to geographic origin rather than to maturity group.

The number of different SSR-alleles was slightly higher in the Chinese than in the European population (Table 3). Remarkably, however, 56 private alleles were found in the European set of cultivars which were not present in the Chinese set. These alleles might have been lost during previous cycles of selection in one region, or they reveal the occurrence of new mutations developing from previous alleles. The considerable number of private alleles found both in the Chinese and European population is also indicating that the populations have been derived from clearly different gene pools with different ancestral lines (Viana et al. 2022). In addition, differences in allele distribution of particular SSR loci between populations might as well indicate signatures of selection for adaptation to particular environments (Tomicic et al. 2015). While the mean number of SSR alleles per locus was 5.5 across the two populations of the present study (Suppl. Table 3) and similar to the sets of Žulj Mihaljević et al. (2020) or Tavaud-Pirra et al. (2009), numbers of alleles were lower in Serbian (Tomicic et al. 2015) and Indian (Kumar et al. 2022) sets of elite cultivars. In contrast, the numbers of alleles were considerably higher in several sets of soybean accessions originating from Korea (Hwang et al. 2020; Lee et al. 2014; Song et al. 2013). Particular SSR loci exhibited a rather high number of

alleles (e.g. Satt281 in Suppl. Table 3, also confirmed by Tavaud-Pirra et al. (2009)), whereas other polymorphic loci have two alleles only across the whole population (e.g. locus SacK149 in Suppl. Table 3, for which the two alleles are associated with low or high cadmium (Cd) uptake from soil (Vollmann et al. 2015), and additional alleles would be of interest in view of potential phenotypic effects). In individual SSR loci, differences in numbers of alleles and other parameters between Chinese and European populations were found (Suppl. Table 4) indicating selection effects in the two regions of adaptation.

Genetic structure

The clear separation between Chinese and European cultivars is also confirmed on the level of structure analysis (Fig. 3). For the European cultivars, the existence of two major ancestral lines (dark blue and green bars in Fig. 3, K=5) was suggested in this analysis. Late maturity European soybeans as represented by Serbian cultivars (Suppl. Table 1) are roughly tracing back to northern U.S. cultivars such as Evans and Hodgson (Hrustic and Miladinovic 2011) which contain ancestral varieties such as Lincoln (pedigree: Mandarin / Manchu) and Richland in their pedigree (Allen and Bhardwaj 1987; SoyBase 2022). As a consequence, the six ancestral varieties Mandarin, Capital, Richland, Lincoln, Strain No. 18 and Mukden have been identified to make about 75% of the parental contribution to south-east European elite cultivars (Tomicic et al. 2015). In contrast, early maturity European soybean cultivars used in central and northern regions of Europe are often tracing back to extremely early germplasm developed in Sweden from germplasm obtained in Sakhalin (Fiskeby, Holmberg varieties) and early maturity Canadian (i.e. Ontario) varieties. Both Canadian and Swiss soybeans such as the widely grown cultivars Maple Arrow (pedigree: Harosoy 63 / Holmberg 840−1−3) or Ceresia (pedigree: Fiskeby V / Harosoy 107−2031−2) have been selected for chilling tolerance and adaptation to cool environments (Yamaguchi et al. 2018). Many modern Canadian, Swiss, German, Polish and Austrian cultivars are related through the use of Swedish early maturity germplasm, e.g. introgression of Fiskeby V through Bicentennial into at least nine North American cultivars, or Maple Arrow being present in the pedigree of almost 20 modern cultivars (SoyBase

2022). Therefore, Canadian, Swiss and German germplasm materials have been described as genetically similar (Hahn and Würschum 2014). Thus, the two ancestral lines structuring the European elite cultivar population (Fig. 3, K=5) might indicate the two different breeding pathways of cultivars from southeast or central-north Europe.

The Chinese elite population is structured into three major ancestral lines which is suggesting a higher level of genetic background variation than for the European population (Fig. 3, K=5, yellow, red and light blue bars). Despite the fact that the series of Heihe or Dongnong cultivars are earlier and Hefeng or Suinong cultivars are clearly later in maturity, the structure within each of these series is not homogeneous. This appears as a clear illustration of a different breeding history between China and Europe: As discussed above, European soybean diversity is largely based on a limited number of distinct plant introductions adapted to small and separated agroecological regions. In contrast, Chinese soybean diversity in the large Northeastern Spring and Northern Spring sowing regions as represented in the present study appears to be more continuous due to multiple ancestral contributions and larger growing areas (Li et al. 2008; Liu et al. 2020b; Wang et al. 2006).

LD decay

Differences in LD decay (Bruce et al. 2019; Contreras-Soto et al. 2017; Viana et al. 2022) may have multiple reasons including size and specifics of population, breeding intensity, or maturity group. The faster LD decay in the European population (Fig. 4) as compared to the Chinese one might again indicate the lower number of ancestors in the European gene pool, because of which relatively more hybridizations per each ancestral line had been carried out thus causing higher rates of recombination; population specific differences in LD decay or greater selection strength in the Chinese population might have contributed to the overall difference as well.

Conclusions

The comparative diversity analysis between Chinese and European elite soybeans has relevant implications

Fig. 4 Comparative estimation of linkage disequilibrium (LD) decay with increasing genetic distance from SNP array data for either all or European (EU) and Chinese (CN) elite soybean cultivars
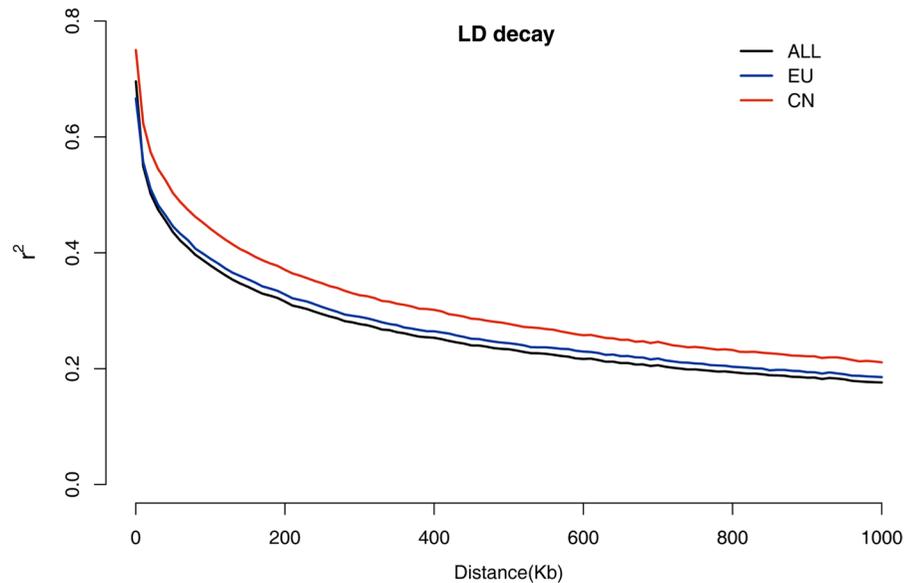
for future soybean breeding in both regions. The level of genetic diversity in elite soybean cultivars is of similar magnitude between the two regions which demonstrates that modern plant breeding for specific target environments within geographical regions can maintain overall genetic variation. Due to regional adaptation needs within Europe, cultivars were separated by maturity classification which was less pronounced for Chinese cultivars. Analysis of population structures suggests that European cultivars are based on two major ancestral lines, whereas Chinese cultivars trace back to more ancestral lines pointing to the rich natural soybean diversity present in China.

Declarations

Conflict of interest The authors declare that they have no conflict of interest related to the subject of research.

References

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19:1655–1664. https://doi.org/10.1101/gr.094052.109

Allen FL, Bhardwaj HL (1987) Genetic relationships and selected pedigree diagrams of North American soybean cultivars. University of Tennessee Agricultural Experiment Station, Knoxville, TN, Bulletin, p 652

Bayer P, Valliyodan B, Hu H, Marsh J, Yuan Y, Vuong TD, Patil G, Song Q, Batley J, Varshney RK, Lam HM,

Edwards D, Nguyen HT (2022) Sequencing the USDA core soybean collection reveals gene loss during domestication and breeding. Plant Genome 15:e20109. https://doi.org/10.1002/tpg2.20109

Bruce RW, Torkamaneh D, Grainger C, Belzile F, Eskandari M, Rajcan I (2019) Genome-wide genetic diversity is maintained through decades of soybean breeding in Canada. Theor Appl Genet 132:3089–3100. https://doi.org/10.1007/s00122-019-03408-y

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4:7. https://doi.org/10.1186/s13742-015-0047-8

Contreras-Soto RI, de Oliveira MB, Costenaro-da-Silva D, Scapim CA, Schuster I (2017) Population structure, genetic relatedness and linkage disequilibrium blocks in cultivars of tropical soybean (*Glycine max*). Euphytica 213(8):1–12. https://doi.org/10.1007/s10681-017-1966-5

Czembor E, Czembor JH, Suchecki R, Watson-Haigh NS (2021) DArT-based evaluation of soybean germplasm from Polish gene bank. BMC Res Notes 14:343. https://doi.org/10.1186/s13104-021-05750-1

FAOSTAT (2022) Food and agriculture data. FAO, Rome. URL: https://www.fao.org/faostat/en/#home (accessed May 5, 2022)

Francis RM (2017) Pophelper: an R package and web app to analyse and visualize population structure. Mol Ecol Resour 17(1):27–32. https://doi.org/10.1111/1755-0998.12509

Gizlice Z, Carter TE Jr, Burton JW (1994) Genetic base for North American public soybean cultivars released between 1947 and 1988. Crop Sci 34(5):1143–1151

Gizlice Z, Carter TE Jr, Gerig TM, Burton JW (1996) Genetic diversity patterns in North American public soybean cultivars based on coefficient of parentage. Crop Sci 36(3):753–765

Hahn V, Würschum T (2014) Molecular genetic characterization of Central European soybean breeding germplasm. Plant Breed 133:748–755. https://doi.org/10.1111/pbr.12212

Hartman GL, West ED, Herman TK (2011) Crops that feed the World 2. Soybean-worldwide production, use, and constraints caused by pathogens and pests. Food Sec 3:5–17

Hrustic M, Miladinovic J (2011) Importance, origin and expansion of soybean. In: Miladinovic J, Hrustic M, Vidic M (eds) soybean. Institute of Field and Vegetable Crops, Novi Sad, Serbia, pp 11–44

Hwang TY, Gwak BS, Sung J, Kim HS (2020) Genetic diversity patterns and discrimination of 172 Korean soybean (Glycine max (L.) Merrill) varieties based on SSR analysis. Agriculture 10(3):77. https://doi.org/10.3390/agriculture10030077

Hyten DL, Song Q, Zhu Y, Choi IY, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB (2006) Impacts of genetic bottlenecks on soybean genome diversity. Proc Natl Acad Sci 103(45):16666–16671

Kamvar ZN, Brooks JC, Grünwald NJ (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front Genet 6:208. https://doi.org/10.3389/fgene.2015.00208

Kisha TJ, Sneller CH, Diers BW (1997) Relationship between genetic distance among parents and genetic variance in populations of soybean. Crop Sci 37:1317–1325. https://doi.org/10.2135/cropsci1997.0011183X003700040048x

Kumar SP, Susmita C, Sripathy KV, Agarwal DK, Pal G, Singh AN, Kumar S, Kumar Rai A, Simal-Gandara J (2022) Molecular characterization and genetic diversity studies of Indian soybean (Glycine max (L.) Merr.) cultivars using SSR markers. Mol Biol Rep 49(3):2129–2140. https://doi.org/10.1007/s11033-021-07030-4

Kurasch AK, Hahn V, Leiser WL, Vollmann J, Schori A, Bétrix C-A, Mayr B, Winkler J, Mechtler K, Aper J, Sudaric A, Pejic I, Sarcevic H, Jeanson P, Balko C, Signor M, Miceli F, Strijk P, Rietman H, Muresanu E, Djordjevic V, Pospišil A, Barion G, Weigold P, Streng S, Krön M, Würschum T (2017) Identification of mega-environments in Europe and effect of allelic variation at maturity *E* loci on adaptation of European soybean. Plant Cell Environm 40:765–778. https://doi.org/10.1111/pce.12896

Lee GA, Choi YM, Yi JY, Chung JW, Lee MC, Ma KH, Lee SY, Cho JW, Lee JR (2014) Genetic diversity and population structure of Korean soybean collection using 75 microsatellite markers. Korean J Crop Sci 59(4):492–497. https://doi.org/10.7740/kjcs.2014.59.4.492

Letunic I, Bork P (2021) Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49(W1):W293–W296. https://doi.org/10.1093/nar/gkab301

Li Y, Guan R, Liu Z, Ma Y, Wang L, Li L, Lin F, Luan W, Chen P, Yan Z, Guan Y, Zhu L, Ning X, Smulders MJM, Li W, Piao R, Cui Y, Yu Z, Guan M, Chang R, Hou A, Shi A, Zhang B, Zhu S, Qiu L (2008) Genetic structure and diversity of cultivated soybean (*Glycine max* (L.) Merr.) landraces in China. Theor Appl Genet 117:857–871. https://doi.org/10.1007/s00122-008-0825-0

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129. https://doi.org/10.1093/bioinformatics/bti282

Liu X, He J, Wang Y, Xing G, Li Y, Yang S, Zhao T, Gai J (2020a) Geographic differentiation and phylogeographic relationships among world soybean populations. Crop J 8(2):260–272. https://doi.org/10.1016/j.cj.2019.09.010

Liu Y, Du H, Li P, Shen Y, Peng H, Liu S, Zhou G, Zhang H, Liu Z, Shi M, Huang X, Li Y, Wang ZMZ, Zhu B, Han B, Liang C, Tian Z (2020b) Pan-genome of wild and cultivated soybeans. Cell 182(1):162–176. https://doi.org/10.1016/j.cell.2020.05.023

Liu Z, Li H, Wen Z, Fan X, Li Y, Guan R, Guo Y, Wang S, Wang D, Qiu L (2017) Comparison of genetic diversity between Chinese and American soybean (Glycine max (L.) accessions revealed by high-density SNPs. Front Plant Sci 8:2014. https://doi.org/10.3389/fpls.2017.02014

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in excel. population genetic software for teaching and research-an update. Bioinformatics 28:2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959. https://doi.org/10.1093/genetics/155.2.945

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Sedivy EJ, Wu F, Hanzawa Y (2017) Soybean domestication: the origin, genetic architecture and molecular bases. New Phytol 214(2):539–553. https://doi.org/10.1111/nph.14418

Song JY, Piao XM, Choi YM, Lee GA, Chung JW, Lee JR, Jung Y, Park HJ, Lee MC (2013) Evaluation of genetic diversity and comparison of biochemical traits of soybean (Glycine max L) germplasm collections. Plant Breed Biotechnol 1(4):374–384. https://doi.org/10.9787/PBB.2013.1.4.374

SoyBase (2022) SoyBase, integrating genetics and genomics to advance soybean research. URL: http://soybase.org. (Accessed March 1, 2022)

Sun R, Sun B, Tian Y, Su S, Zhang Y, Zhang W, Guo B, Li H, Li Y, Gao H, Gu Y, Yu L, Ma Y, Wang J, Yu P, Su E, Li Q, Hu X, Zhang Q, Guo R, Chai S, Feng L, Wang J, Hong H, Xu J, Wen J, Liu J, Li Y, Qiu L (2022) Dissection of the practical soybean breeding pipeline by developing ZDX1, a high-throughput functional array. Theor Appl Genet 135(4):1413–1427

Tavaud-Pirra M, Sartre P, Nelson R, Santoni S, Texier N, Roumet P (2009) Genetic diversity in a soybean collection. Crop Sci 49(3):895–902. https://doi.org/10.2135/cropsci2008.05.0266

Tomicic M, Djordjevic V, Obreht D, Miladinovic J, Brbaklic L, Mikic A, Mikic S (2015) Tracking footprints of selection associated with soybean adaptation to Central-East Europe environments. Euphytica 203(3):701–713. https://doi.org/10.1007/s10681-014-1313-z

Viana JPG, Fang Y, Avalos A, Song Q, Nelson R, Hudson ME (2022) Impact of multiple selective breeding programs on genetic diversity in soybean germplasm. Theor Appl Genet 135:1591–1602. https://doi.org/10.1007/s00122-022-04056-5

Vollmann J, Lošák T, Pachner M, Watanabe D, Musilová L, Hlušek J (2015) Soybean cadmium concentration: validation of a QTL affecting seed cadmium accumulation for improved food safety. Euphytica 203(1):177–184. https://doi.org/10.1007/s10681-014-1297-8

Wang L, Guan R, Zhangxiong L, Chang R, Qiu L (2006) Genetic diversity of Chinese cultivated soybean revealed by SSR markers. Crop Sci 46(3):1032–1038. https://doi.org/10.2135/cropsci2005.0051

Wickham H (2009) ggplot2: Elegant graphics for data analysis. Springer, New York

Yamaguchi N, Ohnishi S, Miyoshi T (2018) Screening for chilling-tolerant soybeans at the flowering stage using a seed yield-and maturity-based evaluation method. Crop Sci 58(1):312–320

Zhang C, Dong SS, Xu JY, He WM, Yang TL (2019) PopLD-decay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics 35(10):1786–1788. https://doi.org/10.1093/bioinformatics/bty875

Žulj Mihaljević M, Šarčević H, Lovrić A, Andrijanić Z, Sudarić A, Jukić G, Pejić I (2020) Genetic diversity of European commercial soybean [Glycine max (L.) Merr.] germplasm revealed by SSR markers. Genet Res Crop Evolut 67(6):1587–1600. https://doi.org/10.1007/s10722-020-00934-3