



# Genetic diversity of European commercial soybean [*Glycine max* (L.) Merr.] germplasm revealed by SSR markers

Maja Žulj Mihaljević · Hrvoje Šarčević · Ana Lovrić · Zoe Andrijanić · Aleksandra Sudarić · Goran Jukić · Ivan Pejić

Received: 23 October 2019 / Accepted: 3 April 2020 / Published online: 18 April 2020  
© Springer Nature B.V. 2020

**Abstract** There are numerous soybean [*Glycine max* (L.) Merr.] breeding programs in Europe focused on development of elite non genetically modified (GM) cultivars for fast growing market of GM-free proteins for animal feed. Due to low variability of visual descriptors and mostly unknown pedigrees, divergent parents' selection for crosses is a great challenge. Another challenge is cultivar distinction and protection of plant breeders' rights of ever-increasing number of cultivars. By using 42 microsatellite

(SSR) markers, we performed characterization of 97 commercial soybean cultivars and experimental lines developed at various research and breeding institutions in Europe (86) and in North and South America (11) in order to assess their genotype distinction power as well as utility for estimating genetic diversity and genetic structure. A set of 27 most polymorphic SSR markers was sufficient to discriminate all 97 genotypes. Discrimination of, by pedigree very related cultivars, was somewhat difficult due to the low polymorphism but still possible. Cluster analysis showed that European germplasm is mainly distributed into clusters reflecting breeding programs and maturity groups. Performed genetic

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s10722-020-00934-3>) contains supplementary material, which is available to authorized users.

M. Žulj Mihaljević · H. Šarčević (✉) · A. Lovrić · Z. Andrijanić · I. Pejić  
Department of Plant Breeding, Genetics and Biometrics,  
Faculty of Agriculture, University of Zagreb,  
Svetošimunska 25, 10000 Zagreb, Croatia  
e-mail: hsarcevic@agr.hr

H. Šarčević · Z. Andrijanić · A. Sudarić · I. Pejić  
Centre of Excellence for Biodiversity and Molecular Plant  
Breeding (CoE CroP-BioDiv), Faculty of Agriculture,  
University of Zagreb, Svetošimunska 25, 10000 Zagreb,  
Croatia

A. Sudarić  
Agricultural Institute Osijek, Južno predgrađe 17,  
31000 Osijek, Croatia

G. Jukić  
Croatian Agency for Agriculture and Food - Centre for  
Seed and Seedling Production, Vinkovačka c. 63c,  
31000 Osijek, Croatia

characterization provides an insight into genetic structure of European soybean germplasm and might serve as a starting point for future breeding decisions. Disclosed SSR data of the analyzed commercial European germplasm can serve for genetic fingerprinting purpose as well as for foundation of public soybean cultivar database.

**Keywords** Genetic diversity · Soybean · SSR · Microsatellites · Cultivar

## Introduction

Cultivated soybean [*Glycine max* (L.) Merr.] is one of the most important crops used for animal feed and human foods. In 2017 soybean hectareage worldwide was 123.55 million ha which makes approximately 8.5% of the total world's agricultural land (FAOSTAT 2019). The biggest producers are United States, Brazil and Argentina and majority of cultivars used are genetically engineered for glyphosate-resistance, accounting for 80% of the total soybean production worldwide (ISAAA 2017). Demand for soybean is growing on a worldwide basis and according to USDA 10-year projections (USDA 2019) it will increase by 23.4%. Europe accounted for only 5% of the total soybean production in 2017 (FAOSTAT 2019) what makes it a big importer of this commodity. Since genetically modified (GM) cultivars are prohibited for growing in the European Union, there are numerous breeding programs aiming at development of elite non-GM cultivars that would ensure GM-free supply of soybean protein for animal feed. Initiatives like European Soya Declaration, signed by 13 EU soybean producers, are promoting non-GM soybean production ([www.donausoja.org](http://www.donausoja.org)).

China is the primary soybean center of diversity, with many Asian countries as secondary centers of diversity. Although soybean has been spread across Asian countries two to three thousand years ago, it was not known in the West until the 18th century (Hymowitz 2008). Higher levels of allelic diversity have been found in wild soybeans than in cultivated ones (Kuroda et al. 2009). According to Carter et al. (2004) genetic diversity has no impact unless it is utilized and in order to utilize the genetic diversity in soybeans, interactions among soybean germplasm

collections and breeding/genetics programs in both the public and private sectors are necessary. Due to low variability of visual descriptors and mostly unknown pedigrees, divergent parents' selection for crosses is a great challenge. Distinction and protection of plant breeders' rights is complicated by great and ever-increasing number of soybean cultivars. To efficiently broaden the genetic base of modern soybean cultivars a detailed insight into genetic diversity of soybean resources is needed. Such insight might be achieved through molecular characterization using DNA based markers, which are more informative, stable and reliable, compared to pedigree analysis and traditionally used morphological markers (Jamali et al. 2019).

Unlike other crops, e.g. grapevine (This et al. 2004) or wheat (Röder et al. 2002), where international collaborations proposed common sets of SSR markers for germplasm genotyping and publicly available SSR databases exist (This et al. 2004), there are no such joint initiatives active in soybean. There were some attempts like suggestion of 13 most informative SSR markers by Song et al. (1999) and construction of Brazilian microsatellite database (Oliveira et al. 2010) or Indonesian initiative (Widaningsih et al. 2014). As of 2016, US PVP Board adopted the option of using molecular markers (SNPs) as supplementary information for establishing cultivar distinction ([www.ams.usda.gov/services/plant-variety-protection/pvpo-advisory-board](http://www.ams.usda.gov/services/plant-variety-protection/pvpo-advisory-board)). Recently, Sohn et al. (2017) proposed selected set of 27 InDel markers for cultivar identification and construction of a publicly shared soybean database.

Regarding the analysis of European germplasm, Tavaud-Pirra et al. (2009) screened diversity of 350 cultivated soybean genotypes including 185 accessions from INRA soybean collection originating from various European countries as well as 32 cultivars and recent breeding lines representing the genetic improvement of soybean in Western Europe from 1950 to 2000. The results indicated relatively low genetic diversity among West European breeding lines and suggested usefulness of INRA soybean collection for broadening the genetic base of West European germplasm. Similar results were reported by Ristova et al. (2010) through the SSR analysis of Southeast European soybean germplasm and Hahn and Würschum (2014) who analyzed a set of Central European cultivars using DArT and SNP markers.

Pedigrees of European commercial cultivars are generally not public, whereas those of some American soybean breeding material (<https://soybase.org/uniformtrial/index.php?page=lines>) or Brazilian germplasm (Wysmierski and Vello 2013) are accessible. This knowledge, together with the understanding of its genetic structure, is a key presumption for further utilization of soybean germplasm in breeding programs, requirements of genomic selection (Grainger et al. 2018) or genome wide association studies (Bandillo et al. 2015).

The objectives of the present study were: (1) to perform molecular characterization of European commercial soybean cultivars and experimental lines developed at various research and breeding institutions, (2) to assess the discriminatory power of the selected SSR markers and (3) to estimate genetic diversity and genetic structure of European commercial germplasm.

## Materials and methods

### Plant material

Analysis involved 97 soybean genotypes, 91 commercially available cultivars registered in Europe and six soybean lines (marked OS) still under cultivar testing for registration (Table 1). Material has been collected for research purpose only from sale representatives, breeders and reliable breeding collections. Material originating from European breeding programs is represented by following number of genotypes: Italy (27), Croatia (23), Romania (11), France (9), Serbia (5), Austria (4), Switzerland (4) and Germany (2). In addition, five cultivars from Canada, four from United States and three from Argentina were also included in the study because of their good performance in Europe and usage in European breeding programs for broadening of genetic base. The studied material is proved to be adapted to European growing conditions and comprised maturity groups from 000 to II. Available information about breeding institutions and origin were retracted from CPVO (Community Plant Variety Office) database ([cpvoextranet.cpvo.europa.eu](http://cpvoextranet.cpvo.europa.eu)). All studied genotypes were non-GM.

### DNA extraction and SSR genotyping

Plant tissue for DNA extraction was collected from indoor germinated seeds. Five fresh young leaves from two-week old seedlings of each genotype were collected, freeze dried and bulked prior to DNA extraction. Each sample was grounded to a fine powder. Total genomic DNA was extracted from 20 mg of freeze-dried bulked leaf tissue using peqGOLD Plant DNA Mini Kit (PEQLAB Biotechnology GmbH, Germany). Initially, 47 microsatellite loci, at least two per soybean chromosome, were chosen based on their polymorphism reported in similar studies, position in the genome and their annealing temperature (Online resource 1). Information about their sequences and motifs were retrieved from SoyBase database (SoyBase 2018) constructed by Cregan et al. (1999) and Song et al. (2004). After screening, five loci were discarded for their bad amplification or inability to clearly distinguish alleles due to the peaks' architecture.

All forward primers were labeled with 6-FAM, VIC, PET, or NED fluorescent dyes. Singleplex PCR amplifications were carried out in a 10 µl reaction mixture containing 10 ng of template DNA, 0.3 µM of forward and reverse primer, 200 µM of each of the four dNTPs, 2 mM of MgCl<sub>2</sub>, 0.75 U Taq polymerase (Sigma-Aldrich, Germany), 0.5 X CES PCR enhancer and 1 X PCR buffer. The following thermal cycling protocol was applied for all loci: pre-cycle at 94 °C for 2 min; 30 cycles of denaturation for 1 min at 94 °C, 45 s of annealing at 45/50/53/55 °C (depending on locus) and 45 s extension at 72 °C; post-cycle of 10 min at 72 °C and then termination at 4 °C. PCR amplifications were carried out in a Veriti thermal cycler (Applied Biosystems, Foster City, CA). Amplified products were separated on an ABI3130 Genetic Analyzer (Applied Biosystems) with a GeneScan-500 LIZ size standard (Thermo Fisher Scientific, Waltham, MA). The fragments (alleles) of the studied cultivars were sized with GeneMapper 4.0 software (Applied Biosystems).

### Data analysis

SSR alleles were recorded as their full length in base pairs, in the whole number approach/proximity. Based upon SSR profiles Polymorphism Information Content (PIC) was calculated according to Botstein et al.

**Table 1** List of analyzed soybean cultivars and breeding lines

Cultivar/line name	Seed provider	Breeding institution	Maturity group	Country of origin
Abelina	AGR <sup>a</sup>	Saatzucht Donau	000	Austria
Adonai	ERSA	ERSA	I	Italy
Aires	ERSA	ERSA	0	Italy
Almas	ERSA	ERSA	I	Italy
Amandine	AGR	AGROSCOPE	000	Switzerland
Amma	ERSA	ERSA	I	Italy
Ananda	ERSA	ERSA	I	Italy
Annette	ERSA	DON MARIO	I-	Argentina
Ascasubi	BC	ERSA	II	Italy
Astafor	AGR	Euralis	I	France
Avatar	ERSA	ERSA	I	Italy
Bahia	BC	ERSA	0	Italy
Blancas	ERSA	ERSA	I	Italy
Buenos	ERSA	ERSA	II	Italy
Buga	AGR	AGR	0	Croatia
Capnor	AGR	Euralis	000	France
Carla TD	AGR	A.R.D.S. TURDA	00	Romania
Castetis	AGR	AGROSCOPE	I	Switzerland
Celina PZO	ERSA	Pflanzenzucht Oberlimpurg	II	Germany
Columna	AGR	NARDI	0	Romania
Crina F	AGR	NARDI	I	Romania
Daciana	AGR	NARDI	0	Romania
Daisy PZO	ERSA	UGRC	I	Canada
Dekabig	ERSA	DEKALB	II	USA
DH 5170	RWA	Sevita International	I	Canada
DH4173	ERSA	RWA	I	Canada
Diamant	AGR	A.R.D.S. TURDA	000	Romania
DM 503	ERSA	DON MARIO	I-	Argentina
DM 513	ERSA	DON MARIO	0+	Argentina
ES Pallador	ERSA	Euralis	I	France
Duchessa	ERSA	ISEA S.R.L.	I-	Italy
Eiko	ERSA	OILSEEDSSF	II	USA
Em Neve	ERSA	Schillinger Genetics	0+	USA
Ema	PIO	PIO	00-0	Croatia
ES Senator	AGR	Euralis	000	France
ESG 121	AGR	Euralis	0	France
Eugen	AGR	A.R.D.S. TURDA	00	Romania
Felix	AGR	A.R.D.S. TURDA	00	Romania
Flavia	AGR	Saatzucht Donau	0	Austria
Gabriela	AGR	AGR	0	Croatia
Galina	RWA	NS SEME	0	Serbia
Ginerva	ERSA	SYNGENTA	0	Italy
Gioconda	ERSA	ISEA S.R.L.	I	Italy
Giunone	ERSA	SYNGENTA	I	Italy
Guru	ERSA	ERSA	I++	Italy

**Table 1** continued

Cultivar/line name	Seed provider	Breeding institution	Maturity group	Country of origin
Hilario	AGR	ERSA	0	Italy
Ika	PIO	PIO	I	Croatia
Isidor	AGR	Euralis	I	France
Korana	PIO	PIO	0	Croatia
Lissabon	AGR	AgReliant Genetics	000	Canada
Lucija	PIO	PIO	00	Croatia
Malaga	AGR	AgReliant Genetics	000	Canada
Mandala	ERSA	ERSA	00	Italy
Mantra	ERSA	ERSA	0+	Italy
Merkur	RWA	NS SEME	00	Serbia
Namaste	ERSA	ERSA	I	Italy
Nirvana	ERSA	ERSA	I+	Italy
Novosađanka	ZSR	NS SEME	I	Serbia
Oana F	AGR	NARDI	00	Romania
Onix	AGR	A.R.D.S. TURDA	00	Romania
Optimus	AGR	NS SEME	I	Serbia
OS Zora	PIO	PIO	I	Croatia
OS-2017-1	PIO	PIO	0	Croatia
OS-2017-2	PIO	PIO	00-0	Croatia
OS-2017-5	PIO	PIO	0	Croatia
OS-2017-6	PIO	PIO	0	Croatia
OS-2017-7	PIO	PIO	0-00	Croatia
OS-2017-8	PIO	PIO	0	Croatia
P21T45	ERSA	PIONEER	I	Italy
Pedro	BC	ERSA	0	Italy
Pepita	ERSA	ERSA	0+	Italy
Perla	AGR	A.R.D.S. TURDA	000	Romania
PR92M22	ERSA	PIONEER	I	Italy
PR92M35	ERSA	PIONEER	I	Italy
Prana	ERSA	ERSA	I	Italy
Proteinka	ZSR	NS SEME	0	Serbia
Proteix	AGR	AGROSCOPE	00	Switzerland
RGT Sinema	ERSA	RAGT	I	France
Ružica	AGR	AGR	I	Croatia
Sabina	ZSR	SCHLESINGER- SAATZUCHT	00	Germany
Sanda	PIO	PIO	0	Croatia
Sara	PIO	PIO	0-I	Croatia
Seka	PIO	PIO	0-I	Croatia
SG Eider	AGR	Saatbau Linz eGen	I	Austria
SG SR Picor	AGR	Saatbau Linz eGen	0	Austria
Sigalia	AGR	RAGT	00	France
Sonja	PIO	PIO	0	Croatia
Sultana	AGR	RAGT	000	France
Tena	PIO	PIO	I	Croatia
Terrapro	AGR	MAES	0	USA

**Table 1** continued

Cultivar/line name	Seed provider	Breeding institution	Maturity group	Country of origin
Tisa	PIO	PIO	1	Croatia
Toma	PIO	PIO	0	Croatia
Tourmaline	AGR	AGROSCOPE	000	Switzerland
Triumf	AGR	NARDI	1	Romania
Xonia	RWA	ERSA	00	Italy
Zagrepcanka	AGR	AGR	1	Croatia
Zlata	AGR	AGR	1	Croatia

AGR University of Zagreb, Faculty of Agriculture; AGROSCOPE Agroscope Changins Wädenswil; A.R.D.S. TURDA Agricultural Research and Development Station TURDA; BC Bc Institute, Zagreb, Croatia; MAES Minnesota Agricultural Research Station; NARDI National Agricultural Research and Development Institute, Fundulea, Romania; PIO Agricultural Institute Osijek; ERSR Regional Agriculture Agency of Friuli Venezia Giulia; RWA Raiffeisen Agro d.o.o.; DON MARIO Asociados Don Mario S.A.; UGRC University of Guelph, Ridgetown College, Ontario; SYNGENTA Syngenta Crop Protection AG; OILSEEDSSF International Oilseed Distributors; ZSR Croatian Agency for Agriculture and Food - Centre for Seed and Seedling Production

<sup>a</sup>All AGR-provided cultivars originating outside of Croatia were collected within the previous collaborative research of several European countries (Kurasch et al. 2017)

(1980) using Parks' (2008) Microsatellite Toolkit Excel add-in. The same program was also used for calculation of the observed heterozygosity ( $H_o$ ), expected heterozygosity or gene diversity ( $H_e$ ) and allele frequencies. Allelic richness ( $A_r$ ) was calculated according to the formula of El Mousadik and Petit (1996) using FSTAT (Goudet 2001). Probability of identity for two assumptions—unrelated (PI) and related ( $PI_{Sibs}$ ) individuals (Waits et al. 2001), was calculated using GenAlEx (Peakall and Smouse 2012). The genetic dissimilarity coefficients between each pair of the 97 genotypes were calculated as  $1 - P_{SA}$ , where  $P_{SA}$  is the proportion of shared alleles (Bowcock et al. 1994). The dissimilarity matrix was further used for clustering of genotypes with UPGMA algorithm. Construction and dendrogram visualization were done in MEGA software (Kumar et al. 2016).

The choice of minimum number of SSR markers which can distinguish genotypes started with the selection of 10 most informative loci according to their  $H_e$  values. The number of markers was incremented until number of matched genotypes reached zero.

The Bayesian model-based method STRUCTURE, version 2.3.4 (Pritchard et al. 2000) was used to implicate the genetic structure and define the number of groups within the analyzed data set. Genotypes prior to the analysis itself were categorized into the

countries of their origin as stated in Table 1. A total of 10 repetitions were performed for each assumed  $k$  that ranged from 1 to 10. Each repetition consisted of a burn in a period of 100,000 steps followed by 100,000 Monte Carlo Markov Chain iterations under the assumption of the admixed model and the correlated allele frequencies. The choice of the most probable number of groups ( $k$ ) was carried out using Structure Harvester (Earl, 2012) by comparing the average probability estimators, calculated as  $\ln [Pr(X|K)]$  for each  $k$  value, as well as by the  $\Delta K$  calculation based on the change in the probability log of data between successive values, as stated in Evanno et al. (2005). Program CLUMPP 1.1.2. (Jakobsson and Rosenberg 2007) was used to align data between independent repetitions for each  $k$  using the “greedy” algorithm with 10,000 repetitions. Application Structure Plot (Ramasamy et al. 2014) was used for the graphical representation of the groups.

## Results

### Genetic diversity

Profiles of 97 genotypes at 42 SSR loci are presented in Online resource 2. The genotypic data were fitted

into allelic classes based mostly on a difference in two or three base pairs but there were some alleles with consistent difference of only one base pair (e.g. Satt556, alleles 199 and 200).

At 42 SSR loci, a total of 251 alleles were recorded. The allele number at individual loci varied from 2 (Satt287) to 11 (Sat\_109, Sat\_418 and Satt534), as presented in Online resource 3, with the average of 5.98 alleles per locus. Out of these 251 alleles, 94 (37%) were rare with frequency less than 0.05 in the full set. Further 89 alleles (35%) had frequency between 0.05 and 0.25, 46 alleles (18.3%) had frequency between 0.25 and 0.50 while 22 alleles (8.8%) had frequency greater than 0.50. The highest frequency (0.81) was observed for 116 base pairs long allele at the Satt583 locus. Genetic diversity parameters and informativeness of analyzed SSR loci are presented in Table 2. When taking in account the sample size, allelic richness ( $A_r$ ) was assessed as the average of 5.95 alleles per locus. Gene diversity ( $H_e$ ) ranged from 0.32 for Satt583 to 0.81 for Satt185 with the average  $H_e = 0.63$ . In comparison, informativeness of the SSR loci expressed as PIC ranged from 0.30 for Satt583 to 0.78 for Satt185 with the average PIC value of 0.57. Mean observed heterozygosity ( $H_o$ ) across loci was 0.03 and it varied from 0.00 for Sat\_197, Sat\_385, Satt338 and Satt583 to 0.08 for Satt534.

The single-locus PI values ranged from 0.07 (Satt185) to 0.48 (Satt583) whereas the  $PI_{Sibs}$  values ranged from 0.36 (Satt185) to 0.71 (Satt583). The cumulative PI across loci was  $3.94 \times 10^{-32}$ , whereas the cumulative  $PI_{Sibs}$  across loci was  $5.02 \times 10^{-14}$ .

All 97 analyzed genotypes showed unique genetic profiles at 42 SSR loci. However, a minimum of 27 SSR loci was needed to distinguish all analyzed genotypes (Fig. 1). Their cumulative  $PI_{Sibs}$  was  $1.39 \times 10^{-10}$ . The value of  $PI_{Sibs} < 0.0001$ , as a proposed conservative upper bound probability for estimating the number of loci necessary to distinguish individuals (Waits et al., 2001), was achieved in the present study already with 10 most informative SSR loci ( $PI_{Sibs} = 8 \times 10^{-5}$ ). However, using these 10 loci we observed as much as 606 genotype matches (identical multilocus SSR profiles) and 2.7-fold more SSR loci were needed to distinguish all 97 genotypes.

## Genetic relationships

Genetic relationships among the studied genotypes was visualized by a cluster analysis of the dissimilarity matrix using SSR data. The respective UPGMA cluster diagram grouped the 96 soybean genotypes into two major clusters (Fig. 2) comprised of 50 mostly late ripening genotypes and 46 mainly early ripening ones. Lissabon, a cultivar originating from one Canadian breeding program, separated as an outlier. The bigger cluster with 50 genotypes was mostly composed of cultivars originating from Italian breeding programs. It also included all cultivars from USA and Argentina. This cluster is further divided into three sub-clusters. First sub-cluster is dominated by 18 Italian cultivars and five others originating from Western European breeding programs. Second sub-cluster was formed with four Southeast European cultivars and three North American introductions. Third sub-cluster was also centered around Italian cultivars, together with some newer Croatian breeding lines. Second major cluster (46 genotypes) contained only two Italian cultivars and, amongst other, two Canadian cultivars. Its further substructure is dominated by two bigger sub-clusters, one containing mostly Croatian cultivars and the other containing mix of early and very early ripening cultivars originating from the breeding programs from Romania, France and Switzerland.

## Genetic structure

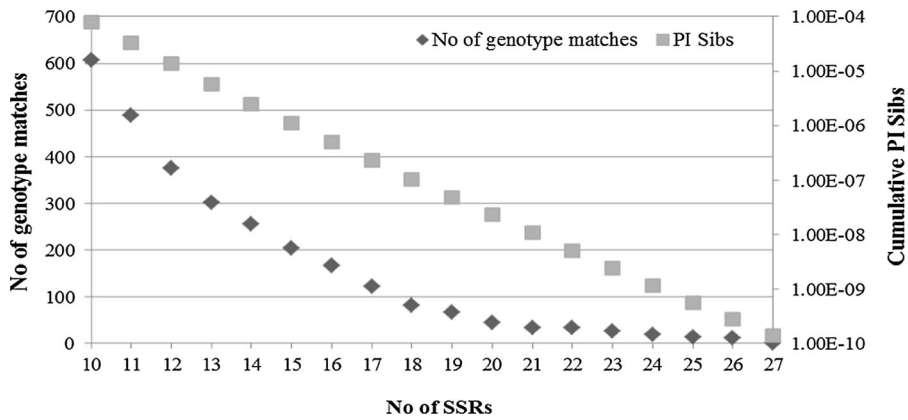
In order to confirm the genetic relationships signalized by UPGMA clustering and to try to comprehend possible genetic structure of European soybean germplasm, model-based approach (STRUCTURE software) was used. There was no clear clustering signal for  $k > 2$  (Fig. 3), so  $k = 2$  was chosen as the best one and used in the attempt to reveal geographic origin of 97 soybean genotypes. From subsequent graphical representation of the share (Fig. 4) it can be noticed that Italian germplasm formed a clear group (ancestral population 1, represented with red color), with 85% of the cultivars in ancestral population 1. Looking separately for each country, Austrian soybean germplasm seems to be of admixed origin. Croatian germplasm originated mostly from ancestral population 2 (represented with yellow color), with

**Table 2** Genetic diversity parameters and informativeness of the analyzed SSR loci: number of analyzed individuals (N), number of alleles per locus ( $N_a$ ), allelic richness ( $A_r$ ), polymorphism information content (PIC) expected heterozygosity ( $H_e$ ), observed heterozygosity ( $H_o$ ), probability of identity—unrelated (PI), probability of identity related ( $PI_{Sibs}$ )

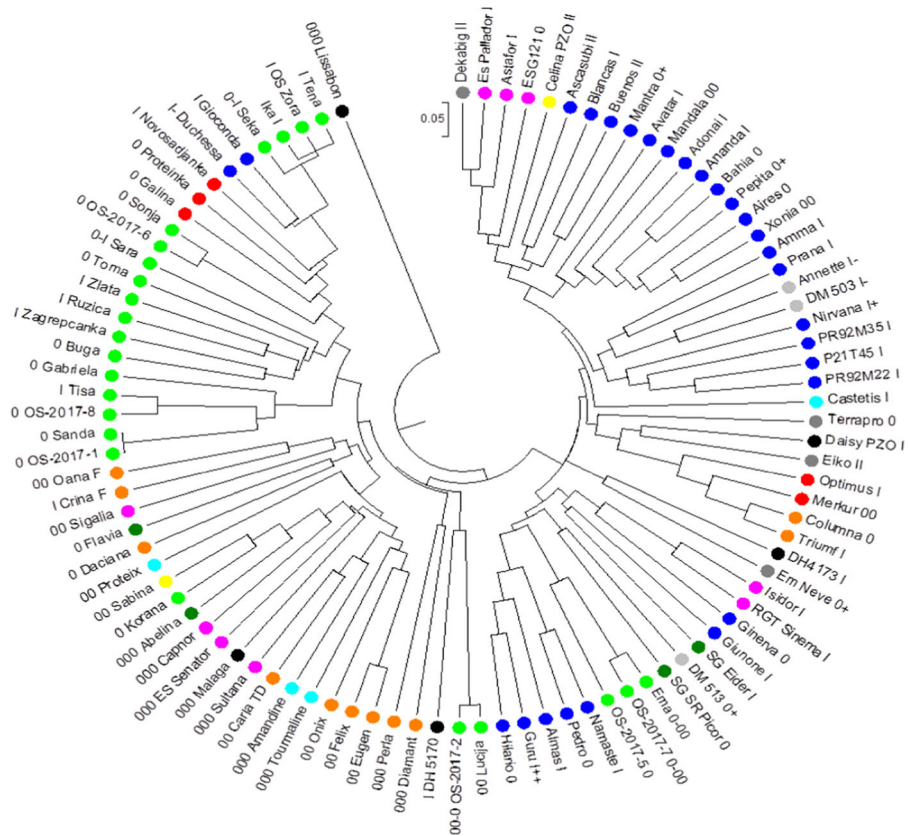
Locus	N	$N_a$	$A_r$	PIC	$H_e$	$H_o$	PI	$PI_{Sibs}$
Sat_109 <sup>15</sup>	96	11	10.97	0.66	0.68	0.02	0.13	0.44
Sat_217 <sup>2</sup>	97	10	9.98	0.76	0.79	0.04	0.07	0.38
Sat_385	97	4	3.97	0.38	0.47	0.00	0.37	0.61
Sat_418 <sup>5</sup>	95	11	10.99	0.73	0.77	0.02	0.09	0.39
Satt005 <sup>11</sup>	97	8	7.85	0.70	0.74	0.07	0.11	0.41
Satt009 <sup>10</sup>	97	8	7.97	0.70	0.74	0.03	0.11	0.41
Satt045	97	7	6.98	0.52	0.57	0.01	0.24	0.53
Satt100 <sup>19</sup>	97	5	5.00	0.62	0.66	0.04	0.16	0.46
Satt142	97	4	3.98	0.49	0.58	0.03	0.27	0.53
Satt147 <sup>20</sup>	97	6	5.98	0.60	0.66	0.04	0.18	0.47
Satt173 <sup>9</sup>	96	9	9.00	0.71	0.75	0.02	0.10	0.40
Satt175 <sup>12</sup>	97	6	6.00	0.69	0.73	0.05	0.12	0.41
Satt181	84	3	3.00	0.40	0.48	0.01	0.35	0.60
Satt185 <sup>1</sup>	97	9	8.97	0.78	0.81	0.01	0.07	0.36
Satt191 <sup>24</sup>	97	4	4.00	0.57	0.64	0.04	0.20	0.48
Satt197	97	5	4.98	0.37	0.40	0.00	0.39	0.65
Satt216	97	3	3.00	0.46	0.56	0.02	0.29	0.55
Satt239 <sup>25</sup>	97	5	5.00	0.56	0.62	0.05	0.21	0.49
Satt242 <sup>27</sup>	97	5	4.97	0.55	0.62	0.02	0.22	0.50
Satt287	97	2	2.00	0.30	0.37	0.01	0.47	0.68
Satt294 <sup>13</sup>	97	5	5.00	0.67	0.72	0.03	0.13	0.42
Satt307	97	5	5.00	0.49	0.56	0.01	0.26	0.54
Satt321	97	3	3.00	0.49	0.56	0.01	0.26	0.54
Satt329 <sup>3</sup>	97	7	6.98	0.75	0.78	0.05	0.08	0.38
Satt338 <sup>16</sup>	97	5	5.00	0.63	0.67	0.00	0.15	0.45
Satt339 <sup>14</sup>	97	6	5.85	0.65	0.71	0.02	0.14	0.43
Satt343 <sup>18</sup>	97	6	5.98	0.62	0.66	0.03	0.15	0.46
Satt373 <sup>7</sup>	97	9	8.97	0.72	0.76	0.01	0.09	0.40
Satt389 <sup>8</sup>	97	5	5.00	0.71	0.76	0.03	0.11	0.40
Satt398 <sup>22</sup>	97	6	5.98	0.59	0.65	0.01	0.19	0.47
Satt406 <sup>26</sup>	97	4	4.00	0.56	0.62	0.04	0.21	0.49
Satt419	97	5	4.97	0.39	0.43	0.02	0.37	0.63
Satt429 <sup>23</sup>	97	7	6.95	0.57	0.65	0.02	0.20	0.48
Satt431	97	5	4.98	0.49	0.57	0.03	0.27	0.53
Satt442 <sup>6</sup>	97	8	7.97	0.73	0.77	0.03	0.09	0.39
Satt454	97	4	4.00	0.41	0.45	0.02	0.35	0.62
Satt463 <sup>4</sup>	97	6	6.00	0.74	0.78	0.01	0.09	0.38
Satt534	97	11	10.73	0.55	0.57	0.08	0.21	0.52
Satt544 <sup>17</sup>	97	6	6.00	0.62	0.67	0.02	0.16	0.46
Satt556 <sup>21</sup>	97	5	4.87	0.59	0.66	0.05	0.19	0.47
Satt581	97	3	3.00	0.34	0.38	0.01	0.43	0.67
Satt583	97	5	4.97	0.30	0.32	0.00	0.48	0.71
Average	96.60	5.98	5.95	0.57	0.63	0.03		
Cumulative over loci							$3.94 \times 10^{-32}$	$5.02 \times 10^{-14}$

<sup>1–27</sup>SSR markers selected for the cultivar identification set, ranked according to their  $H_e$  values





**Fig. 1** Distinction power expressed as number of observed genotype matches and cumulative probability of identity (PI<sub>Sibs</sub>) values as a function of increasing number of SSR loci

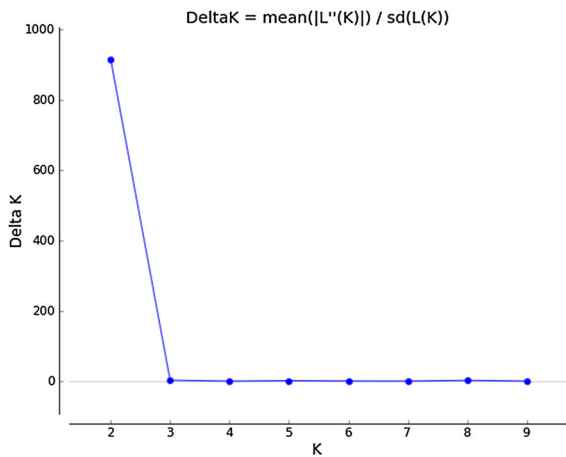


**Fig. 2** UPGMA dendrogram of 97 European soybean genotypes base on 42 microsatellite loci. Colors represent the country of cultivars’ origin. ● Austria ● Argentina ● Canada ● Croatia

● France ● Germany ● Italy ● Romania ● Serbia ● Switzerland ● USA. Next to the accessions name, its maturity group was highlighted. (Color figure online)

only ~ 20% of admixed origin. French soybean germplasm originated either from ancestral population 1 or 2, or it was of admixed origin. Romanian cultivars had mostly ancestry in population 1, while

the rest originated from population 2 or was of admixed origin. Serbian germplasm was also quite diverse with Galina, Proteinka and Novosadanka grouping with ancestral population 2 and Merkur



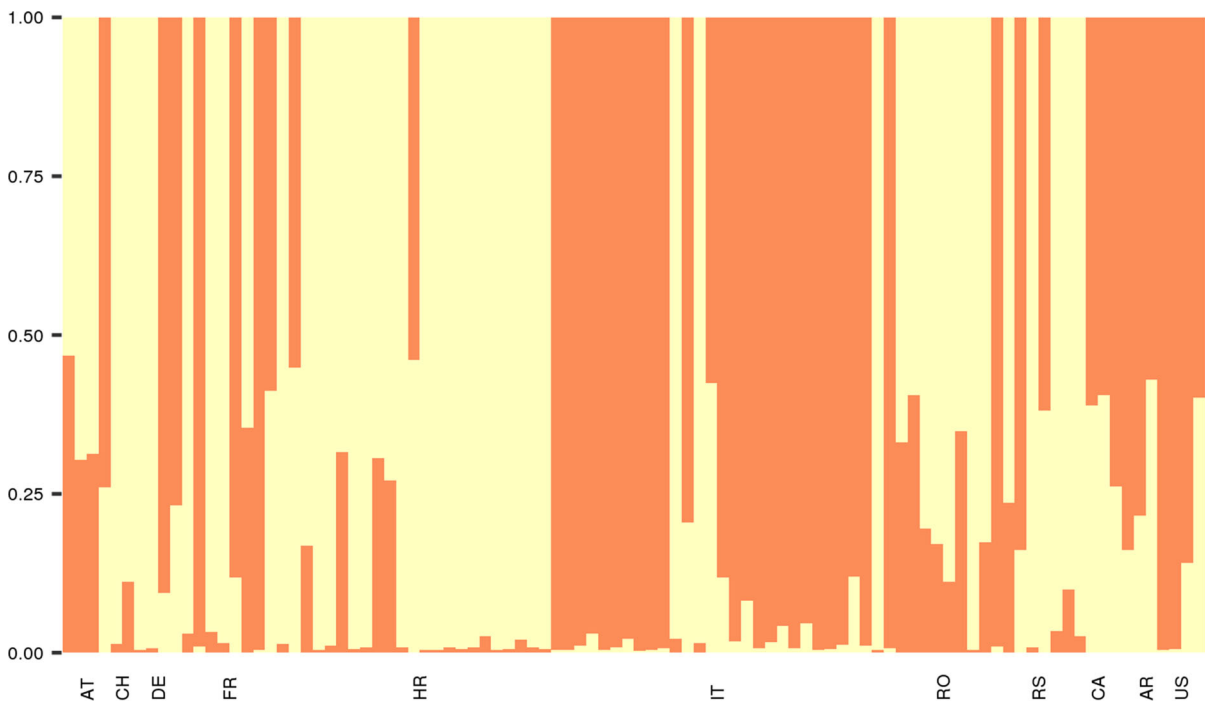
**Fig. 3** Delta K values for  $k = 1$  to 10

and Optimus grouping with ancestral population 1. Swiss germplasm, although represented by only four cultivars, showed its origin mainly from ancestral population 2.

## Discussion

### Genotype identification

In the present study, all 42 used SSR loci were polymorphic and produced unique SSR profiles for 79 cultivars and six breeding lines originating from different European breeding programs as well as for 12 cultivars of non-European origin. Nevertheless, there were some cases when genotypes and probably their parental lines had high proportion of genetic similarity. This applied for Croatian cultivars OS Zora, Ika and Tena, originating from the same breeding program. Named cultivars successfully passed DUS testing and showed sufficient distinctness in morphological traits. These cultivars had high level of genetic similarity (0.952) what implies that they might have been derived from the same cross or that they share at least one common parent. Indeed, according to pedigree (personal communication with breeder) two out of three cultivars in the cluster (Ika and OS Zora) were derived from the same cross, while



**Fig. 4** Graphical representation of the proposed structure of the analyzed set of genotypes for  $k = 2$  according to Structure analysis. Each cultivar is represented by a vertical column painted in accordance with the Q coefficient and the corresponding source group. Genotypes are grouped according to the

countries of their origin (breeding program). Country abbreviations: AR Argentina; AT Austria; CA Canada; HR Croatia; FR France; DE Germany; IT Italy; RO Romania; RS Serbia; CH Switzerland; US USA

one parent of the cultivar Tena was Ika. More precisely, differences among three above mentioned cultivars occur at only three out of 42 SSR loci. Such level of genetic similarity complicates cultivar identification in soybean.

The minimum number of SSR markers needed for clear cultivar distinction in the present study was 27, double more than proposed by Song et al. (1999) where 13 selected markers were enough for distinction of 66 North American soybean cultivars. Although 6 out of these 13 SSR markers overlap with marker set in this study, they were not so informative as in American germplasm.

In many studies cumulative PI or  $PI_{Sibs}$  (probability of identity) were used as an indicator of SSRs' discriminatory power in cultivars' distinction—e.g. wheat (Eujayl et al. 2002), grape (Cipriani et al. 2008), sugarcane (dos Santos et al. 2012), apple (Gross et al. 2012, Gasi et al. 2016) and sunflower (Moreno et al. 2013). Cumulative values of PI or  $PI_{Sibs} < 0.0001$ , originally established in structured natural animal populations by Waits et al. (2001), are usually accepted as sufficient probability levels for estimating the number of loci necessary to distinguish genotypes.

In the present study cumulative  $PI_{Sibs}$  varied between  $8 \times 10^{-5}$  (for 10 SSR markers) to  $1.39 \times 10^{-10}$  (for 27 SSR markers) indicating that already 10 most polymorphic markers should be more than enough to discriminate all 97 genotypes. However, by using only 10 most polymorphic markers we have found 606 genotype matches out of 4656 possible pairwise comparisons in the tested set (Fig. 1). Apparently, widely used statistics ( $PI/PI_{Sibs}$ ), proposed by Waits et al. (2001) for estimating the probability of sampling identical genotypes within populations of intermating individuals, does not work well in defining minimal number of markers for genotypes' discrimination within structured plant populations such as collections of cultivars or gene bank accessions, probably due to absence of free mating. So, the use of cumulative  $PI/PI_{Sibs}$  value as a criterion for selection of minimal number of markers should be used with great prudence.

When taking into account here determined number of SSRs needed for cultivar discrimination as well as their cost per data point, the use of SNP markers should be considered as their alternative. SSRs have more discriminating power due to their multi-allelic nature since one SSR accounts for 1.7–3.2 biallelic

SNPs (Yoon et al. 2007, Cabezas et al. 2011). However, modern high throughput technologies enable screening of large number of SNPs (Song et al. 2015) and therefore detection and selection of the most informative ones. E.g. UPOV's Working Group on Biochemical and Molecular Techniques and DNA-Profiling in Particular singled out 14 highly informative SNPs, that were sufficient to differentiate all 276 cultivars analyzed (Document BMT/17/18 available at [https://www.upov.int/edocs/mdocs/upov/en/bmt\\_17/bmt\\_17\\_18.pdf](https://www.upov.int/edocs/mdocs/upov/en/bmt_17/bmt_17_18.pdf)).

#### Genetic diversity of European commercial soybean germplasm

To make accurate comparisons in level of germplasm genetic diversity estimated by SSRs with other similar studies one should have in mind the genome origin of SSR loci as well as the size of the germplasm sample. The SSR markers employed by the soybean research community for germplasm characterization over the years has been mainly diverse, not standardized. Although the criteria we used for loci selection was based on their high gene diversity ( $H_e$ ) or PIC values, same markers are not expected to be equally polymorphic (informative) when applied to different germplasm. E.g., six SSR markers common to the set recommended by Song et al. (1999) for cultivar identification showed considerably less alleles and had lower  $H_e$  values in our study (Satt009 16 alleles and  $H_e = 0.82$  in Song et al. (1999) versus 8 alleles and  $H_e = 0.63$  in our study). The set analyzed by Song et al. (1999) involved similar number of accessions ( $n = 101$ ), which included American ancestral lines and elite North American germplasm. Considerably less detected alleles and the lower gene diversities of the SSR markers used, together with the double more markers needed to separate European set of soybean cultivars, as compared to North American germplasm, indicates much lower genetic diversity of the European germplasm.

On the other hand, informativeness of five SSR loci common to our and study of Priolli et al. (2002), who analyzed 186 Brazilian elite cultivars, was rather similar. Both studies showed similar mean number of alleles per locus as well as mean gene diversity over all SSR loci  $N_a = 5.98$ ;  $H_e = 0.63$  vs.  $N_a = 5.3$ ;  $H_e = 0.64$ . However, size of data sets differed considerably, 97 vs. 186 analyzed genotypes, respectively,

suggesting the presence of higher genetic diversity within European than Brazilian commercial germplasm.

In terms of mean number of alleles per locus, genetic diversity of 185 European accessions in Tavaud-Pirra et al. (2009) was higher than genetic diversity of 97 European commercial soybean cultivars in this research ( $N_a$  of 9.8 vs.  $N_a$  of 5.98, respectively), which is to be expected having in mind the number of genotypes analyzed in the two studies. However, when considering diversity of 131 accessions from 14 Asian Countries (Abe et al. 2003) and 129 (Wang et al. 2006) and 159 (Li et al. 2011) Chinese accessions, which exhibited 11.9, 12.2 and 14.2 alleles per locus, respectively, as well as 205 Chinese and 39 Japanese accessions (Guan et al. 2010), which exhibited 16.2 alleles per locus, it is clear that only a portion of available genetic diversity is present in commercial European germplasm.

#### Clustering and genetic structure of European commercial soybean germplasm

Cluster analysis revealed 2 major clusters which was in agreement with the results obtained by Structure analysis. As mentioned previously, signals for genetic structure were weak and chosen as best for  $k = 2$ . The method itself assumes that the observed germplasm is a natural population (with equal distribution of accessions), which is of course not true when dealing with commercial cultivars, so weak structuring is expected. Consequently, interpretation of the implied structure has constraints. The genetic structure of the studied soybean germplasm based on the dendrogram expectedly does not interfere strictly with geographic regions due to short history of soybean cultivation in Europe and use of elite germplasm in most breeding programs. However, results suggest that simple structuring might be caused by environmental adaptation of founding germplasm determined by its maturity time. Indeed, it seems that maturity group had very strong impact on clustering, which can be seen e.g. at subcluster containing mix of cultivars (very early MG 000 and 00) originating from Romania, France, Switzerland, Croatia and Austria. Italian soybean breeding program is dominated by ERSA (*Agenzia regionale per lo Sviluppo Rurale del Friuli Venezia Giulia*), whose breeding activities started quite recently in 1987, with its main goal of producing

new cultivars with low content of antinutritional factors (Signor, 2016). Founder cultivars used in this breeding program were provided by the University of Illinois (Urbana, USA) what is confirmed by our results (Fig. 2). Nevertheless, detected separation of the Italian from the rest of the European cultivars, together with their very good agronomic performance, makes this germplasm very interesting for further breeding. Previous introduction of soybean germplasm from South and North America into European breeding programs is also supported (Fig. 2) by its grouping within close sub-clusters (like Dekabig, DM513 and DM503). This introduction from Americas is still in course in modern European soybean breeding programs. Less attention is paid to Asian germplasm. From 1923 to 2005, 1300 soybean cultivars were released by the Chinese breeding programs and their genetic base was traced back to 670 nuclear ancestors (Xiong et al. 2010), unlike genetic base of North American soybean whose 84% was contributed by only 17 founder genotypes (Gizlice et al. 1994). Since China is the center of diversity for soybean it would be advisable to utilize Chinese rich soybean genetic resources in pre-breeding to broaden the genetic base of European cultivars. More than 23,000 cultivated and 7000 wild accessions are conserved in Chinese National Gene bank (CNGB) and its subsequent core collections are a valuable source of favorable traits (Qiu et al. 2013).

#### Conclusions

SSR markers proved, once again, to be a valuable tool in germplasm characterization and genetic diversity assessment. Genetic basis of European soybean germplasm is rather narrow compared to Asian and American germplasm thus, requires a substantial number of SSR markers for reliable cultivar distinction. Here defined set of 27 SSR might serve as a starting tool for soybean cultivar identification in Europe. However, fast progress and lower costs of genotyping technologies might single out new marker types (such as SNPs) as marker of choice for similar future studies. It seems that the structure of current European soybean germplasm was mainly affected by environmental adaptation as indicated by maturity groups. Determined genetic relationships among individual or groups of genotypes might serve as

prediction tool for future breeding decisions. Disclosed SSR data of the analyzed commercial European germplasm can serve as a foundation of public soybean database.

**Acknowledgements** This research was funded by the Environmental Protection and Energy Efficiency Fund with the support of the Croatian Science Foundation of the Republic of Croatia and partly supported by the project KK.01.1.1.01.0005, Centre of Excellence for Biodiversity and Molecular Plant Breeding (CoE CroP-BioDiv), Zagreb, Croatia. Special thanks to Dr. Marco Signor, Regional Agriculture Agency of Friuli Venezia Giulia (ERSA), Udine, Italy, who provided for this research number of genotypes.

#### Compliance with ethical standards

**Conflict of interest** The authors declare no conflict of interest.

#### References

- Abe J, Xu D, Suzuki Y, Kanazawa A, Shimamoto Y (2003) Soybean germplasm pools in Asia revealed by nuclear SSRs. *Theor Appl Genet* 106:445–453
- Bandillo N, Jarquin D, Song Q, Nelson R, Cregan P, Specht J, Lorenz A (2015) A population structure and genome-wide association analysis on the USDA soybean germplasm collection. *The Plant Genome* 8:1–13
- Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314–331
- Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368:455–457
- Cabezas JA, Ibáñez J, Lijavetzky D, Vélez D, Bravo G, Rodríguez V, Carreño I, Jermakow AM, Carreño J, Ruiz-García L, Thomas MR, Martínez-Zapater JM (2011) A 48 SNP set for grapevine cultivar identification. *BMC Plant Biol* 11:153
- Carter TE Jr, Nelson RL, Sneller CH, Cai Z (2004) Genetic diversity in soybean. In: Boerma HR, Specht JE (eds) Soybeans: improvement, production and uses. *Am Soc Agron, Madison*, pp 303–406
- Cipriani G, Marrazzo MT, Di Gaspero G, Pfeiffer A, Morgante M, Testolin R (2008) A set of microsatellite markers with long core repeat optimized for grape (*Vitis* spp) genotyping. *BMC Plant Biol* 8:127
- Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, van Toai TT, Lohnes DG, Chung J, Specht JE (1999) An integrated genetic linkage map of the soybean genome. *Crop Sci* 39:1464–1490
- dos Santos JM, Duarte Filho LSC, Soriano ML, da Silva PP, Nascimento VX, de Souza Barbosa GV, Todaroc AR, Netoc CER, Almeida C (2012) Genetic diversity of the main progenitors of sugarcane from the RIDESA germplasm bank using SSR markers. *Ind Crop Prod* 40:145–150
- Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4:359–361
- El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree (*Argania spinosa* (L.) Skeels) endemic to Morocco. *Theor Appl Genet* 92:832–839
- Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W (2002) Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. *Theor Appl Genet* 104:399–407
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
- FAOSTAT (2019) Statistics division of the Food and Agriculture Organization (FAO) of the United Nations, Rome. [www.fao.org/faostat/en/#data/QC](http://www.fao.org/faostat/en/#data/QC). Accessed 2 June 2019
- Gasi F, Kanlić K, Stroil BK, Pojskić N, Asdal Å, Rasmussen M, Kaiser M, Meland M (2016) Redundancies and genetic structure among ex situ apple collections in Norway examined with microsatellite markers. *HortScience* 51:1458–1462
- Gizlice Z, Carter TE Jr, Burton JW (1994) Genetic base for North American public soybean cultivars released between 1947 and 1988. *Crop Sci* 34:1143–1151
- Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3.). Department of ecology and evolution, Lausanne University, Switzerland. <http://www2.unil.ch/popgen/softwares/fstat.htm>. Accessed 7 Nov 2011
- Grainger CM, Letarte J, Rajcan I (2018) Using soybean pedigrees to identify genomic selection signatures associated with long-term breeding for cultivar improvement. *Can J Plant Sci* 98:1176–1187
- Gross BL, Volk GM, Richards CM, Forsline PL, Fazio G, Chao CT (2012) Identification of “duplicate” accessions within the USDA-ARS national plant germplasm system *Malus* collection. *J Am Soc Hortic Sci* 137:333–342
- Guan R, Chang R, Li Y, Wang L, Liu Z, Qiu L (2010) Genetic diversity comparison between Chinese and Japanese soybeans (*Glycine max* (L.) Merr.) revealed by nuclear SSRs. *Genet Resour Crop Evol* 57:229–242
- Hahn V, Würschum T (2014) Molecular genetic characterization of Central European soybean breeding germplasm. *Plant Breed* 133:748–755
- Hymowitz T (2008) The history of the soybean. In: Johnson LA, White PJ, Galloway R (eds) Soybeans chemistry, production, processing, and utilization. AOCS Press, Urbana, pp 1–31
- ISAAA (2017) Pocket K No. 16: Global Status of Commercialized Biotech/GM Crops in 2017. International Service for the Acquisition of Agri-biotech Applications. [www.isaaa.org](http://www.isaaa.org). Retrieved 8 July 2019
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806

- Jamali SH, Cockram J, Hickey LT (2019) Insights into deployment of DNA markers in plant cultivar protection and registration. *Theor Appl Genet* 132:1911–1929
- Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874
- Kurasch AK, Hahn V, Leiser WL, Vollmann J, Schori A, Béatrix CA, Mayr B, Winkler J, Klemens 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, Pospisil 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 Environ* 40:765–778
- Kuroda Y, Tomooka N, Kaga A, Wanigadeva SMSW, Vaughan DA (2009) Genetic diversity of wild soybean (*Glycine soja* Sieb. et Zucc.) and Japanese cultivated soybeans [*G. max* (L.) Merr.] based on microsatellite (SSR) analysis and the selection of a core collection. *Genet Resour Crop Evol* 56:1045–1055
- Li YH, Smulders MJ, Chang RZ, Qiu LJ (2011) Genetic diversity and association mapping in a collection of selected Chinese soybean accessions based on SSR marker analysis. *Conserv Genet* 12:1145–1157
- Moreno MV, Nishinakamasu V, Loray MA, Alvarez D, Gioco J, Vicario A, Hopp HE, Heinz RA, Paniego N, Lia VV (2013) Genetic characterization of sunflower breeding resources from Argentina: assessing diversity in key open-pollinated and composite populations. *Plant Genet Resources* 11:238–249
- Oliveira MB, Vieira ESN, Schuster I (2010) Construction of a molecular database for soybean cultivar identification in Brazil. *Genet Mol Res* 9:705–720
- Park SDE (2008) Excel Microsatellite Toolkit. Computer program and documentation distributed by the author. Website <http://animalgenomics.ucd.ie/sdepark/ms-toolkit/>. Accessed 18 July 2011
- 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
- Priolli RHG, Mendes-Junior CT, Arantes NE, Contel EPB (2002) Characterization of Brazilian soybean cultivars using microsatellite markers. *Genet Mol Biol* 25:185–193
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Qiu LJ, Xing LL, Guo Y, Wang J, Jackson SA, Chang RZ (2013) A platform for soybean molecular breeding: the utilization of core collections for food security. *Plant Mol Biol* 83:41–50
- Ramasamy RK, Ramasamy S, Bindroo BB, Naik VG (2014) STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. *SpringerPlus* 3:431
- Ristova D, Šarčević H, Šimon S, Mihajlov L, Pejić I (2010) Genetic diversity in southeast European soybean germplasm revealed by SSR markers. *Agriculturae Conspectus Scientificus* 75:21–26
- Röder MS, Wendehake K, Korzun V, Bredemeijer G, Laborie D, Bertrand L, Isaac P, Rendell S, Jackson J, Cooke RJ, Vosman B, Ganal M (2002) Construction and analysis of a microsatellite-based database of European wheat cultivars. *Theor Appl Genet* 106:67–73
- Signor M (2016) Experiences of soybean cropping and breeding in Northern Italy. Presentation. Meeting Sustainable Soya—Sustainable Europe- Budapest, 24–25 Nov 2016
- Sohn HB, Kim SJ, Hwang TY, Park HM, Lee YY, Markkandan K, Lee D, Lee S, Hong SY, Song YH, Koo BC, Kim YH (2017) Barcode system for genetic identification of soybean [*Glycine max* (L.) Merrill] cultivars using InDel markers specific to dense variation blocks. *Front Plant Sci* 8:520
- Song QJ, Quigley CV, Nelson RL, Carter TE, Boerma HR, Strachan JL, Cregan PB (1999) A selected set of trinucleotide simple sequence repeat markers for soybean cultivar identification. *Plant Variet Seeds* 12:207–220
- Song Q, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. *Theor Appl Genet* 109:122–128
- Song Q, Hyten DL, Jia G, Quigley CV, Fickus EW, Nelson RL, Cregan PB (2015) Fingerprinting soybean germplasm and its utility in genomic research. *Genes Genom Genet* 5:1999–2006
- SoyBase (2018) Soybean breeder's toolbox. <http://soybase.org>. Accessed May 2018
- Tavaud-Pirra M, Sartre P, Nelson R, Santoni S, Texier N, Roumet P (2009) Genetic diversity in a soybean collection. *Crop Sci* 49:895–902
- This P, Jung A, Boccacci P, Borrego J, Botta R, Costantini L, Crespan M, Dangel GS, Eisenheld C, Ferreira-Monteiro F, Grando S, Ibáñez J, Lacombe T, Laucou V, Magalhães R, Meredith CP, Milani N, Peterlunger E, Regner F, Zulini L, Maul E (2004) Development of a standard set of microsatellite reference alleles for identification of grape cultivars. *Theor Appl Genet* 109:1448–1458
- USDA - U.S. Department of Agriculture (2019) USDA Agricultural Projections to 2028. U.S. Department of Agriculture Long-term Projections Report OCE-2019-1. Accessed September 2019
- Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Mol Ecol* 10:249–256
- Wang L, Guan R, Zhangxiong L, Chang R, Qiu L (2006) Genetic diversity of Chinese cultivated soybean revealed by SSR markers. *Crop Sci* 46:1032–1038
- Widaningsih NA, Purwanto E, Nandariyah N, Reflinur R (2014) The use of DNA microsatellite markers for genetic diversity identification of soybean (*Glycine max* (L.) Meril.) as a supplementary method in reference collections management. *Indones J Biotechnol* 19:136–145
- Wysmierski PT, Vello NA (2013) The genetic base of Brazilian soybean cultivars: evolution over time and breeding implications. *Genet Mol Biol* 36:547–555
- Xiong D, Zhao T, Gai J (2010) Genetic bases of improved soybean cultivars released from 1923 to 2005 in China—a historical review. *Front Agric China* 4:383–393
- Yoon MS, Song QJ, Choi IY, Specht JE, Hyten DL, Cregan PB (2007) BARCSoySNP23: a panel of 23 selected SNPs for soybean cultivar identification. *Theor Appl Genet* 114:885–899

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