Euphytica (2018)214:108
https://doi.org/10.1007/s10681-018-2197-0

Comparison of methods for the estimation of best parent heterosis among lines developed from interspecific sunflower germplasm

Nada Hladni · Miroslav Zorić · Sreten Terzić · Nataša Ćurčić · Zlatko Satovic · Dragan Perović · Dejana Panković

Received: 8 December 2017 / Accepted: 7 June 2018

Abstract Pre-breeding and elite breeding are two steps in creating high yielding sunflower hybrids that differ in well established procedures and selection methods. However, a methodology that bridge efficient use of introgression lines as product of pre-breeding procedures and their crossing to elite inbreed lines, is not yet very well established. Therefore, the development of cost- and time-efficient methods for the determination of best parent heterosis and the use of best inbred lines in crosses with introgression lines for obtaining high-yielding and stable hybrids is highly desirable. In this regard, sixteen Cytoplasmic Male Sterile (CMS) inbred lines (A) derived from four heterogeneous interspecific lines originating from three annual: *H. debilis silvestris* (DEB-SIL), *H. praecox runyoni* (PRA-RUN), *H. deserticola* (DES) and one perennial *H. resinosus* (RES) wild species were evaluated. Seven agronomic traits were measured over a period of 2 years and 38 DNA loci were analysed, in order to compare four different methods for the estimation of best parent heterosis (BPH). New inbred lines were characterized by Principal Component Analysis (PCA) of morphological traits and Principal Coordinate Analysis (PCoA) of molecular marker data. Line × tester mating design was used to evaluate General Combining Ability (GCA), while Genetic Distance (GC) estimated by markers was evaluated as a predictor of BPH by Locally Weighted Sequential Smoothing (LOESS). Analysis of combining ability is one of the most important tools breeders

N. Hladni · M. Zorić · S. Terzić
Oil Crops Department, Institute of Field and Vegetable Crops, Maksima Gorkog 30, Novi Sad 21000, Serbia

N. Ćurčić
Institute of Food Technologies, Bulevar Cara Lazara 1, Novi Sad 21000, Serbia

Z. Satovic
Department for Seed Science and Technology, Faculty of Agronomy, University of Zagreb, Svetosimunska cesta 25, 10000 Zagreb, Croatia

Z. Satovic
Centre of Excellence for Biodiversity and Molecular Plant Breeding (CoE CroP-BioDiv), Svetosimunska cesta 25, 10009 Zagreb, Croatia

D. Perović
Julius Kuehn-Institute (JKI), Federal Research Centre for Cultivated Plants, Institute for Resistance Research and Stress Tolerance, Erwin-Baur-Strasse 27, 06484 Quedlinburg, Germany

D. Perović · D. Panković (✉)
Faculty of Ecological Agriculture, EDUCONS University, Vojvode Putnika 87, Sremska Kamenica 21208, Serbia
e-mail: dejana.pankovic@educons.edu.rs

Published online: 16 June 2018

use to identify superior inbred lines on the basis of their performance in hybrid combinations. Results obtained in this research show that PCA of morphological and PCoA of molecular marker data on parental lines are generally in agreement with GCA effects for examined traits. GD versus BPH relationships indicate that intermediate to high GD between parental lines was optimal for best heterotic effects of most traits. In this study, we show that the combination of the PCA of morphological data, PCoA of molecular marker data and GD between parental lines is fast and affordable, giving the most important information for parental choice of introgression and elite lines in sunflower breeding programs.

Keywords *Helianthus annuus* L. · Interspecific hybridization · Pre-breeding · Wild species · SSRs · Genetic diversity · Genetic distance · GCA · PCA · LOESS

Introduction

Sunflower has been transformed into a hybrid oilseed crop through the complex evolutionary history, beginning with domestication in North America more than 4000 years ago (Blackman et al. 2011; Smith 2014). Breeding for oilseed traits, self-compatibility, self-pollination and hybrid seed-production resulted in significantly reduced genetic polymorphism in comparison to the wild germplasm. In extensive analysis of genetic population of 433 cultivated accessions from North America and Europe, including a wide-range collection of 24 wild sunflower populations, Mandel et al. (2011) have shown that the cultivated sunflower gene pool nowadays harbors approximately two-thirds (67%) of the total genetic diversity present in wild sunflower. Genetic diversity, assessed by SSR markers, progressively decreased from 0.82 for wild populations, to 0.64 and 0.51 for exotic lines and elite inbreds respectively (Tang and Knapp 2003). Even greater loss of genetic diversity was found on linkage groups that were target of selection. The percentage of retained genetic diversity of LG6, that contains major oil-related QTL cluster, was about 70% for exotic lines versus wild populations, however it decreased to less than 30% for elite versus wild and elite versus exotic lines (Burke et al. 2005).

The phenomenon of heterosis has been widely used in agronomy despite scarce knowledge of the molecular basis and underlying genes. The superiority of F_1 hybrids over their inbred parents is manifested as increased yield, biomass, growth rate, and/or resistance to biotic and abiotic stresses. Recent research on sunflower heterosis at proteome level indicated that improved carbon fixation, chlorophyll stability and efficiency and less photorespiration could contribute to hybrid vigor (Mohayeji et al. 2014). The heterotic potential of different crosses is very variable. General combining ability (GCA) test (Sprague and Tatum 1942) is still the major method for evaluation of inbreed lines for their potential in hybrids (Yao et al. 2013; Reif et al. 2013). As this approach is costly and time consuming there are many efforts on defining the conditions for a priori choice of hybrid parents, based on parental distance by means of agronomic and molecular data. However, the resulting correlation between genetic distance between parents and heterosis was positive, but also negative and sometimes inconclusive, depending on the plant material and measured agronomic traits (Dias Dos Santos et al. 2004; Reif et al. 2013; Krishnamurthy et al. 2013). GDs were not so often used for predicting sunflower hybrid performance (Tersac et al. 1994; Cheres et al. 2000), but rather to describe patterns of genetic diversity between lines (Berry et al. 1994; Hongtrakul et al. 1997). Though GD was correlated with seed yield it was a poor predictor of hybrid performance in sunflower (Cheres et al. 2000). However, the use of GDs for predicting sunflower hybrid performance was mostly examined on adapted (Cheres et al. 2000) and elite sunflower lines (Reif et al. 2013) with a narrow genetic base, which could be the cause of the lack of GD-heterosis association (Dias Dos Santos et al. 2004.). Sunflower inbred lines carrying introgression fragments, due to higher genetic diversity and a lower recombination rate in comparison to elite inbreed lines, are very suitable genetic material to examine the dependence of heterosis on GDs and to predict crosses with better heterotic potential.

Sunflower genome is complex, with different genome structures of sections in Helianthus genus which is causing difficulties in chromosome pairing in interspecific crosses (Atlagić and Terzić 2014). Nevertheless interspecific crosses are often employed in pre-breeding procedures for the purpose of improvement of cultivated sunflower (Hladni and Miklič

2012). Seiler et al. (2017) have identified wild species as sources of pest and disease resistance, abiotic factors resistance, CMS and Restorer fertility (Rf) genes and herbicide resistance. The use of wild species to improve sunflower breeding was quite successful, however the major challenge is to keep targeted alleles from the donor wild species with minimum linkage drag in the offspring of interspecies crosses (Warburton et al. 2017). Progeny of interspecies crosses is the pre-breeding gene pool, a source for the production of new inbred lines that have to be screened for their utility for future breeding programs. Production of new inbred lines from interspecies crosses is itself cumbersome and time consuming and imposes the need for the development of cost- and time-efficient methods for the determination of best parent heterosis and the use of best inbred lines in crosses with introgression lines for obtaining high-yielding and stable hybrids.

One of the preferred tools for analysis of divergence in plant material is multivariate statistics technique Principal Component Analysis (PCA). PCA has been used to create relationships between cultivar traits and provide grouping of cultivars in a 2D or 3D space, based on morphological (Syafii et al. 2015), physiological (Ambachew et al. 2015), biochemical (Abid et al. 2016), and molecular marker data (Zimisuhara et al. 2015; Teich et al. 2014; Wu et al. 2014).

The development and application of molecular techniques and genomics have dramatically improved characterization and deployment of plant genetic resources (Van et al. 2011). Recent results on genomic prediction of *Sclerotinia* resistance (Livaja et al. 2016) and sunflower hybrid oil content (Bonnafous et al. 2016; Mangin et al. 2017) are promising. However, as methods of genomic selection are expensive for small breeding programmes, SSR markers are still very popular among breeders and geneticists due to high informative content, co-dominant nature, abundance and reproducibility (Moreno et al. 2013; Usatov et al. 2014; Fillipi et al. 2015).

The aim of this study was to define cost- and time-efficient method(s) for the determination of best parent heterosis and the use of best inbred lines in crosses with introgression lines for obtaining high-yielding and stable hybrids. For that purpose new divergent sunflower inbred lines were developed from interspecific crosses with three annual (*H. debilis silvestris*, *H. praecox runionii* and *H. deserticola*) and one perennial wild species (*H. resinosus*) and used as a model to examine the potential of GD, estimated by SSR markers, as a predictor for BPH. In order to compare methods for evaluation of breeding material results of BPH prediction based on morphological traits and molecular marker data of new lines was also analyzed by PCA and PCoA. Finally, the potential of evaluated methods is compared to the traditional GCA test of lines and their hybrids using morphological traits.

Materials and methods

Plant materials and field experiments

Interspecific sunflower (*Helianthus* ssp.) germplasm lines were kindly provided by Dr. Gerald Seiler (USDA-ARS, Fargo ND, USA) who developed them by crossing wild species with cultivated sunflower inbred lines. Line RES-834-1 (PI 539897) was developed by Seiler (1991a) using wild perennial species *H. resinosus* (cms HA 89*2/*H. resinosus* (Acc 834) F5 (50 pl.) bulk). Line PRA-RUN-1329-1 (Pl539883) was developed by Seiler (1991b) using wild annual species *H. praecox runyoni* (nms P21*3/*H. praecox* ssp. *runyoni* (ACC 1329) F3 sibbed one generation, selfed one generation). Line DEB-SIL-367-2 (PI 539908) was developed by Seiler (1991c) using wild annual species *H. debilis silvestris* (cms HA 89*3/*H. debilis* ssp. *silvestris* (Acc 367) F3 (32 pl.) bulk). Line DES-1474-2 (PI 539913) was developed by Seiler (1991b) using wild annual species *H. deserticola* (cms 89*2/*H. deserticola* (ACC 1474)//RHA 274 F4 sibbed two generations).

The received interspecific lines were found to be heterogeneous in height and branching type. The heterogeneity allowed developing a larger number of inbred lines by selection and selfing. In the first year individual non branched plants were chosen from the interspecific lines and the following year the pedigree method was used with the head to row sowing. Pedigree method includes choosing individual plants from the source material and following the pedigree of the chosen plants until homozygous lines are obtained. For 3 years, plants were selected based on the phenotypic estimation on the plot and the laboratory's estimates [seed yield per plant (SYP), seed oil content (SOC), total leaf area per plant (TLA), plant height

(PH), head diameter (HD), total seed number per head (TSN) and 1000 seeds weight (TSW)].

After S3 generation, male sterility was introduced to selected B lines by hybridization with commercial CMS inbred line L98 (single source of cytoplasmic male sterility, *Helianthus petiolaris* (CMS) PET1 gene; Leclercq 1969) from the breeding program of the Institute of Field and Vegetable Crops (IFVCNS), Novi Sad, Serbia. All the plants were sterile in the F1 generation and the back crosses with B lines were continued up to BC8. On the other hand, B lines were self-pollinated at each generation until they showed complete fixation. After that, an A line is obtained which contains all the properties of the fertile line plus the property of the CMS.

From this program 16 new CMS lines originating from interspecific lines were included in this study (1RES, 2RES, 5RES-*H. resinosus*; 8DEB-SIL, 9DEB-SIL 10DEB-SIL-*H. debilis silvestris*; 11PRA-RUN, 12PRA-RUN, 13PRA-RUN, 14PRA-RUN-*H. praecox runyoni*;16DES, 17DES D, 18DES D, 19DES D, 20DES-*H. deserticola*). Three Rf restorer lines: RHA-T-1, RHA-T-2, RHA-T-3, with good GCA which were utilized as testers, were selected from the sunflower breeding program of the IFVCNS.

A 2-year trial was set up at the Rimski šančevi experimental field at the IFVCNS (45°19′51″N; 19°50′59″E; altitude-84 m), with soil type calcareous chernozem on the loess terrace, using the line × tester method with three replicates. The trial included in total 67 genotypes (16 inbred lines, 3 restorer lines and 48 F1 hybrids) sown in 4.2 m rows, spaced 0.7 m apart. The analysis of examined traits was performed in three replications, with 10 plants per replication, which were sampled from the middle row of each block. Following traits were assessed: plant height (PH), head diameter (HD), total leaf area per plant (TLA), Seed yield per plant (SYP), total seed number per head (TSN), 1000 seeds weight (TSW) and seed oil content (SOC). TLA per plant (cm^2) was measured using a leaf area meter (LI-300- LiCOR, USA) at the flowering stage. PH and HD were measured (cm) in the field at the stage of physiological maturity. Yield components were determined after manual harvest. SYP was normalized to 11% of seed moisture content. The total seed number per head (TSN) was counted. TSW g was determined in a random sample of cleaned and air-dried seed. The analysis of SOC was carried out non-destructively on pure air-dried seed samples, in a nuclear magnetic resonance (NMR) analyzer.

Molecular analysis

Genomic DNA was extracted from leaf samples of plantlets, frozen in liquid nitrogen according to Panković et al. (2007). PCR amplifications were carried out with 40 ng of genomic DNA in the presence of 0.2 mM of each dNTP, 1 U of Taq DNA polymerase (Fermentas), 1XTaq polymerase buffer, 0.3 µM of each primer, 3 mM $MgCl_2$, 10 µg/µl BSA (Panković et al. 2007). Touch-down PCR was performed under the following conditions: 3 min at 95 °C, 7 cycles for 30 s at 94 °C, 30 s at 64 °C (primer annealing) with temperature increment − 1 °C, 30 s at 67 °C (primer extension), 33 cycles for 30 s at 94 °C, 30 s at 58 °C (primer annealing), 30 s at 67 °C (primer extension), and 20 min of final extension at 67 °C (Tang et al. 2002). PCR products were separated using 2% TBE high-resolution agarose gel electrophoresis.

In total 38 primers, 37 SSRs and one STS, were used for molecular marker analysis of DNA polymorphism (Table 3). SSR primers were selected as either highly polymorphic between wild and cultivated sunflowers (Tang and Knapp 2003), or the closest to previously estimated EST/QTL associations (Lai et al. 2005) and distributed on all linkage groups (Tang et al. 2002, 2003).The position of STS on composite linkage maps of three mapping populations developed in IFVCNS, revealed its proximity to Cleaved Amplified Polymorphic Sequences (CAPS) marker for Pl_6 resistance gene to downy mildew (Saftić-Panković et al. 2008).

Data analysis

Agronomic traits were analyzed using ANOVA model, PCA and GCA effects. Molecular marker data were analyzed by PCoA. Both sets of data, agronomic and molecular markers, were used in combined analysis of relationship between genetic distances between parental lines and heterosis for each of the examined yield component.

Agronomic traits

According to the field experimental design, agronomic traits were measured as explained in plant materials, and analyzed using the linear model for the randomized complete block design with fixed effects (Christensen 2015). The analysis was performed for each season independently. Subsequently, the mean values of the genotypes for each agronomic trait were used to test the differences among the groups using one-way fixed effect ANOVA model. Tukey's honest significant differences (HSD) test was used to test null hypothesis $\mu_i = \mu_j$ for every $i \neq j$ (Christensen 2015). The GCA effects were calculated according to the line × tester method (Singh and Chaudhury 2001) for traits which showed statistically significant differences among crosses.

The PCA (Yan and Rajčan 2002) as a multivariate visualization tool was used to study relationships among sunflower inbred lines and traits for each particular year. A symmetric preserving scaling method was used prior to construction of the biplots (Yan 2002). The interpretation of the biplot is based on the 'inner-product' principle (Kroonenberg 1995). A positive correlation between two sunflower traits is represented by an acute angle between them and an obtuse angle represents a negative correlation, according to the following simple rules, e.g.: cosine 0° = 1 is a maximum positive correlation; cosine 180° = − 1 is a maximum negative correlation and; cosine 90° = 0 is an absence of correlation.

Molecular markers

The polymorphic information content (PIC) was estimated in accordance with formulae published by Powell et al. (1996). The proportion of shared alleles distance (D_{PSA}) between pairs of sunflower genotypes was estimated using PowerMarker v3.23 software (Liu 2002). The D_{PSA} distance matrix was suprimosed into two dimensions using the Principal Coordinates Analysis (PCoA) method.

If not stated differently all data analyses were accomplished within R computing environment (R Core Team 2016).

GD versus heterosis

Robust locally weighted sequential smoothing (LOESS) of the curves was used to examine relations between genetic distances of parental lines and best parent heterosis (Cleveland 1979). LOESS is non-parametric simple method used to fit the smooth curves to empirical data. The relationship between heterosis for each examined agronomic and genetic distance between parental lines was examined as individual model. The goodness of fit of the individual models was measured by estimating the Pearson's linear correlation coefficient between the dependent variable and fitted values of the individual models. The relationship between pairs of variables is not expressed by equation, whereas the value of the correlation coefficients reported as r^2 indicates the quality of the non-parametric regression models.

Results

Mean values and analysis of variance

Seven traits of new CMS inbred lines and Rf lines were measured on one location in three block-based replications on 30 plants during 2 years. Mean values for seed yield per plant (SYP), seed oil content (SOC), total leaf area per plant (TLA), plant height (PH), head diameter (HD), total seeds number per head (TSN) and 1000 seeds weight (TSW) were calculated per interspecific line (Table 1). SYP of CMS inbred lines developed from RES, DEB-SIL and PRA-RUN interspecific line were significantly higher than mean values from DES CMS inbred lines and Rf lines, in both years. However, average SOC was the highest in Rf lines, over 50% in both experimental years. Mean SOC of new CMS lines was the lowest for PRA-RUN interspecific line (36.0% year I; 34.6% year II) and the highest for RES interspecific line of origin (48.2% year I; 45.7% year II). In tester lines mean SOC was 50.8% for year I and 50.2% for year II. Average PH from DEB-SIL CMS inbred lines was consistently higher in both years in comparison to other lines. HD of Rf lines was the lowest in both years of the experiment. New CMS inbred lines originating from DEB-SIL and RES interspecific line had the highest mean values for TSN. The highest mean values for TSW was observed in PRA-RUN CMS inbred lines

Table 1 Mean values for seven agronomic traits (AT), measured on one location and two seasons (years I and II) of new cms inbred lines calculated per interspecific line (RES—*H. resinosus*; DEB-SIL—*H. debilis silvestris*; PRA-RUN—*H. praecox runyoni*; DES—*H. deserticola*) and restorer lines (Rf). The significance of differences was tested across the year by Tukey's test ($p < 0.05$)

New CMS inbred lines					Rf lines
AT	RES	DEB-SIL	PRA-RUN	DES	
Seed yield per plant (g)					
I	50.5a	56.1a	58.1a	35.1b	28.9b
II	69.1a	78.6a	70.8a	39.8b	32.5b
Seed oil content (g/100 g)					
I	48.2ab	44.2b	36.0d	40.7c	50.8a
II	45.7b	42.0bc	34.6d	38.4c	50.2a
Total leaf area per plant (cm^2)					
I	5475b	5982ab	4339c	4556c	6445a
II	7114a	7181a	6199b	5949b	7529a
Plant height (cm)					
I	97.9b	113.1a	105.0ab	81.1c	115.5a
II	100.0b	114.1a	106.2ab	87.8c	104.2ab
Head diameter (cm)					
I	20.8ab	19.6b	21.0a	21.4a	15.4c
II	22.6a	22.2a	23.0a	22.7a	17.5b
Total seed number per head					
I	936a	1001a	731b	787b	760b
II	984a	1006a	743b	772b	836b
1000 seed weight (g)					
I	51.4bc	56.2b	85.4a	47.7c	35.7d
II	69.7b	73.2b	95.1a	51.0c	36.3d

and the lowest average TSW was found in tester lines for both years.

New CMS lines were characterized by: the highest SOC in RES group; highest PH and TSN in DEB-SIL group, and the highest TSW in PRA-RUN group. Analysis of variance revealed that the effect of genotype was significant ($p < 0.05$) for all examined traits in each year of the experiment (Table 2). The effect of genotype was split to effects of parents, parents versus crosses, crosses, line L, tester T and L × T. Only in case when tester lines were examined as a source of variance, the effects were not significant for TLA and HD. For all other traits, the effects were significant ($p < 0.05$).

PCA analysis

PCA was used to compare new CMS lines on the basis of examined traits, and eventually identify lines that are desirable in terms of multiple traits. First two principal components covered 74.3 and 75.2% of total variability for year I and year II, respectively (Figs. 1, 2). Generally all lines formed groups according to the interspecific line they originated from, for both examined years. New lines from the initial cross with interspecific line PRA-RUN and restorer lines formed two most distinguished groups. In other cases there was some overlapping between groups of origins. Lines from the RES group overlapped with DEB-SIL and DES group. CMS lines from PRA-RUN group had the highest TSW and HD in both years, in particular line 13A PRA-RUN with high and stable TSW. In both years, two lines from DEB-SIL interspecies background 8A and 9A exhibited the highest values of two traits: PH and TSN. Restorer lines, especially RHA-T-2, had high and stable seed oil content. PCA indicates that there was a very strong positive correlation between PH and TSN in both years. Strong positive correlation was observed between PH and SYP; SYP and TSW. TLA was in strong correlation with three traits: SOC, TSN and PH. Negative correlation was observed between SOC and three traits: HD, TSW and SYP. The correlations among traits, calculated directly, partly support the PCA data. High positive correlation coefficients of 0.67 and 0.58 ($p < 0.01$) among TLA and SOC were observed in the first and second year respectively. Strong positive correlation (0.7; $p < 0.01$) among SYP and TSW was observed for both years. Positive correlation was also observed between TSW and HD (0.5; $p < 0.05$ and 0.65; $p < 0.01$). Negative correlation coefficients of -0.61 and -0.71 ($p < 0.01$) between HD and SOC and TSW and SOC (-0.74 and -0.64; $p < 0.01$) were observed for the first and second year respectively.

Genetic variability and PCoA based on SSR markers

Among 38 markers used in this investigation, 37 were SSRs and one was a STS marker linked to downy mildew resistance gene HAP2, on LG8 (Table 3). Fifteen linkage groups were covered with at least one marker. Almost half of the selected markers were

Table 2 Mean squares from ANOVA for agronomic traits of 16 interspecific sunflower lines and 3 testers. Seven agronomic traits were: seed yield per plant (SYP), seed oil content (SOC), total leaf area (TLA), plant height (PH), head diameter (HD), total seed number per head (TSN) and 1000 seeds weight (TSW)

SOV	Df	MS						
		SYP	SOC	TLA	PH	HD	TSN	TSW
Year I								
Replicates	2	ns	ns	ns	ns	ns	ns	ns
Genotypes	66	**	**	**	**	**	**	**
Parents	18	**	**	**	**	**	**	**
Parents versus crosses	1	**	**	**	**	**	**	**
Crosses	47	**	**	**	**	**	**	**
Line L	15	**	**	**	**	**	**	**
Tester T	2	**	**	ns	**	ns	**	**
L × T	30	**	**	**	**	**	**	**
Error	132	ns	ns	ns	ns	ns	ns	ns
Year II								
Replicates	2	ns	ns	ns	ns	ns	ns	ns
Genotypes	66	**	**	**	**	**	**	**
Parents	18	**	**	**	**	**	**	**
Parents versus crosses	1	**	**	**	**	**	**	**
Crosses	47	**	**	**	ns	**	**	**
Line L	15	**	**	**	**	**	**	**
Tester T	2	**	*	ns	**	ns	**	**
L × T	30	**	**	**	ns	**	**	**
Error	132	ns	ns	ns	ns	ns	ns	ns

ns no significant

*, **significant at 0.05 and 0.01 probability levels, respectively

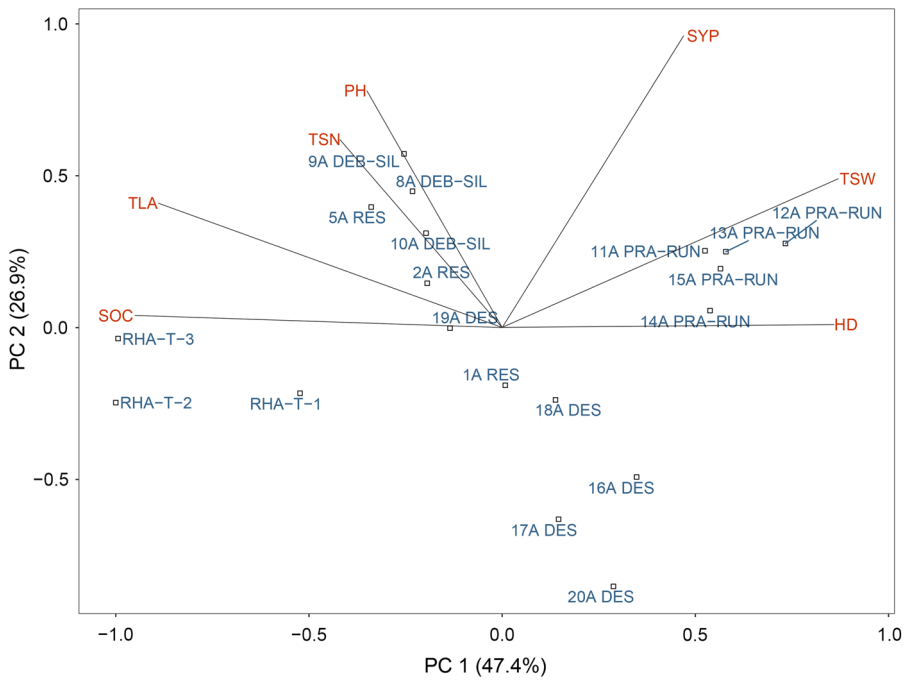

Fig. 1 PCA based on genotype by trait standardized data in the first year

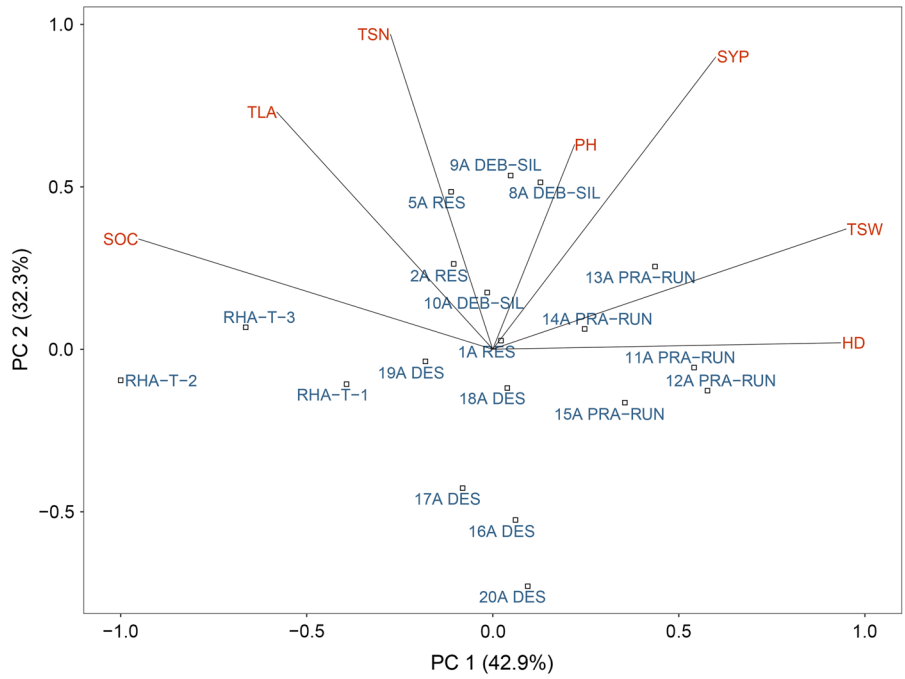

Fig. 2 PCA based on genotype by trait standardized data in the second year

previously mapped in vicinity of loci important for stress response, plant development, carbohydrate metabolism, nitrogen metabolism, and photosynthesis. Several were located near QTLs for different growth parameters, such as relative growth rate, leaf area, stem diameter, seed weight, and head diameter. PIC values of the used SSRs were high and resembled the published values (Tang and Knapp 2003). Only in case of ORS 1159, 546, and 437 the values were lower. Average PIC value was 0.590 (Table 3). Unique haplotypes were revealed by 8 and 9 SSR markers for all restorer lines and CMS inbred lines originating from interspecific line cross with *H. deserticola*, respectively. The highest number of unique haplotypes was 6, and was observed for restorer line RHA-T-2. Two unique haplotypes were observed for one new inbred line from interspecific line cross with *H. debilis silvestris* (10A DEB-SIL). ORS1021, ORS499 and HAP2 were among the most polymorphic and informative markers revealing unique haplotypes for two lines. HAP2 was used to explore Pl6 locus on LG 8. It revealed the presence of gene for downy mildew resistance in restorer line RHA-T-1 (Fig. 3), which was confirmed by resistant haplotype obtained by ORS166 in the same line (results not shown). Similar haplotypes were observed in CMS lines with PRA-RUN background.

Based on molecular data, Dpsa was calculated for all pairs of parental lines used in this experiment. Minimum value was 0.00, for 3 pairs of lines: 1A RES; 2A RES; 5A RES. Maximum value reached 0.92 for 2 pairs of lines: 10ADEB-SIL/11APRA-RUN and 10ADEB-SIL/12APRA-RUN. The average Dpsa value considering all sunflower lines was 0.6. On the other hand, minimum D_{psa}—distance between lines and each tester ranged from 0.50 to 0.60, maximum varied from 0.76 to 0.84, with average of 0.67–0.71.

Dpsa values calculated for all pairs of lines were analyzed using the PCoA method as explained in Materials and methods (Fig. 4). Both principal coordinates explained 58.4% of total variability. Lines have generally formed groups according to the wild parental species in initial interspecies cross. New CMS lines, developed after initial cross with interspecific line PRA-RUN and DES, formed clear distant groups. Male RHA lines grouped around the center of the graph. Lines from DEB-SIL and RES groups were close, one line 10A DEB-SIL was closer to three lines of RES group, which could not be differentiated by the set of markers used in this investigation.

Table 3 SSR primers used for the DNA polymorphism analysis of 19 sunflower parental lines including PIC values. Mapping positions close to QTL, gene or EST loci, involved in listed processes are presented where the data is available (Lai et al. 2005)

Linkage group	Primer	Haplotypes	PIC	Gene or EST locus/function	QTL	Unique haplotype
1	ORS 662	2	0.465			
	ORS552	4	0.693			RHA-T-2
2	ORS 423	4	0.593			
	ORS453	5	0.695			20A DES
	ORS 653	5	0.742			RHA-T-1
3	ORS 1021	5	0.638			10A DEB-SIL
						RHA-T-3
	ORS 432	4	0.647			10A DEB-SIL
	ORS 545	4	0.72	HT107/carbohydrate metabolism		
4	ORS 785	4	0.699			
	ORS 620	3	0.644	HT239/stress response		
	ORS 334	5	0.742			18A DES
	ORS 499	6	0.77			RHA-T-2
						RHA-T-3
	ORS 558	4	0.626			16A DES
	ORS 681	6	0.781	HT221/stress response		
	ORS 644	3	0.654			
5	ORS 1159	2	0.189			
	ORS 546	2	0.388			
	ORS 555	4	0.716			
6	ORS 381	2	0.499	HT185/abiotic stress tolerance		
	ORS 725	3	0.654			
7	ORS 1041	2	0.499			
	ORS 400	3	0.421			
8	HAP 2	5	0.715	Pl6/disease resistance		RHA-T-1
						RHA-T-2
	ORS 166	4	0.582			RHA-T-2
9	ORS 1265	3	0.571			
10	ORS 437	2	0.267			
	ORS 380	3	0.544	HT103/plant development	Relative growth rate; leaf area; stem diameter, initial seed weight	RHA-T-2
	ORS 613	3	0.547	HT227/transport mechanism	Relative growth rate; leaf area; stem diameter, height; peduncle length	18A DES

Table 3 continued

Linkage group	Primer	Haplotypes	PIC	Gene or EST locus/function	QTL	Unique haplotype
	ORS 691	3	0.587		Relative growth rate; leaf area; stem diameter, height; peduncle length	
11	ORS 621	5	0.615			RHA-T-1
	ORS 733	3	0.461		Leaf length, stem diameter	RHA-T-2
12	ORS 502	3	0.578	HT182/abiotic stress tolerance		
13	ORS 317	5	0.738	HT058/carbohydrate metabolism		20A DES
14	ORS 307	3	0.497	HT269/plant defense and senescence		19A DES
				HT251/nitrogen metabolism		
15	ORS 687	4	0.676	HT056/photosynthesis		18A DES
	ORS 7	3	0.421	HT220/abiotic stress tolerance	Head diameter	

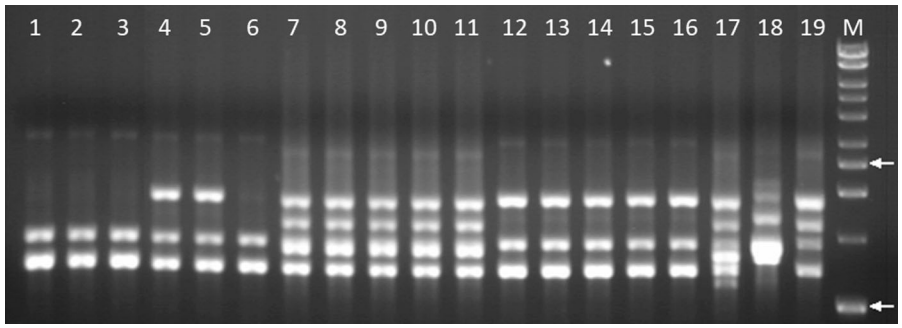

Fig. 3 Amplification profiles obtained with HAP2 primer with genomic DNA isolated from CMS lines developed from interspecies populations RES (lines 1–3), DEB-SIL (lines 4–6), PRA-RUN (lines 7–11), DES (lines 12–16), and restorer lines RHA-T-1 (line 17), RHA-T-2 (line 18), RHA-T-3 (line 19). Lane M 1/4 1 kb DNA ladder (Pharmacia Biosciences). Fragments of 1000 and 2500 bp are marked with arrows. Fragments were separated using 1% agarose gel

General combining ability (GCA)

The analysis of GCA values for examined traits showed that the A lines descending from interspecific line and RHA testers differed significantly in both years (Tables 4, 5). CMS lines developed from RES interspecific line had highly significant ($p < 0.01$) positive GCA values for SOC and highly significant negative values for PH in both years. However, CMS lines from DEB-SIL background had high significant positive GCA for TSN and significant negative values for TSW in both years. The best positive GCA for SYP, TLA, PH and TSW in both years were observed in most lines that originated from PRA-RUN interspecific line. Only one line, 12A PRA-RUN, had a highly significant negative GCA value for SYP and TLA. Lines from interspecific line PRA-RUN had a highly significant negative GCA value for SOC and TSN. Positive highly significant GCA effects for HD were observed in lines from interspecific line DES in both years. Genotypes with highly significant negative GCA values, for SYP, PH and TSW originated from interspecific line DES in both years. The best line from interspecific line DEB-SIL is 8A which demonstrates a highly positive GCA for multiple traits (SOC, TLA, PH and TSN) in both years. Highly significant positive values for five traits (SY, TLA, PH, HD and TSW) were found in line 11A PRA-RUN, which was the best in this group. Significant positive values of GCA for SOC, HD, TSN were found in CMS inbred line 18A DES (Tables 4, 5).

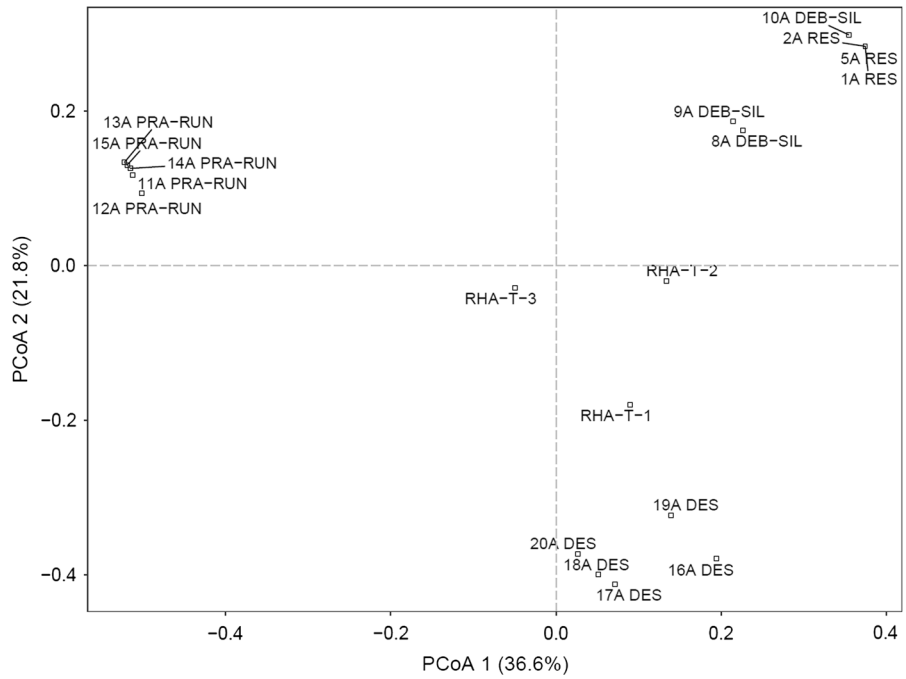

Fig. 4 Plot of the first and second principal coordinate analysis (PCoA) based on D_{SA} distance matrix

Inbred lines demonstrating the highest significant positive GCA value for SYP were 11A, 14A PRA-RUN; for SOC were lines 1A, 2A, 3A RES; for TLA were lines 11A, 13A PRA-RUN; for PH were lines 14A,11A PRA-RUN; for HD were lines 19A, 20A DES; for TSN were lines 8A,10A DEB-SIL and for TSW were lines 12A, 13A PRA-RUN.RHA-T-2 line had high significant positive GCA values for multiple traits (SYP, SOC, PH, HD and TSN) and was the best among Rf lines in both years.

Genetic divergence and heterosis

Mean values of best parent heterosis (BPH) were calculated through each restorer line for seven examined traits (Table 6). Almost all hybrid combinations exhibited significant BPH ($t \geq 0.05$) for following traits: SYP, PH, HD and TSN. The best BPH values for almost all traits were obtained for hybrid combinations with RHA-T-2. Hybrids with RHA-T-1 had the highest BPH for TLA only.

The relationship between best parent heterosis and genetic distance, based on molecular markers, was examined for each of the yield traits by LOESS nonparametric technique (Fig. 5). Basically, the individual models for the relationship from experimental data were tested by estimating the Pearson's linear correlation coefficient between the dependent variable and fitted values of the individual models. The value of the correlation coefficients reported as r^2 indicates the quality of the non-parametric regression analysis. It ranged from about $r^2 \sim 0.4$ for SYP, HD and TSN, to $r^2 \sim 0.7$ for SOC. Generally the fitted curves exhibited polyphasic kinetics and differed among examined yield traits. For most examined traits, such as SY, SOC, TLA, TSN and TSW heterosis decreased in the range of GD from about 0.5 to about 0.6. When GD between parental lines was higher than 0.7, the heterosis of SYP, SOC and TSW decreased further, while heterosis for TLA and TSN increased significantly. However, the best parent heterosis for PH and HD increased with GD between parental lines in similar manner (Fig. 5).

Discussion

Cultivated sunflower generally has limited genetic variability, especially for major agronomic traits (Zambelli et al. 2015), therefore the creation of a new sunflower ideotype demands broadening of the genetic diversity of sunflower and requires an increased use of wild *Helianthus* species in breeding programs (Škorić et al. 2012). The fact that sunflower

Table 4 GCA values for seven agronomic traits (*SYP* seed yield per plant, *SOC* seed oil content, *TLA* total leaf area, *PH* plant height, *HD* head diameter, *TSN* total seed number per head and *TSW* 1000 seeds weight) of sunflower inbred lines in the first year. Lines were developed from interspecies populations with annual wild species: *H. debilis silvestris* (DEB-SIL), *H. preacox* ssp. *runionii* (PRA-RUN) and *H. deserticola* (DES); and perennial wild species *H. resinosus* (RES). Lines 17–19 are restorer lines-testers

No.	Lines	SYP	SOC	TLA	PH	HD	TSN	TSW
1	1A RES	− 11.24	2.90**	− 426.8	− 8.82	− 2.37	117.77**	− 10.34
2	2A RES	− 14.44	2.85**	− 241.1	− 3.48	− 1.86	81.96**	− 12.07
3	5A RES	− 14.86	2.53**	− 494.1	− 6.37	− 1.39	− 9.90	− 9.90
4	8A DEB-SIL	− 0.11	0.72**	361.5**	11.91**	− 0.71	52.03**	− 2.21
5	9A DEB-SIL	5.11**	− 0.22	− 697.1	10.57**	− 1.31	142.57**	− 2.16
6	10A DEB-SIL	7.59**	1.12**	− 59.2	− 0.32	0.03	214.79**	− 2.86
7	11A PRA-RUN	16.65**	− 1.55	1882.3**	14.41**	1.39**	− 97.76	10.83**
8	12A PRA-RUN	− 2.29	− 3.40	− 604.9	12.14**	− 0.89	− 227.71	17.35**
9.	13A PRA-RUN	12.52**	− 1.82	1867.3**	12.02**	− 0.87	− 107.90	7.71**
10	14A PRA-RUN	16.63**	− 1.49	387.2**	16.85**	− 0.20	− 146.34	14.13**
11	15A PRA-RUN	14.58**	− 3.97	1274.4**	4.85**	1.71**	− 109.98	13.09**
12	16A DES	− 10.86	− 2.00	20.4	− 19.04	1.00**	− 186.84	− 0.81
13	17A DES	− 10.73	1.47**	− 471.7	− 11.43	0.65**	− 49.70	− 7.07
14	18A DES	− 2.96	1.04**	− 754.2	− 8.76	0.59**	195.41**	− 3.49
15	19A DES	− 2.14	− 0.09	− 641.1	− 9.15	2.19**	51.42**	− 2.92
16	20A DES	− 3.46	1.92**	− 1402.8	− 15.37	2.05**	80.19**	− 9.31
	LSD 0.05	1.335	0.166	97.77	0.888	0.207	12.219	0.338
	LSD 0.01	2.002	0.249	146.66	1.332	0.310	18.928	0.512
17	RHA-T-1	− 1.69	− 0.68	178.0	1.38	− 0.50	− 147.82	3.66**
18	RHA-T-2	5.81**	0.62**	− 159.1	4.79**	0.53**	253.37**	− 4.43
19	RHA-T-3	− 4.12	0.05	− 18.9	− 6.17	− 0.03	− 105.55	0.77**
	LSD 0.05	0.578	0.072	42.338	0.385	0.090	5.291	0.147
	LSD 0.01	0.867	0.108	63.506	0.577	0.134	7.94	0.220

*≤ 0.05; **≤ 0.01

is among the leading crops of major importance for global food security in the use of traits from crop wild relatives (Seiler et al. 2017) emphasize interspecies hybridization as an effective approach for the increase of genetic variability of cultivated sunflower and its improvement. However, the process is long and cumbersome. For example, the starting plant material for this study was obtained after the progeny of the first interspecies cross was back-crossed twice with sunflower inbred lines and then selfed for 3–5 generation and some of lines were also sib-crossed for 1–2 generations (Seiler 1991a, b, c). In this study, lines from the starting plant material were selfed for three generations, which was followed by eight backcrosses as explained in detail in Materials and methods. Cost- and time-efficient methods for the determination of best parent heterosis and the use of best inbred lines in crosses for obtaining high-yielding and stable hybrids are generally desirable, but in case of introgression lines it becomes a must.

Morphological traits of parental inbred lines, PCA and GCA effects

The PCA used to compare new CMS lines on the basis of examined traits, indicated that all lines grouped according to their origin generally, however lines originating from PRA-RUN interspecific line and restorer lines formed two most distinguished groups (Figs. 1, 2). Both PCA and correlation coefficients between traits indicate a strong positive correlation between SYP and TSW, which is in agreement with the results of Yalcin et al. 2007 and Behradfar et al. 2009. Mean values of studied traits observed for the

Table 5 GCA values for seven agronomic traits (*SYP* seed yield per plant, *SOC* seed oil content, *TLA* total leaf area, *PH* plant height, *HD* head diameter, *TSN* total seed number per head and *TSW* 1000 seeds weight) of sunflower inbred lines in the second year. Lines were developed from interspecies populations with annual wild species: *H. debilis silvestris* (DEB-SIL), *H. preacox* ssp. *runionii* (PRA-RUN) and *H. deserticola* (DES); and perennial wild species *H. resinosus* (RES). Lines 17–19 are restorer lines-testers

No.	Lines	SYP	SOC	TLA	PH	HD	TSN	TSW
1	1A RES	− 3.93	0.67**	678.3**	− 8.64	− 0.58	10.91	1.50**
2	2A RES	3.88**	2.78**	142.4**	− 5.25	0.56**	130.20**	− 0.39
3	5A RES	− 8.30	1.62**	− 338.0	− 4.58	− 1.58	− 140.96	1.78**
4	8A DEB-SIL	13.37**	− 1.53	112.9**	7.08**	− 1.10	241.13**	− 1.21
5	9A DEB-SIL	8.95**	− 0.94	− 1904.4	11.25**	− 0.87	92.87**	− 6.10
6	10A DEB-SIL	− 1.26	1.47**	− 259.1	2.92**	− 0.93	84.61**	0.72
7	11A PRA-RUN	24.92**	− 3.52	2562.4**	12.64**	0.01	59.92**	14.47**
8	12A PRA-RUN	− 13.92	− 1.21	− 1161.3	13.31**	− 1.39	− 211.40	8.44**
9	13A PRA-RUN	26.82**	− 2.14	2113.3**	6.64**	− 0.57	38.20**	17.61**
10	14A PRA-RUN	18.10**	− 2.68	280.0**	16.47**	− 0.97	− 118.96	9.48**
11	15A PRA-RUN	3.23**	− 1.66	750.8**	6.99**	0.20**	− 110.98	2.59**
12	16A DES	− 24.78	0.05	− 645.4	− 16.38	0.52**	− 177.34	− 10.97
13	17A DES	− 15.67	2.17**	− 476.5	− 11.64	0.19*	− 51.70	− 8.89
14	18A DES	− 5.30	0.99**	− 333.9	− 10.47	1.50**	170.97**	− 7.75
15	19A DES	− 10.67	1.52**	− 457.2	− 11.25	2.31**	18.61*	− 10.76
16	20A DES	− 15.43	2.41**	− 1064.6	− 9.08	2.67**	− 36.08	− 10.53
	LSD 0.05	0.898	0.178	41.01	1.328	0.128	15.308	0.831
	LSD 0.01	1.347	0.268	61.51	1.992	0.192	22.962	1.247
17	RHA-T-1	1.83**	− 1.12	191.2**	0.83	− 0.10	− 52.89	4.01**
18	RHA-T-2	3.68**	0.63**	− 338.20	3.17	0.22**	186.22**	− 5.24
19	RHA-T 3	− 5.50	0.49**	147.0**	− 3.99	− 0.11	− 133.324	1.24**
	LSD 0.05	0.389	0.077	17.756	0.575	0.055	6.629	0.360
	LSD 0.01	0.583	0.116	26.634	0.863	0.083	9.943	0.540

*≤ 0.05; **≤ 0.01

examined populations and Rf testers were higher in the second year, except for SOC. However, in this experiment the interaction between genotype and environment was low, i.e. the ranking of lines according to origin for most traits was similar in both years (Table 1). Restorer lines had the highest mean values of SOC, especially RHA-T-2, and grouped near the SOC vector in both years (Figs. 1, 2). Highly significant positive values of GCA for SOC for the same restorer line was also observed. (Tables 4, 5). SOC of CMS lines originating from perennial wild species *H. resinosus* was only about 6% lower than in Rf, while GCA for SOC was also highly significant positive. According to Breton et al. (2010), perennial species have a high potential to breed sunflower for any trait, but progenies of sunflower line x perennial usually have low oil content that requires several years of improvement. Our results showed that SOC can be improved during the prebreeding process and the early generations of inbred lines. In the literature, there is little research data on oil content with lines obtained through interspecies hybridization with perennial sunflower species, since it is difficult to obtain such crosses in order to improve agronomic traits such as SOC. CMS lines with annual wild background had from 14% up to 30% lower SOC than Rf lines, with mostly negative and insignificant GCA for SOC except the lines 10A DEB SIL, 17A and 20A DES which had highly positive significant GCA values. These results are in agreement with the findings of Seiler (2012) who reported that the cultivated sunflower shows a significant variability in oil content, although its amount is generally low in wild species.

Table 6 Mean values and standard deviations of best parent heterosis (BPH) over each restorer line for seven traits (*SYP* seed yield per plant, *SOC* seed oil content, *TLA* total leaf area, *PH* plant height, HD head diameter, *TSN* total seed number per head and *TSW* 1000 seeds weight)

	RHA-T-1	RHA-T-2	RHA-T-3
SYP/plant (g)			
BPH ± SD	126.69 ± 61.97	134.31 ± 76.27	109.53 ± 46.72
H*	14	16	16
SOC (g/100 g seed)			
BPH ± SD	− 7.07 ± 4.70	− 8.30 ± 3.81	− 2.52 ± 4.05
H*	7	7	2
TLA (cm^2)			
BPH ± SD	34.19 ± 20.79	14.84 ± 14.83	5.55 ± 17.07
H*	5	1	1
PH (cm)			
BPH ± SD	27.32 ± 10.41	45.41 ± 10.80	23.48 ± 11.87
H*	15	16	15
HD (cm)			
BPH ± SD	27.45 ± 10.72	42.04 ± 14.86	24.74 ± 10.23
H*	15	13	15
TSN			
BPH ± SD	110.80 ± 33.82	132.54 ± 26.17	80.44 ± 15.14
H*	16	16	16
TSW (g)			
BPH ± SD	7.46 ± 14.76	− 2.56 ± 14.70	2.84 ± 13.88
H*	6	4	6

The number of hybrids showing significance over best parent heterosis ($t \geq 0.05$) is also presented (H*). In total 48 sunflower hybrids were analysed

The highest values of PH were observed in lines of DEB-SIL background, that grouped at PH vector in both years (Figs. 1, 2) and had highly significant GCA for this trait (Tables 4, 5). PH plays a major role in the creation of new hybrids with high genetic potential for seed yield, but it is strictly linked to total leaf area and petiole length, which are relevant for seed yield per plant (Hladni 2010). PCA indicates strong positive correlation between PH and SYP (Figs. 1, 2), but significant positive correlation coefficient was found between PH and TLA. PH is important trait because it influences the stability of the plant via tolerance to lodging and certain diseases (Kaya 2016). Producing short sunflowers with increased yield was proposed by Vear (2016), who reported that short sunflowers could be sown at higher density, with a crop with numerous small heads, which covers the ground more quickly in spring and matures more rapidly, thereby permitting earlier harvest. If the goal is the change of sunflower plant architecture, genotypes originating from DES interspecific line with shorter PH (Tables 1, Figs. 1, 2) and negative GCA values for PH (Figs. 4, 5) could be desirable in breeding programs.

Strong positive correlation coefficients between TSW and HD was observed in both years and the lines of originating from PRA-RUN grouped between these vectors (Figs. 1, 2). This was also confirmed by GCA values determined for the same plant material. Beside these traits PRA-RUN lines were the best general combiners for SYP, TLA and PH. Highly significant positive GCA value for five traits (SYP, TLA, PH, HD and TSW) were present in line 11A PRA-RUN and it was the best from this group.

It is very difficult to combine all positive traits in one crossing combination. The inbred line 8A and 9A from DEB-SIL interspecies background showed a highly significant positive GCA value for PH and TSN and were the best in a PCA for those two traits in both years. As expected the parents with higher mean values are better general combiners, and lines with lower mean values lower general combiners. For example, CMS inbred line originating from DES with the best GCA for HD has the highest mean values for that trait, while the line with the worst GCA for PH and TSW has the lowest mean value for that trait. Highly significant positive GCA values for SYP were

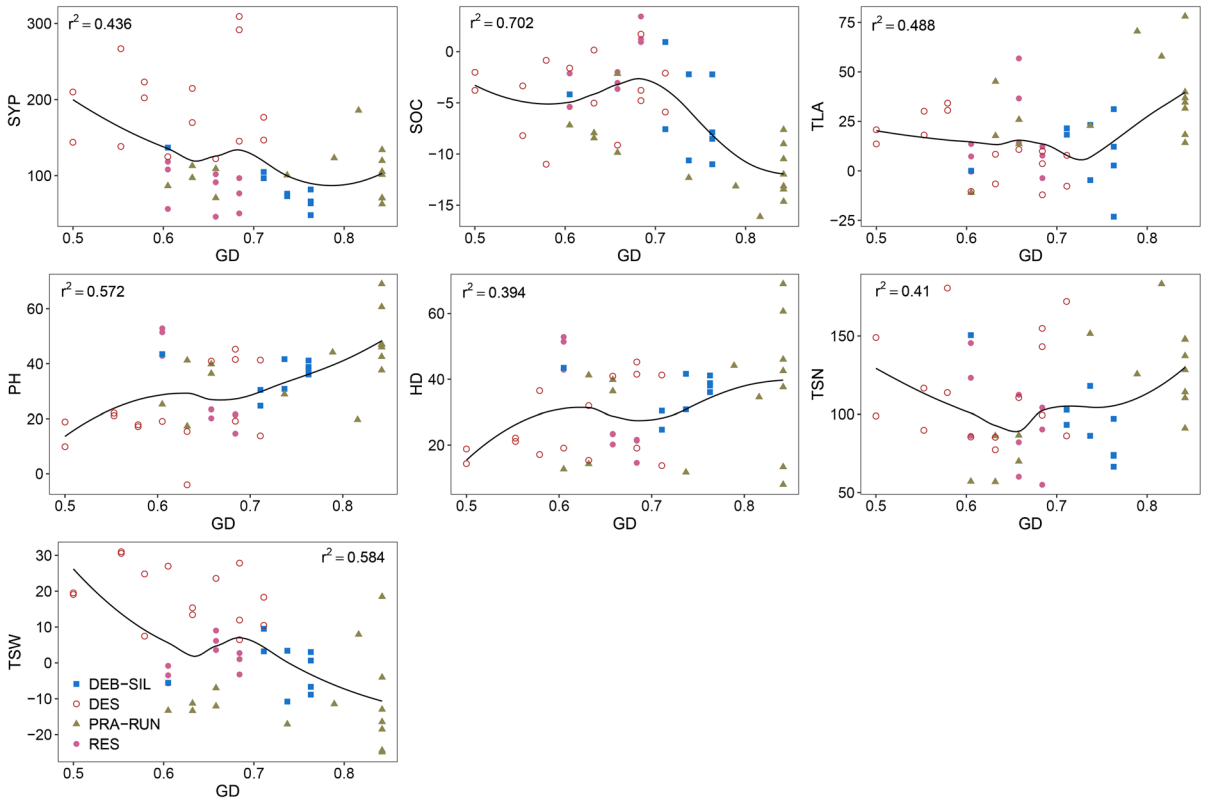

Fig. 5 Relationships between best parent heterosis and genetic distance based on molecular markers for each of the examined traits (seed yield-SYP, seed oil content-SOC, total leaf area-TLA, plant height-PH, head diameter-HD, total seed number per head-TSN and 1000 seeds weight-TSW). Hybrids from crosses with new CMS lines of the same interspecies origin (RES, PRA-RUN, DES and DEB-SIL) are presented with different symbols. Curves were fitted to empirical data by LOESS nonparametric technique. On each graph r^2 indicates the quality of the nonparametric regression models

observed in lines 11A, 14A PRA RUN; for SOC 2A RES for TLA 11A, 13A PRA RUN; for PH 14A, 11A PRA RUN; and for HD 19A, 20A DES, 16A DES.

PCoA of molecular marker data of parental inbred lines

In this paper, the genomic DNA polymorphism was investigated by limited number of SSRs, which were distributed on all linkage groups and carefully selected either for their previously estimated high polymorphic values (Tang et al. 2003; Tang and Knapp 2003) or expressed sequence tags/quantitative trait loci (EST/QTL) associations (Lai et al. 2005). Fillipi et al. (2015) have compared analysis of genetic variability obtained on 42 SSR loci and/or 384 SNPs. Though the number of SSR markers was a lot smaller than SNPs, both types of markers revealed moderate genetic variability of germplasm collection composed of sunflower inbred lines, open-pollinated and composite populations. High levels of differentiation among accessions of sunflower open-pollinated and composite populations were reported by Moreno et al. (2013). The pairwise genetic distance up to 0.8, was detected for RILs derived from the initial cross of two inbred lines with interspecies background (Darvishzadeh 2012).

We have observed maximum genetic distance of 0.92, between two pairs of inbred lines: 10A DEB-SIL and 11A PRA-RUN and 10A_DEB-SIL/12A_PRA-RUN. Line 10A_DEB-SIL which originates from the wild sunflower species *H. debilis* is considered to be the most primitive of the annual sunflowers (Rieseberg and Soltis 1987). Its subspecies silvestris is endemic to Texas, USA. Molecular phylogenetic evidence indicates that *H. praecox* is derived from *H. debilis* (Rieseberg 1991). Line 10A_DEB-SIL grouped with lines from RES-1 interspecies population, rather than with other DEB-SIL lines (Fig. 4). RES-1 lines are

descending from *H. resinosus*, the only perennial hexaploid wild species used in this investigation. Several crosses, first for the production of the interspecific lines (Seiler 1991a, b, c) and the following for the development of new CMS lines, changed the GDs of new lines in comparison with the phylogenic relations of their wild progenitors. The lowest GDs between cultivated Rf *H. annuus* lines and new CMS lines was determined for DES lines (0.62), developed from *H. deserticola,* which is a natural diploid hybrid of *H. annuus* and *H. petiolaris* (Seiler 2007).

Out of the 37 SSR primers, in total 17 revealed unique haplotypes for restorer inbred lines and new CMS inbred lines originating from DES interspecific line. However, it was not possible to distinguish inbred lines with RES background by selected SSR primers, though all linkage groups were covered and nearly 50% of them were mapped near genes important for stress response, plant development, carbohydrate metabolism, nitrogen metabolism and photosynthesis. Pl_6 locus on LG8 was among most polymorphic loci. Its analysis confirmed the presence of typical profile of Pl_6 downy mildew resistance gene in restorer line RHA-T-1 (Panković et al. 2007). On the other hand, restorer line RHA-T-3 and CMS lines from PRA-RUN interspecific line had similar profile for this locus. Since this restorer line is resistant to downy mildew race 730, the obtained results indicate firstly the presence of genes from wild species in this line, and secondly that CMS lines from PRA-RUN population should be included in downy mildew resistance program (Škorić, personal communication).

GD versus BPH

It is generally accepted that by crossing of genetically similar lines, it is not possible to get high heterosis, and that higher genetic distance between parental lines is a precondition for the expression of good specific combining ability (SCA) (Badu-Apraku et al. 2011, Kane et al. 2013). It was observed that the correlations of marker distances of the parents with both mid-parent and best-parent heterosis for agronomic traits were too low to be of predictive value in maize (Wegary et al. 2013; Ndhlela et al. 2015). This relationship is not so often studied in sunflower. Darvishzadeh (2012) has found significant positive correlation between GD of sunflower lines and mean parent heterosis for SYP and chlorophyll contents in well-watered conditions. However, in water-stress significant negative correlations with leaf number, head weight, aerial part dry weight and plant height were observed. In this study, the relationship between GD based on molecular marker data and best parent heterosis for each of the seven yield traits, was examined by LOESS nonparametric simple strategy used to fit the smooth curves to empirical data. It is beneficial fitting technique as it does not require specification of the relationship between the dependent and independent variables. It is particularly valuable in case of the presence of outliers, like extreme parameter values. Mostly, it is used as a scatter plot smoother, but also, it can be generalized very easily to multivariate data (Jacoby 2000). According to our results the kinetics of polyphasic fitted curves differed among the yield traits. BPH for SYP, SOC and TSW generally decreased with GD. For these traits the best parental lines would be with GD of about 0.5. However BPH for HD and PH generally increased with GD. The relationship between BPH for TLA and TSN and GD was more complex, the decrease up to 0.6 is followed by the increase up to 0.9. So for HD, PH, TLA and TSN the best would be to select parental lines with bigger GD. Yet intermediate to high GD between parental lines was optimal for best heterotic effects of most traits. Higher number of parental combinations would be necessary for LOESS analysis followed by determination of correlation coefficients in separate ranges of GDs. Similar results were obtained on chilli by Krishnamurthy et al. (2013). Nonlinear regression of mid parent heterosis versus total genetic divergence of parental chilli lines revealed intermediate parental divergence as best for the occurrence of higher frequencies of heterotic crosses.

Genomic selection is the most precise in predicting the breeding values of wheat lines (Heffner et al. 2011). However, recent results on genomic prediction of sunflower hybrid performance are only partially in accordance to this. For example, Reif et al. (2013) used 572 AFLP markers in the best linear unbiased prediction model for genomic prediction of sunflower hybrid performance and observed that genomic prediction was accurate only in the group of related sunflower lines. Reif et al. (2013) concluded that prediction accuracy of hybrid performance based on GCA effects was high and could not be increased by genomic prediction. Similarly, Mangin et al. (2017)

have shown that genomic predictions using whole genome sequencing, increased breeding efficiency in comparison to the classical GCA modeling for oil content, only in case when either one or both sunflower parental lines were not well-characterized. Genomic selection with several hundreds or thousands of markers is suitable for regular crosses with a regular recombination frequency. In case of prediction of heterotic potential of lines derived from interspecies crosses there are some specificities in comparison to the elite lines. The size of the introgressed regions slowly decrease in the crossing cycles since the number of recombination events is lower. Therefore, SSR markers specific for introgressed regions stay stable for many breeding cycles. In such way, limited number of molecular markers, or even only one diagnostic marker per introgressed fragment could be sufficient in selection procedure, as observed by Perovic et al. (2009) for Sbm1 introgression in wheat. Therefore, fewer specific markers are needed for investigation of introgressed genomic regions. In this study we observed many unique haplotypes for lines analysed with only 38 SSRs and PCoA, illustrated grouping of lines according to the wild parental species in initial interspecies cross.

Conclusions

In this study four different methods were applied to test breeding performance of inbreed introgression lines derived from crosses to four wild sunflower relatives. Results show that in the interspecific hybridization of cultivated sunflower high number of unique haplotypes obtained by SSR markers could be useful for fast and cheap selection in breeding schemes. Pl_6 locus was one of the most variable and informative loci among examined sunflower lines indicating that lines with PRA RUN background should be involved in breeding programs for downy mildew resistance. Lines originating from the same wild parent were the best general combiners for SYP, TLA, TSW and PH. CMS lines derived from perennial wild species *H. resinosus* background were the best for SOC. Lines of DEB-SIL background would be the best choice if high PH and correlated high SYP are targeted. In case the target is changed sunflower plant architecture towards shorter PH, genotypes originating from *H. deserticola* would be appropriate.

Results obtained in this research show that PCA of morphological data on parental lines are generally in agreement with GCA effects for examined traits. Lower costs and needed time makes the first method of choice for the parental choice of introgression and elite lines in sunflower breeding programs. PCoA of molecular marker data showed broad genetic diversity between lines. GD versus BPH relationships indicate that intermediate to high GD between parental lines was optimal for best heterotic effects of most traits. GDs between parental lines over 0.7 could diminish heterosis for SOC and TSW. The combination of the PCA of morphological data, PCoA of molecular marker data and GD between parental lines is fast and affordable and in the same time gives the most important information for parental choice.

Acknowledgements This research was partly financed by the project 31025: Development of new varieties and production technology improvement of oil crops for different purposes, from the Ministry of Education and Science Republic of Serbia. The authors are grateful to Dr. Gerald Seiler (USDA-ARS, Fargo ND, USA) for providing interspecies populations and to Prof. Dragan Škorić (Serbian Academy of Sciences and Arts, Novi Sad Branch), and Dr. Dejan Dodig (Maize Research Institute, Zemun Polje, Serbia) for fruitful discussions during preparation of the manuscript.

References

Abid MA, Malik W, Yasmeen A, Quayyum A, Zhang R, Liang C, Guo S, Ashraf J (2016) Mode of inheritance for biochemical traits in genetically engineered cotton under water stress. AOB Plants 8:1–15

Ambachew D, Mekbib F, Asfaw A, Beebeand SE, Blair MW (2015) Trait associations in common bean genotypes grown under drough stress and field infestation by BSM bean fly. Crop J 3:305–316

Atlagić J, Terzić S (2014) Sunflower genetic resources: interspecific hybridization and cytogenetics in prebreeding. In: Arribas JI (ed) Sunflowers: growth and development, environmental influences and pests/diseases. Nova Science Publishers, New York, pp 95–130. https://www.novapublishers.com/catalog/product_info.php?products_id=48246

Badu-Apraku B, Lum AF, Akinwale RO, Oyekunle M (2011) Biplot analysis of diallel crosses of early maturing tropical yellow maize inbreds in stress and nonstress environments. Crop Sci 51:173–188. https://doi.org/10.2135/cropsci2010.06.0366

Behradfar A, Gorttapeh H, Zardashty MR, Talat F (2009) Evaluation correlated traits for seed and oil yield in sunflower through path analysis in under condition relay cropping. Res J Biol Sci 4:82–85

Berry ST, Allen RJ, Barnes SR, Caligari PDS (1994) Molecular marker analysis of *Helianthus annuus* L. 1. Restriction

fragment length polymorphism between inbred lines of cultivated sunflower. Theor Appl Genet 89:435–441

Blackman BK, Scascitelli M, Kane NC, Luton HH, Rasmussen DA, Bye RA, Lentz DL, Rieseberg LH (2011) Sunflower domestication alleles support single domestication center in eastern North America. PNAS 108(34):14360–14365

Bonnafous F, Langdale N, Sunrise Consortium, Mangin B (2016) Inclusion of dominance effect in genomic selection model to improve predictive ability for sunflower hybrid performance. In: Proceedings of the 19th international sunflower conference, Edirne, p 285

Breton C, Serieys H, Berville A (2010) Gene transfer from wild *Helianthus* to sunflower: topicalities and limits. OCL 17(2):104–114. Article disponible sur le site http://www.ocl-journal.org. http://dx.doi.org/10.1051/ocl.2010.0296

Burke JM, Knapp SJ, Rieseberg LH (2005) Genetic consequences of selection during the evolution of cultivated sunflower. Genetics 171:1933–1940

Cheres MT, Miller JF, Crane JM, Knapp SJ (2000) Genetic distance as a predictor of heterosis and hybrid performance within and between heterotic groups in sunflower. Theor Appl Genet 100:889–894

Christensen R (2015) Analysis of variance, design, and regression: linear modeling for unbalanced data, 2nd edn. Chapman & Hall/CRC Press, Boca Raton

Cleveland WS (1979) Robust locally weighted regression and smoothing scatter plots. Am Stat Assoc 74(368):829–836

Darvishzadeh R (2012) Phenotypic and molecular marker distance as a tool for prediction of heterosis and F_1 performance in sunflower (*Helianthus annuus* L.) under well-watered and water-stressed conditions. Aust J Crop Sci 6(4):732–738

Dias Dos Santos LA, De Toledo Picoli EA, Rocha RB, Alfenas AC (2004) A priori choice of hybrid parents in plants. Genet Mol Res 3(3):356–368

Fillipi CV, Aquirre N, Rivas JG, Zubrycki J, Puebla A, Cordes D, Moreno MV, Fusari CM, Alvarez D, Heinz RA, Hopp HE, Paniego NB, Lia VV (2015) Population structure and genetic diversity characterization of a sunflower association mapping population using SSR and SNP markers. BMC Plant Biol 15(52):1–12

Heffner EL, Jannink JL, Iwata H, Souza H, Sorrells ME (2011) Genomic selection accuracy for grain quality traits in biparental wheat populations. Crop Sci 51:2597–2606

Hladni N (2010) Genes and sunflower yield. Foundation Andrejević, Beograd, p 116

Hladni N, Miklič V (2012) Old and new trends of using genetic resources in sunflower plant breeding with the aim of preserving biodiversity. In: Proceedings of the 4th Joint UNS-PSU international conference on biosci: biotechnology and biodiversity, Novi Sad, pp 109–120

Hongtrakul V, Huestis GM, Knapp SJ (1997) Amplified fragment length polymorphism as a tool for DNA fingerprinting sunflower germplasm: genetic diversity among oilseed inbred lines. Theor Appl Genet 95:400–407

Jacoby WG (2000) LOESS: a nonparametric, graphical tool for depicting relationships between variables. Elect Stud 19(4):577–613

Kane N, Burke J, Marek L, Seiler GJ, Vear F, Baute G, Knapp S, Vincourt P, Rieseberg LH (2013) Sunflower genetics, genomics and ecological resources. Mol Ecol Resour 13:10–20

Kaya Y (2016) Chapter 4 - Sunflower, In: Gupta SK (ed) Breeding oilseed crops for sustainable production. Academic Press, San Diego, pp 55–88. https://doi.org/10.1016/B978-0-12-801309-0.00004-5

Krishnamurthy SL, Rao AM, Reddy KM, Ramesh S, Hittalmani S, Rao MG (2013) Limits of parental divergence for the occurrence of heterosis through morphological and AFLP marker in chilli (*Capsicum annuum* L.). Curr Sci 6:25

Kroonenberg PM (1995) Introduction to biplots for G × E tables. Department of Mathematics, Research Report 51. University of Queensland, Australia. http://three-mode.leidenuniv.nl/document/biplot.pdf

Lai Z, Livingstone K, Zou Y, Church SA, Knapp SJ, Andrews J, Rieseberg LH (2005) Identification and mapping of SNPs from ESTs in sunflower. Theor Appl Genet 111:1532–1544

Leclercq P (1969) Une sterilite cytoplasmique chez le tournesol. Ann Amelior Plantes 19:99–106

Liu J (2002) PowerMarker: a powerful software for marker data analysis. Raleigh, North Carolina State University, Bioinformatics Research Center. http://statgen.ncsu.edu/powermarker/

Livaja M, Unterseer S, Erath W, Lehermeier C, Wieseke R, Plieske J et al (2016) Diversity analysis and genomic prediction of *Sclerotinia* resistance in sunflower using a new 25 K SNP genotyping array. Theor Appl Genet 129:317–329

Mandel JR, Dechaine JM, Marek LF, Burke JM (2011) Genetic diversity and population structure in cultivated sunflower and a comparison to its wild progenitor, *Helianthus annuus* L. Theor Appl Genet 123:693–704. https://doi.org/10.1007/s00122-011-1619-3

Mangin B, Bonnafous F, Blanchet N, Boniface M-C, Bret-Mestrie E, Carrère S et al (2017) Genomic prediction of sunflower hybrids oil content. Front Plant Sci 8:1633. https://doi.org/10.3389/fpls.2017.01633

Mohayeji M, Capriotti AL, Cavaliere C, Piovesana S, Samperi R, Stampachiacchiere S, Toorchi M, Lagana A (2014) Heterosis profile of sunflower leaves: a label free proteomics approach. J Proteom 99:101–110

Moreno MV, Nishinakamasu V, Loray MA, Alvarez D, Gieco 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 Resour Charact Util 11(3):238–249

Ndhlela T, Herselman L, Semagn K, Magorokosho C, Mutimaamba C, Labuschagne MT (2015) Relationships between heterosis, genetic distances and specific combining ability among CIMMYT and Zimbabwe developed maize inbred lines under stress and optimal conditions. Euphytica 204(3):635–647

Panković D, Radovanović N, Jocić S, Šatović Z, Škorić D (2007) Development of Co-dominant amplified polymorphic sequence markers for resistance to sunflower downy mildew race 730. Plant Breed 126(4):440–444

Perovic D, Förster J, Devaux P, Hariri D, Guilleroux M, Kanyuka K, Lyons R, Weyen J, Feuerhelm D, Kastirr U, Sourdille P, Röder M, Ordon F (2009) Mapping and diagnostic marker development for soil-borne cereal mosaic virus resistance in bread wheat. Mol Breed 23:641–653

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1(7):215–222

R Core Team (2016) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. Accessed Oct 2017

Reif JC, Zhao Y, Würschum T, Gowda M, Hahn V (2013) Genomic prediction of sunflower hybrid performance. Plant Breed 132(1):107–114

Rieseberg LH (1991) Homoploid reticulate evolution in *Helianthus* (Asteraceae): evidence from ribosomal genes. Am J Bot 78:1218–1237

Rieseberg LH, Soltis DE (1987) Phosphoglucomutase in *Helianthus debilis*: a polymorphism for isoenzyme number. Biochem Syst Ecol 15(5):545–548

Saftić-Panković D, Šatović Z, Radovanović N, Jocić S, Miklič V (2008) Positioning of CAPS markers for resistance to downy mildew on linkage maps as determined in three sunflower mapping populations. In: Proceedings of the 43rd Croatian and 3rd international symposium on agriculture, Opatija, Croatia, pp 357–361

Seiler GJ (1991a) Registration of six interspecific sunflower germplasm lines derived from wild perennial sunflower. Crop Sci 31:1097–1098. https://doi.org/10.2135/cropsci1991.0011183X003100040074x

Seiler GJ (1991b) Registration of 13 downy mildew tolerant interspecific sunflower germplasm lines derived from wild annual species. Crop Sci 31:1714–1716. https://doi.org/10.2135/cropsci1991.0011183X003100060093x

Seiler GJ (1991c) Registration of 15 interspecific sunflower germplasm lines derived from wild annual species. Crop Sci 31:1389–1390. https://doi.org/10.2135/cropsci1991.0011183X003100050081x

Seiler GJ (2007) Wild annual *Helianthus anomalus* and *H. deserticola* for improving oil content and quality in sunflower. Ind Crops Prod 25:95–100

Seiler GJ (2012) Utilization of wild *Helianthus* species in sunflower breeding. In: Škorić D et al (eds) Sunflower genetics and breeding. International monography. Serbian Academy of Science and Arts Branch, Novi Sad, p 520

Seiler GJ, Qi LL, Marek LF (2017) Utilization of sunflower crop wild relatives for cultivated sunflower improvement. Crop Sci 57:1–19

Singh RK, Chaudhury BD (2001) Biometrical techniques in breeding and genetics. Saujanya Books, Delhi, p 350

Škorić D, Seiler GJ, Zhao L, Chao-Chien J, Miller JF, Charlet LD (2012) Sunflower genetics and breeding. International monography. Serbian Academy of Science and Arts Branch, Novi Sad, p 520

Smith B (2014) The domestication of *Helianthus annuus* L. (sunflower). Veg Hist Archaeobot 23:57–74. https://doi.org/10.1007/s00334-013-0393-3

Sprague GF, Tatum LA (1942) General versus specific combining ability in single crosses of corn. J Am Soc Agron 34:923–932

Syafii M, Cartika I, Ruswandi D (2015) Multivariate analysis of genetic diversity among some maize genotypes under maize-*Albizia* cropping system in Indonesia. Asian J Crop Sci 7:244–255

Tang S, Knapp SJ (2003) Microsatellites uncover extraordinary diversity in native American land races and wild populations of cultivated sunflower. Theor Appl Genet 106:990–1003

Tang S, Yu JK, Slabaugh MB, Shintani DK, Knapp SJ (2002) Simple sequence repeat map of the sunflower genome. Theor Appl Genet 105:1124–1136

Tang S, Kishore VK, Knapp SJ (2003) PCR-multiplexes for a genome-wide framework of simple sequence repeat marker loci in cultivated sunflower. Theor Appl Genet 107:6–19

Teich I, Verga A, Balzarini M (2014) Assessing spatial genetic structure from molecular marker data via principal component analysis: a case study in a *Prosopis* sp. forest. Adv Biosci Biotechnol 5:89–99

Tersac M, Blanchard P, Brunel D, Vincourt P (1994) Relations between heterosis and enzymatic polymorphism in populations of cultivated sunflowers (*Helianthus annuus* L.). Theor Appl Genet 88:49–55

Usatov AV, Klimenko AI, Azarin KV, Gorbachenko OF, Markin NV, Tikhobaeva VE, Kolosov YA, Usatova OA, Bakoev S, Makarenko M, Getmantseva L (2014) The relationship between heterosis and genetic distances based on SSR markers in *Helianthus annuus*. Ame J Agric Biol Sci 9(3):270–276

Van K, Kim DH, Shin JH, Lee SH (2011) Genomics of plant genetics resources: past, present, and future. Plant Genet Res 9(2):155–158

Vear F (2016) Changes in sunflower breeding over the last fifty years. OCL 23(2):D202

Warburton ML, Rauf S, Marek L, Hussain M, Ogunola O, de Jesus Sanchez Gonzalez J (2017) The use of crop wild relatives in maize and sunflower breeding. Crop Sci 57:1227–1240

Wegary D, Vivek B, Labuschagne M (2013) Association of parental genetic distance with heterosis and specific combining ability in quality protein maize. Euphytica 191(2):205–216

Wu K, Yang M, Liu H, Tao Y, Mei J, Zhao Y (2014) Genetic analysis and molecular characterization of Chinese sesame (*Sesamum indicum* L.) cultivars using insertion–deletion (InDel) and simple sequence repeat (SSR) markers. BMC Genet 15(35):1–15

Yalcin K, Goksel E, Sezgin D, Veli P, Tahir G (2007) Determining the relationships between yield and yield attributes in sunflower. Turk J Agric 31:237–244

Yan W (2002) Singular-value partition for biplot analysis of multi environment trial data. Agron J 94:990–996

Yan W, Rajčan I (2002) Biplot analysis of test sites and trait relations of soybean in Ontario. Crop Sci 42:11–20

Yao WH, Zhang YD, Kang MS, Chen HM, Liu L, Yu LJ, Fan XM (2013) Diallel analysis models: a comparison of certain genetic statistics. Crop Sci 53:1481–1490. https://doi.org/10.2135/cropsci2013.01.0027

Zambelli A, Leon A, Garces R (2015) Chapter 2: Mutagenesis in sunflower. In: Martínez-Force E, Dunford NT, Salas JJ (eds) Sunflower chemistry, production, processing, and utilization, pp 27–52. https://aocs.personifycloud.com/PersonifyEBusiness/Default.aspx?TabID=251&productId=2976692

Zimisuhara B, Valdiani A, Shaharaduddin NA, Quamaruzzaman F, Maziah M (2015) Structure and principal components analyses reveal an intervarietal fusion in Malaysian mistletoe fig (*Ficus deltoidea* Jack) populations. Int J Mol Sci 16:14369–14394