Indian J. Genet., 79(2) 444-452 (2019)
DOI: 10.31742/IJGPB.79.2.8

Agromorphological and molecular analysis discloses wide genetic variability in sunflower breeding lines from USDA, USA

K. T. Ramya, A. Vishnuvardhan Reddy and M. Sujatha*

ICAR-Indian Institute of Oilseeds Research, Rajendranagar, Hyderabad 500 030

(Received: December 2018; Revised: March 2019; Accepted: April 2019)

Abstract

The present study investigates genetic divergence among 84 fertility restorers and 32 cytoplasmic male sterile (CMS) lines of sunflower augmented from USDA, USA along with the popular Indian parental lines using simple sequence repeats (SSR). Thirty-nine polymorphic SSR primers produced 139 alleles with an average of 3.56 alleles per locus. The polymorphic information content ranged from 0.23 to 0.69 with an average of 0.45. The average genetic distance was 0.45 and 0.42 for the R and CMS lines, respectively. Dendrogram based on the dissimilarity coefficient matrix grouped the CMS and R lines into separate clusters except for Cluster A which consisted of all CMS lines along with five R lines. Genetic distance matrix estimated from three sets of mitochondrial primers (BOX, ERIC and REP) grouped the 32 CMS lines into eight clusters. The results suggest the existence of considerable genetic diversity among the restorer and CMS lines of sunflower obtained from USDA, USA.

Key words: Genetic diversity, *Helianthus annuus,* hybrid breeding, SSR

Introduction

The cultivated sunflower (*Helianthus annuus* L.) is an important oilseed crop and ranks next to soybean and groundnut in the total world production of oilseeds. Commercial single-cross hybrid sunflower in a hybrid seed industry has dramatically expanded since the first single-cross hybrid which was released in 1980 (Seetharam 1980). Large-scale cultivation of hybrids based on a single CMS source may lead to genetic vulnerability. Till date, 46 hybrids have been released from both public and private sectors but hybrid development in the public sector in India is restricted with the use of not more than 15 restorer lines and four CMS lines. The low genetic gain of hybrids over the last five decades is due to the use of limited breeding material which is also narrow in its genetic base. The availability of genetic diversity within breeding germplasm is important for maintaining a competitive edge in the commercial market. One of the practical uses of genetic diversity analysis is to describe heterotic groups and patterns (Laude and Carena 2015). Gentzbittel et al. (1994) used RFLPs to study the genetic relationships between sunflower inbred lines to determine unique restorer and maintainer germplasm pools. SSRs or microsatellites have also been successfully used in the determination of genetic distance in sunflower (Yu et al. 2002; Gvozdenoviæ et al. 2009; Smith et al. 2009).

Characterization of public sunflower inbred lines for genetic similarity will be useful for future sunflower germplasm enhancement and utilization across the globe. In sunflower, there are more than 72 sources of cytoplasmic genic male sterile (CMS) sources. However, due to the poor availability of the effective restorers, the other CMS sources are not widely used. Sunflower being an introduced crop to India, the breeding program has to rely on foreign gene banks to broaden the genetic base. A continuous effort is being made for germplasm augmentation with the cooperation from USDA-ARS, Ames, USA. Germplasm procured from USDA was tested for adaptability under the local environment and to identify suitable lines for initiating the crossing programme. Therefore, the objectives of the present study were 1) to select parental lines for developing high yielding hybrids from the germplasm received from USDA, USA, 2) to study diversity among the germplasm to classify genotypes and describe heterotic groups and patterns by a set of SSR primers, and 3) to analyse the extent of genetic diversity of CMS lines by rep-PCR.

*Corresponding author's e-mail: mulpurisujata@yahoo.com
Published by the Indian Society of Genetics & Plant Breeding, A-Block, F2, First Floor, NASC Complex, IARI P.O., Pusa Campus, New Delhi 110 012; Online management by www.isgpb.org; indianjournals.com

Materials and methods

Plant materials

A set of 84 fertility restorers and 32 CMS lines were used in the present study. Among the 84 lines, 78 were received from Sunflower Research Unit, USDA-ARS, Ames, USA and six were the locally developed indigenous R lines (Table 1). Similarly, among 32 CMS

Table 1. R lines and inbreds procured from USDA ARS Ames USA

S.No.	R Line	Pedigree
1	RHA276	CMS PI343765/HA119//HA62-4-5/2/ T66006-2-2-11-3-2
2	RHA278	CMS PI343765/HA119//HA62-4-5/2/ T66006-2-2-11-3-2
3	RHA293	3*Commander /Mennonite RR
4	RHA309	RHA 293 Selection
5	RHA328	RHA 294 /3/ R811 // HA 292 / RHA 282
6	RHA329	RHA 298 and 299/NC
7	RHA330	RHA 298 and 299/NC
8	RHA389	C2 RFSS (DMRRS)
9	RHA401	RHA 274/RHA 1185-14-3-4-1
10	RHA408	ROMANIA R-LINE SCL POP-1
11	RHA282	Boneta Giant Manchurian/Mennonite RR
12	RHA294	CMS PI 343863/Bonita Giant-Manchurian/HA61/Mennonite RR
13	RHA295	CMS Mennonite RR- 18-1*3/T66006-2
14	RHA409	ROMANIA R-LINE SCL REC CYCLE C2
15	RHA299	CMS P-21 VRI/T70050 T70050=PI 343765/HA119//HA62-4-5/2/T66008-2
16	RHA419	RHA 373 / ARG 1575-2
17	RHA427	RHA 409/R 376*2/*H. annuus* IMI resistant
18	RHA325	R811-3
19	RHA326	RHA 293/RHA 294
20	RHA327	RHA 294/3/R811//Jumbo Israel/R268
21	RHA331	RHA 298 and 299/ NC
22	RHA332	RHA 298 and 299/ NC
23	RHA333	RHA 298 and 299/ NC
24	RHA334	RHA 298 and 299/ NC
25	RHA340	HA 89 **H. argophyllus* 415
26	RHA344	RHA 274 *2/PHO
27	RHA345	RHA 274 *2/PHO
28	RHA 346	RHA 274 *2/PHO
29	RHA 347	RHA 274 *2/PHO
30	RHA 348	RHA 274 *2/PHO
31	RHA354	RHA 310 *2/PHO
32	RHA355	RHA 310 *2/PHO
33	RHA358	RHA 274 *3/DDR Short height
34	RHA359	RHA 274/DDR
35	RHA436	RHA340 / RHA344 High Oleic
36	RHA437	RHA340 / RHA344 High Oleic
37	RHA364	RHA418/RHA 419/3/ RHA801// RHA365/PI 413047
38	RHA365	SELECT
39	RHA373	RHA 274/82ROM-R31
40	RHA374	ARG-R43
41	RHA376	RHA 296/RHA 266
42	RHA377	RHA 299//SOREM HT 58/RHA 801
43	RHA381	USDA 1869-3/*3 RHA 274
44	RHA386	82 ROM. R-LINE BULK
45	RHA387	RHA 274/83 ROM. R-LINE B.
46	RHA388	RHA 274/FELIX
47	RHA395	C4 RFSS (DMRRS)
48	RHA454	RHA 447//RHA440/PSC 8 (High ol.)
49	RHA398	RHA 274/BCD LINE BULK
50	RHA399	RHA 274/ODESSKIJ 91
51	RHA400	AUSTRALIA 85 R-LINE POP.
52	RHA415	RHA 274/2696-1 (HIGH LINOL.)
53	RHA416	RHA 274/2696-1 (HIGH LINOL.)
54	RHA418	RHA 801/RHA 274 // MYHOCO H-9
55	RHA420	RHA 373 / ARG 1575-2
56	RHA428	RHA 801 // RHA 365 / PI 413157
57	RHA438	RHA340 / RHA344 High Oleic
58	RHA439	RHA377/AS3211 Sclerotinia Tolerant
59	RHA440	RHA377/AS4379 Sclerotinia Tolerant
60	RHA443	RHA426/RHA419//RHA440*
61	RHA447	RHA377/RHA348 High Oleic
62	RHA450	RHA324/Primrose
63	RHA461	RHA418/RO 12-13//RHA 274/Dobritch
64	RHA462	RHA439/IS PH RES.
65	RHA463	RHA440/PSC 8
66	RHA455	RHA440/HO IS R-line (High oleic)
67	RHA854	RHA 273 Selection
68	RHA855	CMS HA 89/RHA 273
69	RHA857	S310/RHA 297
70	RHA858	P1161/RHA 298
71	RHA859	NSH 43/RHA 299
72	ID-LRLYC	Late R-line Yield Composited (33 S4 families)
73	ID-ERLYC	Early R-line Yield Composited (39 S3 families)
74	NDBR1	Impira INTA white *2/3SP/2/HA89
75	NDBR2	Impira INTA white *2/3SP/2/HA89.
76	ND-ERLYS	Early R-line Yield Synthetic
77	RHA360	RHA 274/Donskoi
78	P 93-R	Selection from GP-2-378-5
79	RHA 6-D-1	Selection from GPR-102
80	GKVK-1	NA
81	RHA-1-1	Selection from MRHA-1
82	RHA 271	CMS PI343765/HA119//HA62-4-5 /2/ T66006-2-1-31-1=T70020
83	PS 2056	*H.annus /H.petiolaris//H.annus*
84	RHA-95-C-1	Selection from GPR-102

Source for pedigree: https://www.ag.ndsu.edu/fss/ndsu-varieties/usda-sunflower-inbred-lines

* = Imidazolinone Herbicide Resistant; NC = Non oilseed composite; PHO = Pervents high oleic

lines, 28 were received from USDA-USA and four were the locally developed CMS lines from India (Table 2). Seed multiplication of augmented lines was carried out during April 2015 at the experimental farm of ICAR-Indian Institute of Oilseeds Research (ICAR-IIOR), Hyderabad, India. Restorer and CMS lines along with their maintainers were evaluated over three seasons *viz*., rainy season of 2015, 2016 and winter of 2016 at the experimental farm, ICAR-IIOR, Hyderabad. Augmented design with four checks, namely, P 93-R, RHA6-D-1, GKVK-1 and RHA-1-1 for R lines and two checks for CMS lines, *viz.,* HA234 and ARM243A, was used in the experiment. Standard agronomic practices were followed for the maintenance of the crop.

Table 2. CMS lines procured from USDA

S.No.	CMS Line	Pedigree/Identifiers
1	HA65	Ames 3959
2	HA89	VNIIMK 8931 Sel
3	HA112	Ames 3967
4	HA115	Ames 3971
5	HA116	Ames 3974
6	HA208	Ames 3986
7	HA224	Armavirski 9345(P.I.265102) /HA8
8	HA228	Ames 3992
9	HA234	2*Smena//HA6/HA8
10	HA236	Ames 3997
11	HA248	Ames 4003
12	HA249	Ames 4005
13	HA250	Ames 4007
14	HA253	Ames 4009
15	HA259	Ames 4011
16	HA286	Isreal Selection
17	HA288	Mennonite RR Seletio
18	HA291	INRA 6501 Sel
19	HA302	Peredovik 304 (P.I. 372173) Sel
20	HA303	Voshod (P.I. 371936) Sel
21	HA232	2*Smena//HA6/HA8
22	HA292	3*Commander / Mennonite RR
23	HA64	VNIIMK 1646 Sel = cm 316
24	HA234*	2*Smena//HA6/HA8
25	ARM243	Armavirski Selection
26	2023A	*H. agrophyllus/H. annus* (cv. Morden)//Morden
27	HA89A-1*	VNIIMK 8931 Sel
28	HA300	Peredovik 301 (P.I.372172) sel
29	243A	Ames 3999
30	430A	HA821*2/ /CMS HA89*2/*H. paradoxus* (PAR1673)
31	124A	VNIIMK 8883
32	133A	Ames 3979

Source: https://www.ag.ndsu.edu/fss/ndsu-varieties/usda-sunflower-inbred-lines
* = Maintained at IIOR Hyderbad

SSR analysis

For each line, leaves were harvested from five field-grown six-week-old plants and bulked for DNA isolation. DNA was extracted following the CTAB method (Webb and Knapp 1990) and quantified on 0.8% agarose gel electrophoresis using a known concentration of DNA as standard. Using the information on marker distribution on a genetic linkage map, 39 SSR primers of the ORS series (Tang et al. 2003) were selected. Polymerase chain reaction (PCR) was performed according to Yu et al. (2002). The reaction mix (20 µl) contained 20 ng of DNA template, 5.5 pmol of each primer, 2.5 mM $MgCl_2$, 125 mM of each dNTP, 1X PCR buffer, and 0.8 U of *Taq* DNA polymerase (Genei, Bangalore, India). The PCR amplification profile included an initial denaturation step at 94°C for 2 min, followed by 45 cycles of denaturation at 94°C for 30 s, annealing for 30 s, and extension at 72°C for 1 min, with a final extension at 72°C for 5 min in a GeneAmp PCR System 9700 thermocycler (Perkin Elmer – Applied Biosystems). Annealing temperature varied between 51 and 60°C for the primers used. The PCR-amplified products were run on 3.5% agarose gel electrophoresis (Genei, Bangalore, India) in 1X Tris-acetate EDTA buffer with ethidium bromide (50 ng/ml) for 2 h at 70 V for detection of allelic variations. A 50/100 bp ladder was used for determining the allele size.

Rep-PCR analysis

PCR conditions were followed as described by Dinesh Kumar and Nizampatnam (2013). PCR amplification was carried out in a 20 µl reaction mixture that contained 10 mM Tris-HCl (pH 8.3), 0.4 mM of each dNTP, 10 pM of each primer, 1.5 U *Taq* DNA polymerase (Genei, Bangalore, India) and 20 ng of genomic DNA. The PCR conditions were 94°C for 3 min, followed by 45 cycles of DNA amplification [20 s at 92°C, 1 min at 52°C for BOX and ERIC primers (1 min at 38°C for REP primers) and 8 min at 68°C] and a 15 min incubation at 68°C. Amplified products were separated by electrophoresis on 2% agarose gel for 7-8 h at

constant voltage (2 V/cm). The amplicons were visualized under UV light after staining with 0.01% ethidium bromide. The band size of the three BOX, ERIC and rep-PCR amplicons were determined by the *Hin*dIII and *Eco*RI digest of λ-phage DNA, 100 bp marker. All clearly visible, polymorphic bands were scored for the band size. All PCR reactions were replicated to confirm the consistency of results.

Data analysis

Presence or absence of a fragment was coded in a binary data matrix as 1 or 0, respectively. Pair-wise genetic similarities (S_{ij}) between accessions i and j were estimated using the similarity coefficient of Nei and Li (1979). The PIC values as a marker discrimination power were calculated according to Anderson et al. (1993).

Dissimilarity matrix was used for clustering of genotypes using selected SSR markers, based on unweighted neighbour joining method using Powermarker software (version 3.25, 2014). For scoring and analysis of data from the Rep-PCR analysis (BOX, ERIC and REP primers), the total number of monomorphic and polymorphic bands which were clear, unambiguous and reproducible were scored for the tested primers. Data scoring was carried out using a binary number system for '1' as the presence and '0' as the absence of fragment (band) for both primers. A measure of similarity among 32 CMS lines was established as matrices of genetic similarity compiled using the SIMQUAL function Jaccard's coefficients. Dendrograms representing the genetic relationships among all CMS lines were generated from the similarity matrices by applying unweighted pair-group arithmetic mean method (UPGMA) (cluster analysis) with the SAHN function system. Dendrograms were visualized on MEGA Version 6 (2013) after bootstrapping using UPGMA. The dissimilarity matrix based on SSR primers and rep-PCR primers was compared following Mantel (1967) test using NTSYSpc Ver. 2.20 program (Sneath and Sokol 1973).

Results and discussion

Trait variability

The crop establishment of R lines and CMS lines received from USDA ARS, Ames, USA was good under the Indian conditions. The mean data of three

Table 3. Cluster-wise phenotypic traits such as days for 50% flowering (DF), days to maturity (DM), plant height (PH cm), Head diameter (HD cm), Seed yield/plant (PY), Test weight (TW 100 seed wt. (g) and Oil content (OC) in percentage for R lines

	R1 (n=5)			R2 (n=16)			R3 (n=20)			R4 (n=8)			R5 (n=7)			R6 (n=6)			R7 n=17)			Overall (n=84)		
Variable	Mean	SE	Range	Mean	SE	Range	Mean	SE	Range	Mean	SE	Range	Mean	SE	Range	Mean	SE	Range	Mean	SE	Range	Mean	SE	Range
DF	61.3	3.1	57-73.3	59	1.3	47.7-67.7	65	1	58.7-74.3	61.2	1.9	53.7-68.3	64.8	1.3	59-69.7	56.8	1.7	52.7-62.7	63.2	1	54.3-70.3	62.2	0.58	47.7-74.3
DM	89.2	2.2	85.3-98	88.2	1.8	71-100	94.3	1.2	86.7-106.3	90.5	1.8	84.3-97.3	93.9	1.9	87-102.7	86	1.3	82.7-90	91.9	1.1	83.3-99.3	91.3	0.64	71-106
PH (cm)	104.6	7.7	89.5-133	82.8	6.3	36.8-121.7	104.7	4.7	66.7-138	83.6	2.9	69.7-97	105.6	7.1	90.1-132.4	80.8	7.1	54.8-104.3	93	4.4	53.5-140	94.5	2.34	36.8-140
HD	8	1.1	5.5-10.8	7.3	1.4	5.5-10.8	9.1	2.5	4.4-56	6	0.3	4.8-7.5	6.5	0.7	5.4-10.3	5.9	0.3	5-6.9	6.9	0.3	4.7-9.9	7.56	0.65	4.4-10.8
PY	8.3	1.4	6.4-13.9	4.5	0.6	1.51-0.5	4.5	0.5	1.4-11.1	3.3	0.5	2.3-5.6	4.2	1.1	1.4-10.1	3.8	0.5	2.5-6.2	3.2	0.3	1.4-6.1	4.27	0.25	1.37-13.9
TW	3.2	0.8	1.6-5.2	2.1	0.3	0.8-4.2	1.7	0.2	0.8-3.7	1.8	0.3	1-3.2	2.7	0.4	1.4-4	2.7	0.2	2.2-3.8	2.4	0.2	1-4.3	2.21	0.1	0.8-5.2
OC	35.3	1.1	30.1-43.1	32.1	0.6	25.8-40	29.5	0.6	25.9-34.9	32.3	0.9	29.6-35.6	33.2	1.2	29.2-37.6	30.2	1.1	25.9-34.2	29.4	0.8	24.6-35.7	31.1	1.35	24.6-43.1

n = Number of lines in cluster; SE = Standard error; Min = minimum; Max = Maximum

Table 4. Cluster-wise phenotypic traits such as days for 50% flowering (DF) days to maturity (DM) plant height (PH cm) Head diameter (HD cm) for CMS lines

	A1 (n=15)			A2 (n=4)			A3 (n=13)			Overall		
Variable	Mean	SE	Range	Mean	SE	Range	Mean	SE	Range	Mean	SE	Range
DF	66.4	2.2	58-83	67.3	3.3	60-76	65.2	1.2	60-74	66.0	1.1	58.0-83
DM	96.4	2.2	88-113	97.3	3.3	90-106	95.3	1.2	90-104	96.0	1.1	88.0-113
PH	102.4	6.9	67-128	91.4	7.5	70-105	105.3	5.3	68-131	102.2	3.8	67.0-131
HD	13.3	0.4	11.2-15.8	13.0	1.0	10.4-14.8	12.0	0.3	10-13.8	12.7	0.3	10.0-15.8

seasons are presented in Tables 3 and 4. The medium duration hybrids are most preferred among the Indian sunflower farmers. The genotype which matures within 85-100 days (with 55 to 70 days for flowering) is considered as a medium duration genotype. The duration for anthesis ranged from 7 to 14 days with an average of 8 days. Monohead (RHA436, RHA376, RHA428 and RHA95-C-1) and multihead (RHA276, RHA346, RHA436, RHA386, RHA395, RHA454 and PS 2056) lines showed shorter flowering duration. RHA418 showed longest flowering duration of 14 days. Superior agronomic performance, in terms of water use efficiency, photosynthetic rate and seed yield, was observed for the R lines RHA326, RHA344, RHA345, RHA346, RHA355, RHA373, RHA386, RHA387, RHA388, RHA395, RHA398, RHA416, RHA418 and RHA857. Three CMS lines *viz*., CMS HA228, CMS HA248 and CMS HA202 showed better performance for water use efficiency and seed yield per plant (Anonymous 2018).

Molecular analysis

SSR and SNP primers were used to study the phylogenetic relationship between wild and cultivated sunflower accessions (Tang and Knapp 2003; Kolkman et al. 2007). These studies have revealed that genetic diversity has been reduced due to domestication and selection in sunflower. In the present study, a set of 139 alleles were detected across 84 restorer lines at 39 SSR loci (Table 5). Allele frequencies per locus were not homogeneous for most of the markers and at each locus one allele was predominantly observed. The polymorphic information content values, a reflection of allele diversity and frequency among the restorer lines were high for all the SSR loci tested except for ORS 381. Overall genetic distance ranged from 0.33 to 0.66 with an average of 0.50. An average genetic distance within R lines was 0.49 with the range of 0.19 to 0.59. Similarly, an average genetic distance within CMS lines was 0.42 with the range of 0.33 to 0.53. The average genetic distance between CMS lines and R lines was 0.56 with a maximum of 0.86 and a minimum of 0.19. Highest genetic distance (0.86) was between CMS 430A and RHA857, RHA359 and ID-LRLYC; CMS 243A and RHA388, CMS 250A and RHA859, RHA443. The average genetic distance between CMS and locally developed R lines from Indian source was 0.61. Indigenously developed R line, GKVK-1 showed maximum genetic distance with CMS lines ARM243A, HA116A, HA248A and HA249A with an average genetic distance of 0.63. Similarly, P 93R-1 showed maximum genetic distance (0.76) with HA133A, HA430A, HA253A and ARM243A with an average genetic distance of 0.59. Popular local R line RHA6D-1 showed maximum genetic distance (0.80) with CMS lines HA133A, HA430A, ARM243A, HA208A and HA302A with an average distance of 0.59. The genetic distance between RHA1-1 and CMS lines HA133A, HA2023A, HA430A and ARM243A was also maximum (0.80) with an average of 0.59 for the 32 CMS lines. Similarly, locally developed CMS lines ARM243A and 2023A had maximum genetic distance with RHA1-1, RHA364 and RHA857.

Dendrogram constructed based on the dissimilarity coefficient matrix, grouped the CMS and R lines into separate clusters (Fig. 1). Cluster A consisted of all CMS lines along with five R lines. Nevertheless, the CMS lines were further divided distinctly into three subgroups, namely, A1, A2 and A3. Cluster A1 consisted of 15 lines, cluster A2 included four lines and A3 had 13 CMS lines. The CMS lines 2023A and 430A were grouped in cluster A2 and ARM243A in A3. Locally maintained CMS line HA89A for 40 years and recently augmented USDA line HA89A differed at three loci and were grouped in cluster A1. CMS line HA208A was genetically distant from other counterparts. Three confectionary type CMS lines splintered into a small sub-group in cluster A1. Phenotypic data of CMS lines in each cluster are

Table 5. Thirty-nine sunflower simple sequence repeat (SSR) markers exhibiting linkage groups major allele frequency number of alleles gene diversity and polymorphic information content

Marker	LG	Major allele frequency	Allele no.	Gene diversity	PIC
ORS610	1	0.52	6	0.64	0.58
ORS371	1	0.60	2	0.48	0.36
ORS837	1	0.72	3	0.43	0.38
ORS229	2	0.50	3	0.52	0.40
ORS1065	2	0.52	4	0.52	0.40
ORS1036	3	0.74	2	0.38	0.31
ORS924	3	0.77	3	0.36	0.30
ORS695	4	0.57	3	0.50	0.38
ORS785	4	0.72	4	0.45	0.42
ORS505	5	0.61	6	0.57	0.52
ORS533	5	0.49	7	0.68	0.63
ORS774	5	0.46	3	0.61	0.53
ORS381	6	0.85	3	0.25	0.23
ORS966	7	0.53	3	0.61	0.54
ORS400	7	0.72	3	0.43	0.39
ORS456	8	0.28	4	0.74	0.69
ORS1161	8	0.46	3	0.64	0.57
ORS299	8	0.72	4	0.45	0.42
ORS844	9	0.33	5	0.73	0.68
ORS887	9	0.62	3	0.48	0.37
ORS437	10	0.36	3	0.66	0.59
ORS878	10	0.59	3	0.50	0.40
ORS1146	11	0.52	2	0.50	0.37
ORS607	11	0.72	4	0.44	0.41
ORS761	12	0.72	3	0.44	0.39
ORS778	12	0.72	3	0.41	0.35
ORS656	12	0.72	3	0.43	0.38
ORS534	13	0.65	3	0.46	0.36
ORS707	13	0.75	3	0.38	0.31
ORS694	14	0.41	4	0.69	0.63
ORS1180	14	0.44	4	0.66	0.60
ORS254	15	0.53	2	0.50	0.37
ORS857	15	0.54	2	0.50	0.37
ORS996	16	0.61	3	0.54	0.47
ORS807	16	0.60	3	0.49	0.39
ORS303	16	0.72	3	0.41	0.35
ORS561	17	0.44	6	0.70	0.66
ORS735	17	0.45	4	0.64	0.56
ORS297	17	0.44	7	0.68	0.63
Mean		0.58	3.56	0.53	0.45

summarized in Table 4. The 84 restorer lines were divided into seven clusters (Fig. 1). Clusters, R2, R3 and R7 were large groups with 16, 20 and 17 lines, respectively. Five restorer lines were grouped along with CMS lines and included RHA340 in cluster A1; RHA855, RHA276 and RHA278 in A2 and RHA299 in A3. However, the R line PS 2056 developed through interspecific hybridization between *H. petiolaris* and cultivated *H. annuus*, was grouped in a different cluster and the distance was also high up to 0.53 with other R lines of Indian source. In this study, 13 confectionery R lines were clustered in R6 and R7. Phenotypic data of R lines in each cluster are summarized in Table 3. Cluster R6 consisted of early flowering and maturing lines, while R3 consisted of late flowering and maturing accessions.

The present study included R lines and CMS lines of popular hybrids from India, USA and other countries along with five trait specific inbreds. Unfortunately, a limited number of SSR primers were used on a large set of genotypes in this study. There were no unique alleles identified for the lines studied. The genetic similarity among 25 oilseed type sunflower lines ranged from 0.58 to 0.98 with an average of 0.70 (Rönicke et al. 2005), and genetic similarity among 24 oilseed type sunflower lines ranged from 0.70 to 0.91 at the AFLP marker loci investigated (Hongtrakul et al. 1997). The average genetic distance was high in the present study (0.52) than that in the previous studies (Hongtrakul et al. 1997; Yue et al. 2009). The average PIC value is equivalent to genetic diversity estimated as a measure of genetic variation by the formula of Weir (1996). Generally, genetic dissimilarity coefficients agree well with the pedigree data. For example, the dissimilarity coefficient was highest between RHA330 and RHA355 which had different ancestry. The dissimilarity coefficient between RHA415 and RHA416 was the lowest, and they were directly selected from the same cross, RHA274/2696-1 (HIGH LINOL.). The SSR primers used in the study successfully differentiated confectionery and oil type restorer lines. All the 16 confectionery type lines were grouped into separate clusters (R6 and R7). Genetic distance among the confectionery type lines was lower than the oil type restorer lines. The results of the current study indicated that substantial genetic variability has been captured and maintained in this valuable public inbred line collection agreeing with the report of Yue et al. (2009). Manivannan et al. (2010) assessed the genetic diversity of 26 CMS lines involving various cytoplasmic sources using 20 SSR

Fig. 1. Dendrogram of 116 sunflower restorer and CMS lines along with Indian lines based on allelic variation determined by 39 SSR primers

different groups. Hybridization between CMS lines and R lines can be planned using the dendrogram for production of commercial hybrids. Genetic diversity plays an important role in the choice of parents because hybrids between lines of diverse origin generally display greater heterosis than those between closely related parents (Cheres and Knapp 1998). Cheres et al. (2000) demonstrated strong correlation between genetic distance and seed yield using AFLP fingerprints. Bernardo (1992) showed that genetic distance does not accurately predict hybrid performance unless the DNA markers used in the analysis are associated with (linked to) genes affecting the trait. Therefore, SSR markers linked to yield-related traits can be used to predict the performance of the hybrids. SSRs are a powerful tool for fingerprinting inbred lines, markers and found very low diversity among commercially used CMS lines. Commercial sunflower hybrids are produced by utilizing heterosis between the CMS lines and fertility restorer lines. Thus, assigning public CMS or R lines to different heterotic groups assumes priority for sunflower breeders. A heterotic group is a collection of closely related inbred lines. The co-ancestries within a heterotic group are usually high, whereas the co-ancestries between two heterotic groups comprising a heterotic pattern are usually low (Laude and Carena, 2015). Clustering based on molecular profiles is in agreement with the pedigree data, for example, lines that originated from the same cross were grouped in a single cluster (R6). It is also interesting to note that, the highly similar inbreds which are either the progenies of the same crosses or directly selected from the other inbred lines are the closest pairs in the dendrogram. These similar pairs of inbred lines are quite useful in genetic studies on mapping and cloning genes governing traits with differences between the lines. However, some inbred lines with the same ancestries were clustered into producing genetic maps, and marker-assisted selection in sunflower (Tang et al. 2002; Yu et al. 2002).

BOX, ERIC and REP primer analysis of male sterile lines

Number of scorable amplicons was 17, 16 and 10 for BOX, ERIC and REP primers, respectively. Average polymorphic information content was 0.48 while it was 0.49 for BOX and ERIC and 0.46 for REP primers (Table 6). The three sets of primers (BOX, ERIC and

Table 6. Allele information on 32 CMS lines using *Rep Box* and *Eric* primers

Primer	Major allele frequency	Allele number	Gene diversity	PIC
BOX	0.06	28.00	0.96	0.49
ERIC	0.09	29.00	0.96	0.49
REP	0.19	14.00	0.88	0.46
Mean	0.11	23.67	0.94	0.48

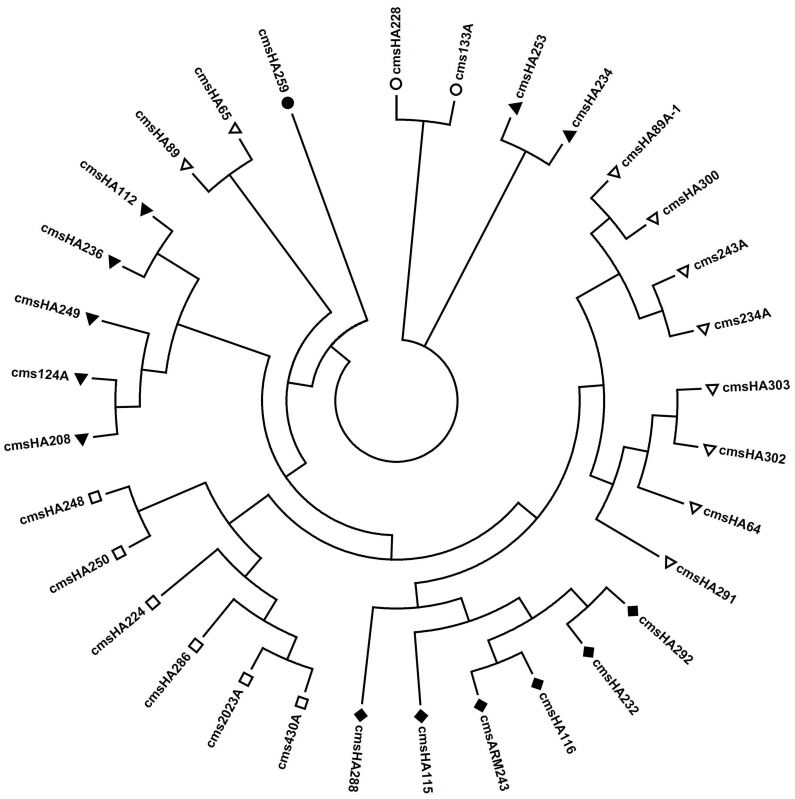

Fig. 2. Dendrogram of 28 sunflower male sterile lines along with four Indian lines based on allelic variation determined by mitochondrial based Eric Box and rep primers

REP*)* produced eight clusters for 32 CMS lines (Fig. 2). Cluster numbers 1, 2 and 7 consisted of only two CMS lines, while CMS HA259 formed a separate cluster. All the three primers resulted in distinct banding patterns differentiating CMS HA89 of USDA from CMS HA89-1 which was maintained under Indian conditions. BOX, ERIC and REP primers are conserved sequences found in mitochondrial genome, therefore rep-PCR fingerprinting is robust technology for fingerprinting of cytoplasmic male sterile lines. Dinesh Kumar and Nizampatnam (2013) tested rep-PCR using three sets of commonly used primers, BOX, ERIC and REP, and both inter-specific and intra-specific mitochondrial genome diversities in *Carthamus* were identified. Horn (2001) studied the organisation of mtDNA for 28 sources of cytoplasmic male sterility (CMS) and a fertile line (normal cytoplasm) of sunflower by Southern hybridization. Clustering using rep-PCR was not in accordance with pedigree data, unlike SSR primer based clustering. Seven CMS lines were not clustered into any cluster. Furthermore, the correlation coefficient (r=0.009; P=0.842) for the two clustering matrices tested by Mantel's test showed an insignificant correlation between genetic distances derived using SSR primers and rep-PCR primers. SSR primers and rep-PCR primers (BOX, ERIC and REP primers) clustered CMS lines in a different manner.

In summary, the present study assessed the genetic diversity of 116 publicly released germplasm of USDA-ARS and India. There is considerable genetic diversity among the exotic R lines and CMS lines. Seventy-eight R lines and 28 CMS lines of popular public hybrids of USDA-ARS along with respective maintainers have established well to Indian environment. The information about the genetic diversity and relationships among breeding lines and varieties is not only useful for germplasm conservation and inbred line identification, but also for the selection of parental lines for hybrid breeding in crops, including sunflower (Senior et al*.* 1998; Meena et al. 2013). Seven R lines for early flowering and 15 lines for early maturity were identified. Three lines were identified for high oil content (>40%) under Indian conditions. These lines can be utilized for the development of hybrids with diverse genetic sources. The USDA lines which are agronomically well-adopted lines in Indian environment condition serve as a new source of diversity and can accelerate hybrid development not only in India but also in other sunflower growing countries.

Authors' contribution

Conceptualization of research (MS, RKT); Designing of the experiments (RKT); Contribution of experimental materials (RKT, MS); Execution of field/lab experiments and data collection (RKT, MS); Analysis of data and interpretation (RKT, MS, AV); Preparation of manuscript (RKT, MS, AV).

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

The authors thank Dr. Laura Fredrick Marek, USDA-ARS-MWA-PIRU; Iowa State University, USA, for providing the seed materials.

References

Anderson J. A., Churchil G. A., Autrique J. E., Tanksley S. O. and Sorrels M. E. 1996. Optimizing parent selection for genetic linkage maps. Genome, **36**: 181-186.

Anonymous 2018. Sunflower Annual Report. ICAR-Indian Institute of Oilseeds Research, Hyderabad pp 210.

Cheres M. T. and Knapp S. J. 1998. Ancestral origins and genetic diversity of cultivated sunflower: coancestry analysis of public germplasm. Crop Sci., **38**: 1476-1482.

Cheres M. T., Miller J. F., Crane J. M. and Knapp S. J. 2000. Genetic distance as a predictor of heterosis and hybrid performance within and between heterotic groups in sunflower. Theor. Appl. Genet., **100**: 889-894.

Dinesh Kumar V. and Nizampatnam N. R. 2013. Rep-PCR could identify both inter and intra specific mitochondrial genome differences in the genus *Carthamus*. Plant Mol. Biol. Rep., **31**: 1150-1156.

Dong G. J., Liu G. S. and Li K. F. 2007. Studying genetic diversity in the core germplasm of confection sunflower (*Helianthus annuus* L.) in China based on AFLP and morphological analysis. Russ. J. Genet., **43**:762-770.

Dwivedi K. K., Radhakishna A., Kumar S., Srivastava M. K., Gupta M. G. and Kaushal P. 2015. Development of an ISSR-derived SCAR marker linked to apospory in buffel grass (*Cenchrus ciliaris* L.). Indian J. Genet., **75**: 271-273.

FAOSTAT 2016. http://www.fao.org/faostat/en/#data/QC

Gentzbittel L., Zhang Y. X., Vear F., Griveau B. and Nicolas P. 1994. RFLP studies of genetic relationships among inbred lines of the cultivated sunflower, *Helianthus annuus* L.: Evidence for distinct restorer and maintainer germplasm pools. Theor. Appl. Genet., **89**: 419-425.

GOI. 2016. Agricultural Statistics at a Glance, Directorate of Economics and Statistics, Department of Agriculture and Cooperation, Ministry of Agriculture, Government of India.

Gvozdenoviæ S., Saftiæ-Pankoviæ D., Jociæ S. and Radiæ V. 2009. Correlation between heterosis and genetic distance based on SSR markers in sunflower (*Helianthus annuus* L.). J. Agril. Sci., **54**: 1-10.

Hongtrakul V., Huestis G. M. and Knapp S. J. 1997. Amplified fragment length polymorphisms as a tool for DNA fingerprinting sunflower germplasm: Genetic diversity among oilseed inbred lines. Theor. Appl. Genet., **95**: 400-407.

Horn R. 2001. Molecular diversity of male sterility inducing and male fertile cytoplasms in the genus *Helianthus.* Theor. Appl. Genet., **177**: 457-468.

Kolkman J. M., Berry S. T., Leon A. J., Slabaugh M. B., Tang S., Gao W., Shintani D. K., Burke J. M. and Knapp S. J. 2007. Single nucleotide polymorphisms and linkage disequilibrium in sunflower. Genetics, **177**: 457-468.

Laude T. P. and Carena M. J. 2015. Genetic diversity and heterotic grouping of tropical and temperate maize populations adapted to the Northern US Corn Belt. Euphytica, **204**: 661-677.

Manivannan N., Vijaya B. A. and Vindhiya V. P. 2011. Genetic diversity analysis in male sterile lines of sunflower by SSR markers. Indian J. Genet. Plant Breed., **71**: 276-278.

Mantel N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res*.,* **27**: 209-220.

Meena H. P., Sujatha M. and Varaprasad K. S. 2013. Achievements and bottlenecks of heterosis breeding of sunflower (*Helianthus annuus* L.). India. Indian J. Genet., **73**: 123-130.

Rohlf F. J. 2005. NTSYS-pc: numerical taxonomy and multivariate analysis system Version 2.2. Exeter Software: Setauket NY.

Sarada C., Alivelu K., Sambasiva Rao V., Sudakara Babu S. N. and Varaprasad K. S. 2015. Oilseeds Statistics a Compendium. ICAR-Indian Institute of Oilseeds Research Hyderabad India. Pp 165.

Seetharam A. 1980. Hybrid sunflowers. Indian Farming, **24**: 15.

Senior M. L., Murphy J. P., Goodman M. M. and Stuber C. W. 1998. Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. Crop. Sci., **38**: 1088-1098.

Smith J. S. C., Hoeft E., Cole G., Lu H., Jones E. S., Wall S. J. and Berry D. A. 2009. Genetic diversity among sunflower inbreds and hybrids: assessing probability of ancestry and potential for use in plant variety protection. Crop Sci., **49**: 1295-1303.

Sneath P. H. A. and Sokol R. R. 1973. Numerical Taxonomy the Principles and Practice of Numerical Classification. WH Freeman and Co. San Francisco.

Tang S. and Knapp S. J. 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 J. K., Slabaugh M. B., Shintani D. K. and Knapp S. J. 2002. Simple sequence repeat map of the sunflower genome. Theor. Appl. Genet., **105**: 1124-1136.

Webb D. M. and Knapp S. J. 1990. DNA extraction from a previously recalcitrant plant genus. Plant Mol. Biol. Rep., **8**: 180-185.

Weir B. S. 1996. Genetic Data Analysis II. Sinauer Associates Inc Sunderland MA.

Yu J., Mangor J., Thompson L., Edwards K. J., Slabaugh M. B. and Knapp S. J. 2002. Allelic diversity of simple sequence repeats among elite inbred lines of cultivated sunflower. Genome, **45**: 652-660.

Yue B., Cai X., Vick B. A. and Hu J. 2009. Genetic diversity and relationships among 177 public sunflower inbred lines assessed by TRAP markers. Crop Sci., **49**: 1242-1249.