INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
20–0038/2020/24–1–51–58
DOI: 10.17957/IJAB/15.1407
http://www.fspublishers.org

Full Length Article

Population Structure, Linkage Disequilibrium and Association Mapping Study through SSR Markers in Sunflower (*Helianthus annuus*)

Danish Ibrar[1*], Rafiq Ahmad[2], Muhammad Ayub Khan[1], Asif Javaid[1], Zubair Ahmed[1], Munir Ahmad[2] and Talat Mahmood[2]

[1]*Pakistan Agricultural Research Council, Islamabad, Pakistan*
[2]*Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan*
[*]For correspondence: danish.uaar@gmail.com
Received 07 January 2020; Accepted 01 February 2020; Published 20 April 2020

Abstract

Sunflower is one of the most important oilseed crops of world. Exploitation of heterosis has been very successful in cross pollinated crops and sunflower being cross pollinated is ideal for hybrid breeding. Exploitation of hybrid vigor with precision saves time and resources, and this can be achieved by using new genetic based crop improvement techniques like association mapping. Association mapping has proved very successful in designing molecular marker based speedy crop improvement. Agro-morphological traits are important for sunflower breeders in selection of genotypes with high performance and other yield contributing traits. Ten agro-morphological characters *i.e.*, days to flower initiation and completion, plant height, stem curvature, head diameter, number of leaves per plant, leaf area, hundred seed weight, seed yield per plant and oil contents were used to study the genetic diversity in 109 sunflower lines at morphological level. A high genetic variability was observed among sunflower lines in field conditions. Population structure and linkage disequilibrium was estimated after genotyping through 40 SSR (simple sequence repeats) markers, that amplified a total of 65 DNA bands to locate the QTLs (quantitative trait loci) controlling the important yield contributing agro-morphological traits. Based on population structure analysis, sunflower genotypes showed two sub-populations. For identification of important QTL's controlling the morphological traits, genotypic and phenotypic data combination showed eleven SSR bands had a significant association ($P \leq 0.05$) with 6 out of 10 tested agro-morphological characters. These identified marker trait associations are expected to be helpful for sunflower breeders in designing a successful breeding strategy. © 2020 Friends Science Publishers

Keywords: Sunflower breeding; Microsatellites; Mixed linear model; Genetic diversity; Association mapping; Linkage disequilibrium

Introduction

Marker assisted selection in plant breeding has been found to be very effective in devising more focused and robust crop improvement programs. As this technique provided more precision and accuracy in selection thus saves time, resources and efforts required for the development of new variety/hybrid (Filippi *et al.* 2015). An important part of this methodology is to locate the markers in the genome that are tightly linked to the quantitative trait loci (QTLs) controlling phenotype of the plant (Darvishzadeh 2016). QTL mapping in crop plants is normally achieved by two frequently utilized techniques *i.e.*, QTL mapping and association mapping (Ilyas *et al.* 2018). In conventional QTL mapping, two parents are crossed in an organized way and the association between phenotypic traits and mapped marker loci allows the identification of QTLs. Because of few recombination events, the QTL of interests may not be tightly associated with the marker identified (Myles *et al.* 2009). Whereas, association mapping which is a relatively new approach of mapping QTLs, identifies relationship between gene polymorphism and phenotypic variation in existing germplasm collections without the development of mapping populations (Fusari *et al.* 2012). It proves to be efficient in detecting the markers that are tightly linked to a specific QTL (Abdurakhmonov and Abdukarimov 2008). Association mapping utilizes the population structure and linkage disequilibrium (LD) information, hence, also known as LD mapping (Thornsberry *et al.* 2001). Association mapping provides high resolution mapping by arresting all the meiosis occurring in the breeding history of the crop. Moreover, it is cost effective and time saving technique as compared to QTL mapping or linkage analysis (Ambreen *et al.* 2018).

Microsatellites/SSR markers are frequently being employed for population structure analysis in association mapping studies because of their proven ability for generation of more information content as compared to

To cite this paper: Ibrar D, R Ahmad, MA Khan, A Javaid, Z Ahmed, M Ahmad, T Mahmood (2020). Population structure, linkage disequilibrium and association mapping study through SSR markers in sunflower (*Helianthus annuus*). *Intl J Agric Biol* 24:51–58

biallelic markers (Fusari *et al.* 2012; Filippi *et al.* 2015). In Pakistan, sunflower (*Helianthus annuus* L.) was introduced during 1960's along with safflower (*Carthamus tinctorius*) and soybean (*Glycine max*) to increase the local edible oil production. During 2015–16 sunflower was cultivated on an area of 866003 hectares in Pakistan and 35000 tons of vegetable oil was extracted (Ibrar *et al.* 2018). Worldwide, sunflower is the fourth biggest vegetable oil producing crop after palm oil (*Elaeis guineensis*), soybean and canola (*Brassica napus*) (Rauf *et al.* 2017). From hybrid breeding prospective, it is considered as second most important crop after maize (*Zea mays*) (Seiler *et al.* 2017). Domestication of sunflower was started in Pre-Colombian times, but the breeding efforts for oil types were started in 18th century (Mandel *et al.* 2011). Heterosis on commercial scale was exploited in sunflower after the incorporation of CMS (cytoplasmic male sterility) genes by Leclerq (1969) and discovery of male fertility restoration genes by Kinman (1970) and Lochner (2011). Assessment of genetic diversity for various agro-morphological traits is a pre-requisite for manipulating and introgression precisely for achieving the crop improvement objectives under optimum and less than optimum conditions (Hussain *et al.* 2018; Noble *et al.* 2018)

In this study 109 sunflower genotypes panel were evaluated for genetic diversity, linkage disequilibrium and population structure so that to detect the SSR loci associated with nine important agro-morphological traits through association mapping.

Despite of considerable progress being made in plant breeding by conventional approaches; the need to save time and resources, increase accuracy and pace of improvement had urged plant breeders to use new and improved breeding strategies by combining new advancement made in the field of genetics and phenomics. Association mapping has been proved an effective approach being utilized globally for marker assisted breeding program. Sunflower is among the most important oilseed crops but only few association mapping studies has been conducted on this crop so far (Filippi *et al.* 2015). Therefore, it is needed to characterize sunflower genotypes to upsurge the level of understanding regarding sunflower worldwide genetic map. In this study 109 sunflower genotypes were used to document the level of their genetic diversity, linkage disequilibrium and population structure through microsatellite markers. The SSR loci located in proximity of the genes controlling morphological traits studied could be highlighted at chromosomal level through association mapping analysis. The information gathered in this study will be helpful for plant breeders working on the improvement of morphological traits through directed and precise breeding approaches.

Materials and Methods

Plant material and phenotyping

Present study was performed on 109 sunflower lines (Table 1) maintained by Oilseeds Research Program (ORP) of National Agricultural Research Center (NARC), Islamabad. For phenotypic evaluation, these sunflower lines were planted in open field conditions at NARC, Islamabad during spring 2016 following augmented block design. Each sunflower line was planted in a 5 m row with row to row distance of 75 cm and plant to plant distance of 25 cm. NPK fertilizers was applied @ 150:60:60 kg/ha. Complete doze of phosphorus and potassium was applied as basal along-with half of the nitrogen and remaining half of nitrogen was applied at first irrigation. Thinning was done after 12–15 days to ensure proper plant population. Weeding was done manually twice to keep crop weed free during the growth period. Ten morphological parameters *viz.* days to flower initiation (days taken from date of sowing till 5% of the plant of an entry starts flowering), days to flower completion (days taken from sowing till 95–100% plants of an entry initiate flowering), days to maturity (days taken from sowing till 95% of the plants turn their brackets brown), plant height (height of the plant from soil surface to base of head at maturity), stem curvature (plant height subtracted from height of the head from the ground surface), number of leaves per plant, leaf area, hundred seed weight, seed yield and oil content were recorded for the phenotyping of sunflower material.

DNA extraction and genotyping with SSR markers

Total genomic DNA was extracted from 10–12 old sunflower seedlings following CTAB (Cetyl trimethylammonium bromide) DNA extraction protocol (Murray and Thompson 1980). DNA extracted was diluted in 50 in μL of TE buffer for working solution and stored at -4°C. Purity and concentration of genomic DNA was checked by running it on 1% agarose gel.

Overall, 40 SSR markers were employed for genotyping (Table 3). These microsatellites were selected from the 17 linkage groups in sunflower identified by Yu *et al.* (2002) so that a uniform representation of sunflower genome could be ensured. For PCR analysis 20 μL of reaction mixture was prepared from 1–1.2 μL DNA solution, 2 μL Taq Buffer, 2.5 μL $MgCl_2$, 2 μL dNTP's mixture, 0.2 μL Taq Polymerase enzyme, 10.8 μL dd.H_2O and 0.8 μL each of forward and reverse primers (Primers were first diluted with dd.H_2O for making their working solution). A touchdown cycling program was employed to reduce the spontaneous amplification of the PCR product. Cycling protocol comprises of initial denaturation at 94ºC for 5 min, followed by 30 cycles of 94ºC for 30 s, annealing temperature for 30 sec (it ranges from 55ºC to 62ºC for different primers), extension temperature of 72ºC for 40 sec, with final extension at 72ºC for 5 min. The PCR products then obtained were run on a 2% agarose gel for visualization of the amplified segments.

Linkage disequilibrium

Pairwise LD among the SSRs was calculated using the TASSEL program (v.3.0.174) based on the allele frequency correlation (r^2) and LD was drawn to represent the pairwise LD measurements graphically.

Population structure

For population structure analysis Bayesian clustering approach was followed for the SSR genotyping data in the Structure program (v.2.3.4) (Falush *et al.* 2003). For calculation of ancestry fractions of each cluster an admixture model and correlated allele frequencies were applied to each accession, 10 independent runs for each K-value (1–10) were completed with burn-in period of 10,000 followed by 10,000 Markov Chain Monte Carlo repetitions. The delta K method was implemented in Structure Harvester Program (Earl 2012) to determine the most suitable K-value. An unrooted neighbor joining phylogenetic tree was drawn using TASSEL program (v.3.0.174).

Association mapping

Association mapping among the phenotypic data of nine morphological traits and the genotypic data of 40 SSR markers was performed using the program TASSEL (v 3.0.174). SSR markers with known linkage groups and their corresponding positions were used. MLM (mixed linear model) of association mapping that uses both population structure and kinship matrix was employed so that to minimize the probability of false association that may arise in GLM (general linear model) based method. The association between marker and trait was considered significant at $P \leq 0.05$ (Ambreen *et al.* 2018).

Results

Phenotypic variability and population structure analysis

High level of phenotypic variability was observed in the field condition among the sunflower genotypes for all the studied traits (Table 2). Phenotypic data collected exhibited a broad variation among the sunflower studied panel making it an ideal population for documenting the genotypic variability. The data of ten morphological traits was then later combined with the genotypic data revealed by the SSR markers genotyping to highlight the underlying genes controlling these traits. Forty SSR markers were used for the population structure analysis that amplified a total of 65 bands. The admixture model-based analysis provided information about the optimal number of sub-populations. As the value of K increased from 1 to 10 (Fig. 1) LnP(D) also increases continuously and maximum inflection was noticed as the value of K changed from 1 to 2 (Fig. 2). This optimum number of k was further validated by ΔK, a second order statistics. ΔK value also showed a peak at K=2 (Fig. 3). This shows that there were two sub-populations in our samples, based on SSR genotyping data. Further analysis of these 2 sub-populations revealed that these two populations contained maintainer and CMS lines separately.

Neighbor-joining tree

An unrooted neighbor joining phylogenetic tree diagram was generated in TASSEL (v.3.0.174) to compute the level of relatedness among the sunflower accessions. Sunflower accessions study panel was divided into three clusters namely CI, CII and CIII (Fig. 4). CI contains A and B lines mainly with some mixture of open-pollinated and few R-lines as well. Likewise, CIII mainly consists of R-lines, whereas, C-II contains few R-lines along-with some open-pollinated lines. This tree diagram validated that female lines are quite distinct from male (R) lines. However, diversity within the clusters was limited.

Linkage disequilibrium analysis

Linkage disequilibrium was assessed among the entire forty SSR markers over 109 sunflower accessions. A linkage disequilibrium graph was generated on the basis of squared correlation of allele frequencies. The distribution assembly of squared correlation of allele frequencies (r^2) is shown in Fig. 5. Loci in red, green and blue dots exhibited high level of LD having their p-value less than 0.0001, 0.001 and 0.01 respectively.

Association mapping analysis

Linked markers along-with their *P*-values are shown in Table 4. In this study 11 markers showed a significant association with the underlying QTLs controlling six studied traits i.e., head diameter, leaf area, seed yield, days to flower initiation, days to flower completion and hundred seed weight while no significant marker trait association was detected for plant height, stem curvature, oil content and number of leaves per plant. More than one marker was found to have strong correlation with head diameter and days to flower initiation. A scatter plot diagram was drawn to illustrate the markers expressing strong association with the traits studied with threshold value at $P \leq 0.05$ (Fig. 6).

Discussion

Rapid domestication and urge for more productive cultivars housed with superior quality attributes had resulted in more yields but at the cost of loss in the genetic diversity. To address the threats faced by the crops with narrow genetic base like sunflower, scientists are establishing and maintaining large and genetically divergent germplasm material. In this present research, 109 diverse sunflower

Table 1: List of Sunflower Accessions used in present study

S. No.	Accession No.	Source	S. No.	Accession No.	Source
1	CMS-HAP-12	*NARC, Islamabad*	56	RHP-38	*NARC, Islamabad*
2	CMS-HAP-56	*NARC, Islamabad*	57	RHP-77	*NARC, Islamabad*
3	CMS-HAP-101	*NARC, Islamabad*	58	RHP-82	*NARC, Islamabad*
4	CMS-HAP-54	*NARC, Islamabad*	59	RHP-42	*NARC, Islamabad*
5	CMS-HAP-103	*NARC, Islamabad*	60	RHP-73	*NARC, Islamabad*
6	CMS-HAP-24	*NARC, Islamabad*	61	RHP-74DN	*NARC, Islamabad*
7	CMS-HAP-110	*NARC, Islamabad*	62	RHP-7485	*NARC, Islamabad*
8	CMS-HAP-112	*NARC, Islamabad*	63	RHP-7490	*NARC, Islamabad*
9	CMS-HAP-25	*NARC, Islamabad*	64	RHP-7495	*NARC, Islamabad*
10	CMS-HAP-111	*NARC, Islamabad*	65	RHP-7498	*NARC, Islamabad*
11	CMS-HAP-10	*NARC, Islamabad*	66	RHP-74100	*NARC, Islamabad*
12	CMS-HAP-114	*NARC, Islamabad*	67	RHP-74105	*NARC, Islamabad*
13	CMS-HAP-115	*NARC, Islamabad*	68	RHP-74107	*NARC, Islamabad*
14	CMS-HAP-03	*NARC, Islamabad*	69	RHP-74108	*NARC, Islamabad*
15	CMS-HAP-99	*NARC, Islamabad*	70	RHP-74110	*NARC, Islamabad*
16	CMS-HAP-125	*NARC, Islamabad*	71	RHP-74112	*NARC, Islamabad*
17	CMS-HAP-118	*NARC, Islamabad*	72	RHP-74115	*NARC, Islamabad*
18	CMS-HAP-116	*NARC, Islamabad*	73	RHP-74120	*NARC, Islamabad*
19	CMS-HAP-121	*NARC, Islamabad*	74	RHP-74125	*NARC, Islamabad*
20	CMS-HAP-117	*NARC, Islamabad*	75	RHP-74128	*NARC, Islamabad*
21	CMS-HAP-122	*NARC, Islamabad*	76	RHP-74130	*NARC, Islamabad*
22	CMS-HAP-120	*NARC, Islamabad*	77	RHP-71	*NARC, Islamabad*
23	CMS-HAP-123	*NARC, Islamabad*	78	SFP-14	*NARC, Islamabad*
24	CMS-HAP-102	*NARC, Islamabad*	79	SFP-12	*NARC, Islamabad*
25	CMS-HAP-08	*NARC, Islamabad*	80	SFP-10	*NARC, Islamabad*
26	CMS-HAP-119	*NARC, Islamabad*	81	SFP-40	*NARC, Islamabad*
27	HAP-12	*NARC, Islamabad*	82	SFP-42	*NARC, Islamabad*
28	HAP-56	*NARC, Islamabad*	83	SFP-38	*NARC, Islamabad*
29	HAP-101	*NARC, Islamabad*	84	SFP-18	*NARC, Islamabad*
30	HAP-54	*NARC, Islamabad*	85	SFP-36	*NARC, Islamabad*
31	HAP-103	*NARC, Islamabad*	86	SFP-31	*NARC, Islamabad*
32	HAP-24	*NARC, Islamabad*	87	SFP-37	*NARC, Islamabad*
33	HAP-110	*NARC, Islamabad*	88	SFP-24	*NARC, Islamabad*
34	HAP-112	*NARC, Islamabad*	89	SFP-09	*NARC, Islamabad*
35	HAP-25	*NARC, Islamabad*	90	SFP-41	*NARC, Islamabad*
36	HAP-102	*NARC, Islamabad*	91	SFP-19	*NARC, Islamabad*
37	HAP-10	*NARC, Islamabad*	92	SFP-22	*NARC, Islamabad*
38	HAP-114	*NARC, Islamabad*	93	SFP-25	*NARC, Islamabad*
39	HAP-116	*NARC, Islamabad*	94	SFP-43	*NARC, Islamabad*
40	HAP-123	*NARC, Islamabad*	95	SFP-33	*NARC, Islamabad*
41	HAP-111	*NARC, Islamabad*	96	SFP-46	*NARC, Islamabad*
42	HAP-99	*NARC, Islamabad*	97	SFP-08	*NARC, Islamabad*
43	HAP-122	*NARC, Islamabad*	98	SFP-07	*NARC, Islamabad*
44	HAP-120	*NARC, Islamabad*	99	SFP-16	*NARC, Islamabad*
45	HAP-03	*NARC, Islamabad*	100	SFP-26	*NARC, Islamabad*
46	HAP-08	*NARC, Islamabad*	101	SFP-13	*NARC, Islamabad*
47	RHP-68	*NARC, Islamabad*	102	SFP-35	*NARC, Islamabad*
48	RHP-72	*NARC, Islamabad*	103	SFP-20	*NARC, Islamabad*
49	RHP-53	*NARC, Islamabad*	104	SFP-32	*NARC, Islamabad*
50	RHP-73-1	*NARC, Islamabad*	105	RHP-83	*NARC, Islamabad*
51	RHP-46	*NARC, Islamabad*	106	RHP-84	*NARC, Islamabad*
52	RHP-76	*NARC, Islamabad*	107	RHP-88	*NARC, Islamabad*
53	RHP-41	*NARC, Islamabad*	108	RHP-86	*NARC, Islamabad*
54	RHP-81	*NARC, Islamabad*	109	RHP-89	*NARC, Islamabad*
55	RHP-69	*NARC, Islamabad*			

lines that included A, B R and OPV's were studied. Ten morphological traits studied showed a high level of genetic variability in field conditions (Table 2) making this panel of sunflower genotypes an ideal fit for determining the marker-trait association by combining the phenotypic and genotypic data. For genotyping 40 SSR markers that yielded 65 scorable DNA bands. SSR markers have been found to be very effective as they showed more resolution power than SNPs (Emanuelli *et al.* 2013).

SSR based genomic data is the most common method of studying the genetic diversity and population structure analysis. Genomic SSR's are useful marker types because of their abundance in genome, higher polymorphic content and reproducibility (Filippi *et al.* 2015). It has also been reported in recent studies that SSR markers produce same results as obtained through SNPs from GBS (Souza *et al.* 2018).

In depth knowledge of population structure in important to avoid any spurious or false associations

Table 2: Mean, range and standard deviation studied traits among sunflower genotypes

Crop traits	Range	Mean	Standard deviation	Genotypes with highest value	Genotype with lowest value
DFI	69 –94	79.86	5.85	RHP-7485	CMS-HAP-102
DFC	72 – 102	91.45	8.12	RHP-7495	CMS-HAP-102
PH	121.8 – 246.25	185.46	27.13	SFP-42	CMS-HAP-111
SC	15.5 – 36.7	27.68	5.45	SFP-12	SFP-31
HD	9.25 – 20.9	14.50	2.93	HAP-10	RHP-69
L/P	22 – 4189	30.84	3.99	SFP-33	HAP-111
L/A	139.08 – 276.48	202.31	39.13	SFP-37	CMS-HAP-12
HSW	2.76 – 6.94	4.92	0.98	CMS-HAP-111	RHP-42
SY/P	23.9 – 62.7	37.14	7.31	CMS-HAP-54	RHP-41
OC	35.6 – 59.93	44.18	4.99	CMS-HAP-12	SFP-33

DFI: Days to flower initiation; DFC: Days to flower completion; PH: Plant height; SC: Stem curvature; L/P: Leaves per plant; L/A: Average leaf area; HSW: 100-seed weight, SY/P: Seed yield per plant; SY/ha: Seed yield per hectare OC: Oil contents

Table 3: List of SSR markers used to study the association mapping and population structure in sunflower

Primer Name	Linkage Group	Forward Sequence	Reverse Sequence
ORS-605	1	CGCGTGATGTGACGATTATT	ACGGAGCAAAGTTTCGAGGT
ORS-543	1	CCAAGTTTCAGTTACAATCCATGA	GGTCATTAGGAGTTTGGGATCA
ORS-371	1	CACACCACCAAACATCAACC	GGTGCCTTCTCTTCCTTGTG
ORS-453	2	CCTGTGAGCTACAATACTCCCACA	GATTCTGATTAGGCGGTGGT
ORS-1053	2	TTTCATCACATTAGACCATAGACCA	GGCTTTCCTTCGTGGTTTGTAT
ORS-752	3	CACTGATGAACAAGTGCGAGA	ATGATTCCCATACCCACCAA
ORS-924	3	TAAATCGCCATACCACTCCATC	TATCAGCAGGAAGAACGCCTAAT
ORS-366	4	AACCAACTGAGCATTCTTGTGA	GCGCTAGGTTAAAGAGGACAAA
ORS-1068	4	AATTTGTCGACGGTGACGATAG	TTTTGTCATTTCATTACCCAAGG
ORS-337	4	TTGGTTCATTCATCCTTGGTC	GGGTTGGTGGTTAATTCGTC
ORS-1024	5	GGGAAGTGGGCTTGTCTATGTAT	AACACACCGAAATCACCTATGAA
ORS-533	5	TGGTGGAGGTCACTATTGGA	AGGAAAGAAGGAAGCCGAGA
ORS-608	6	CATGGAAAGCCGAGTTCTCT	CGTGCGTGATTAACATACCC
ORS-1256	6	GATGTTGATGTTGGTGAAGTTGC	CTCCGTCACCTTAAGCACTTGTA
ORS-400	7	CGAACCCGTCTGTACCGTTT	ACTTCGTTCACAAGGCACAA
ORS-700	7	GTACCCACCACGCTTAACCA	AGTCTTCCACAGCAACGTCA
ORS-830	8	CAAGTGCATTAGGTGGTTCTAACA	GCCCTCTGACTGTTGTATGACTG
ORS-599	8	TTCCCTATCACACGCCTCTC	GAAAGGAAGTAGCGGTGGTG
ORS-882	9	AAACCGGCATGTAAGATATTCG	ATCGGGAGCAGAAGAAGAGTATG
ORS-617	9	GGTACTTGGTATTCATGGGTCAT	GACACCGCCAACTTAACACTT
ORS-795	9	CGCTAGTTACACCGCAGATG	TGTCCACAGGTTGAAGATCG
ORS-613	10	GTAAACCCTAGGTCAATTTGCAG	ATCTCCGGAAAACATTCTCG
ORS-1088	10	ACTATCGAACCTCCCTCCAAAC	GGATTTCTTTCATCTTTGTGGTG
ORS-433	10	CCGAGGTTTGATCGCTATTT	AGCGTTTGTGATTTGATTACGA
ORS-769	11	GTTTATTTATGTAGAAATGTTCTGGAA	ATGTGGTGGTAAGGGTTGTTG
ORS-697	11	TTGGGCTGTGGTTCCTTAAC	AAGAGATGGGAGTGTTGATGC
ORS-1085	12	GACCTCAAGGCATGCTAACACTC	ACTAAGTGTGTGGACGGGGAAA
ORS-1040	12	CTGCTGATCGTTTCTTGGATAGA	TGCTAATCCTTCTAATCAACTTCCAC
ORS-879	13	GAACCTCCCTTTGTCTGCATATC	CTCCGGTTGCTGTTGATGTCT
ORS-781	13	GTCAACCCATGACCCAAACC	GATGTGGAGGAGAGAGGGTGT
ORS-511	13	TGGCTCAGATTAAGTTCACACAG	CGGGTTGCGAGTAACAGGTA
ORS-307	14	CAGTTCCCTGAAACCAATTCA	GCAGTAGAAGATGACGGGATG
ORS-1086	14	TTGTTTGTCGCACACTCAAGATT	ATTATCGGCACATCTTTGGATTT
ORS-857	15	ACATCCGAACGAAGGACAATC	CAAGAAAGTATGTCACCCAATAGCA
ORS-562	15	CACACACACAAACCCTAGCTCT	CAATCATATCGAGCACACATCA
ORS-768	16	CCACTCATCATCAAGCCTAACA	AGGTGGTGCTGGTTGTAGGT
ORS-1064	16	TGAATGATCTATGAGTGGTGATGG	ACTCGCAGTGGTAAGTCGTTAGG
ORS-495	16	CCAGGATTAGGTAGCTTAGTTCG	GCGATCTGAGGTTGACTCGT
ORS-811	17	CCTTCTCCTCAATCTTTGGCTA	AGGAATGAAATGGGTGTGTGT
ORS-845	17	GGTGCCCTATCTTCATTCTCTG	CTAAAGGGTATCACACATTTGACATT

(Flint-Garcia *et al.* 2005). Based on the diagrams of ΔK and LnP(D), the sunflower accessions were divided into two sub-populations as highest peak was observed at K = 2. Mandel *et al.* (2013) studied 433 sunflower lines and fund two sub-populations based on the optimum K value. In sunflower population studies CMS and R lines tends to group separately from each other and this trend has been observed previously by (Lochner 2011; Ibrar *et al.* 2018)

Linkage disequilibrium is an important analysis while performing marker assisted selection and association mapping analysis. It is considered as a non-random association of alleles at different loci present on the same chromosome. It assumes the co-segregation of a specific trait and DNA marker and by using this information locates

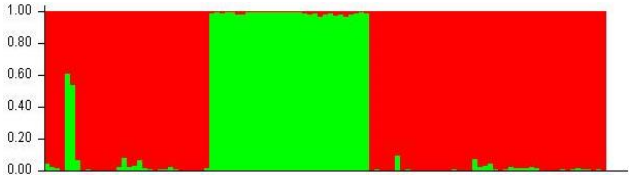

Fig. 1: Population structure of 109 sunflower accessions based on SSR genotyping data at K=2 (Structure 2.3.4)

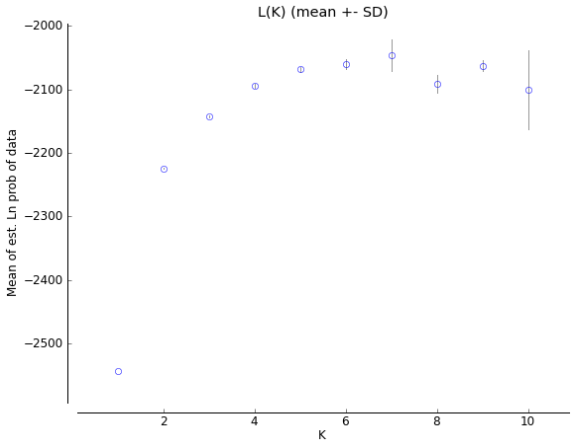

Fig. 2: Diagram of LnP(D) of possible clusters (K) from 1 to 10

the QTLs or major genes (Darvishzadeh 2016). Resolution of association mapping depends on the structure of LD across the genome. LD is dependent on various factors like outcrossing, inbreeding, population size, physical separation and recombination frequency between loci, mutation rate, selection, historical sub-division and admixture of populations and genomic rearrangements (Ilyas *et al.* 2018). Presence of linkage disequilibrium is a pre-requisite for any association mapping studies as LD determines the significance of association among QTL's and the phenotype (Maulana *et al.* 2018). Mandel *et al.* (2011) also reported two distinct populations of sunflower one primarily composed of R lines and the other of B or female lines. This distinction among R and B lines is may be due to their separate breeding history and origin (Lochner 2011). The rare mixing of B lines in R line group may arise because of continuous breeding efforts as sunflower breeders had introgressed superior traits from one genotype to the other for the development of superior inbred lines for hybrid development (Fick and Miller 1997).

This study provided us to detect QTLs controlling some important morphological traits like head diameter, days to flower initiation, leaf area, 100 seed weight, days to flower completion and seed yield per plant. These traits expressed a strong correlation with seed yield in sunflower (Arshad *et al.* 2010; Jalil *et al.* 2014); thus these markers can be used for designing sunflower breeding programs for increased seed yield alongwith short duration. However for determination of marker-trait association with other

Table 4: List of traits along with significantly associated markers ($P \leq 0.05$)

Trait	Marker	*P*- value
Head diameter	ORS-1088	0.0054
Head diameter	ORS-371	0.0054
Leaf area	ORS-1085	0.0079
Seed yield per plant	ORS-769	0.019
Head diameter	ORS-608	0.025
Head diameter	ORS-608	0.025
Head diameter	ORS-608	0.025
Days to flower initiation	ORS-433	0.027
Days to flower completion	ORS-433	0.027
Hundred seed weight	ORS-811	0.028
Days to flower initiation	ORS-605	0.033

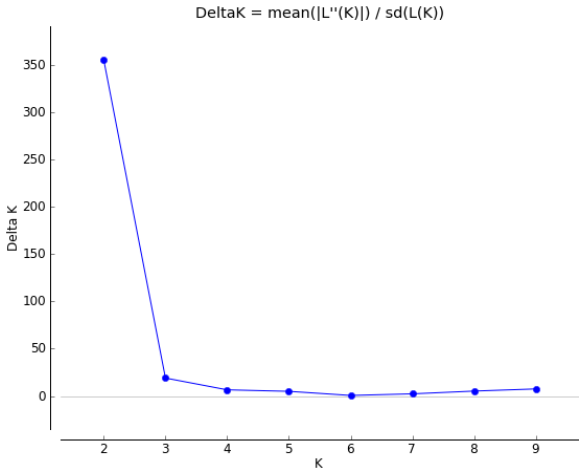

Fig. 3: ΔK based on the rate of change of LnP(D) between successive K values

important characters like oil content and plant height further evaluation should be carried out. Neighbor joining tree shows two sub-populations in the sunflower studied panel along-with a minor cluster containing OPVs. However, out of two major clusters one contains mostly A and B lines while other predominantly contains R lines. Similar clustering pattern has previously been reported by Lochner (2011) and Ibrar *et al.* (2018).

Conclusion

As very few association mapping studies have been conducted on this crop; therefore, in this effort forty SSR markers was utilized to identify marker-trait association. MLM based approach of association mapping coupled with Q+K model identified eleven SSR markers that have been found to be in proximity ($p \leq 0.05$) of the genes controlling six phenotypic traits i.e., head diameter, days to flower initiation and completion, leaf area, seed yield and 100-seed weight. The marker-trait association identified could be used in designing sunflower improvement/breeding programs with more precision and efficacy thus saving the time and resources needed for new cultivar/hybrid development.

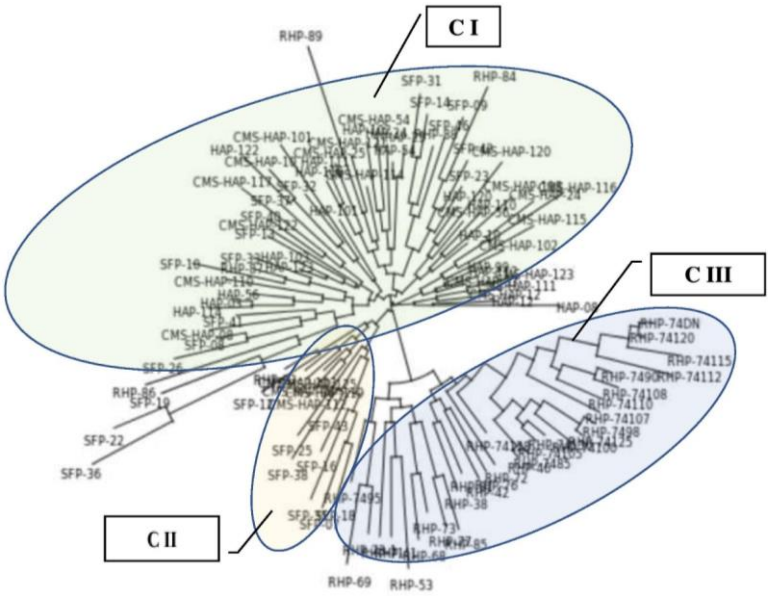

Fig. 4: An unrooted neighbor joining phylogenetic tree of 109 sunflower accessions based on SSR data. C I (Cluster I) consists of mainly CMS and maintainer lines, C II (Cluster II) consists of OPV's and C III (Cluster III) contains restorer lines

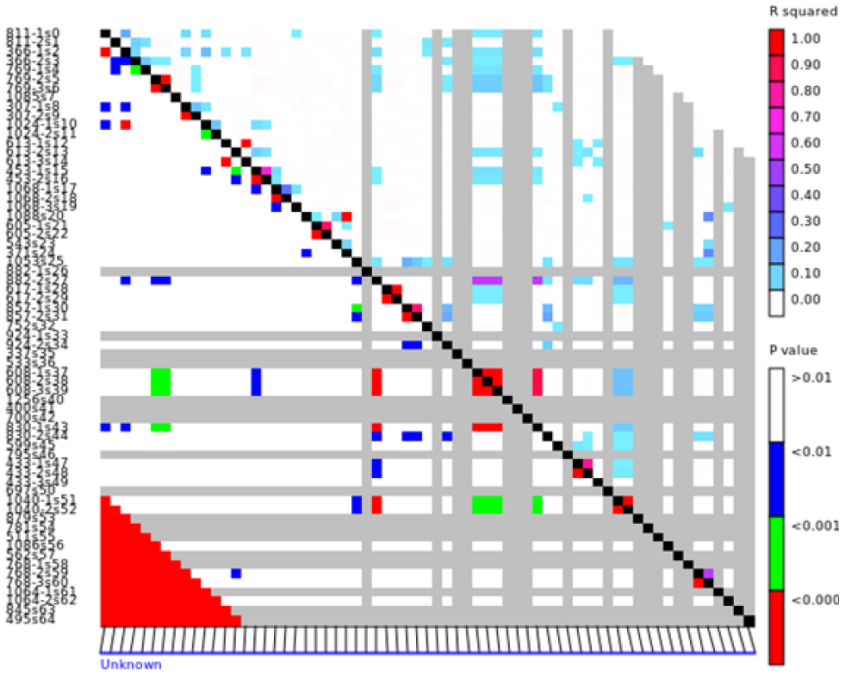

Fig. 5: LD distribution pattern based on squared correlation of allele frequencies (r^2). Each cell represents the comparison of two pairs of marker sites with color codes for the presence of significant linkage disequilibrium

Acknowledgements

The author is highly thankful to Higher Education Commission of Pakistan (HEC) for providing funds to conduct this research under the Indigenous PhD Fellowship Program, Phase-II Batch-I. Plant genetic resources institute, NARC, Islamabad for providing the lab facilities and Oilseeds Research Program NARC for providing the sunflower accessions and field facilities.

References

Abdurakhmonov IY, A Abdukarimov (2008). Application of association mapping to understanding the genetic diversity of plant germplasm resources. *Intl J Plant Genomics* 2008; Article 574927

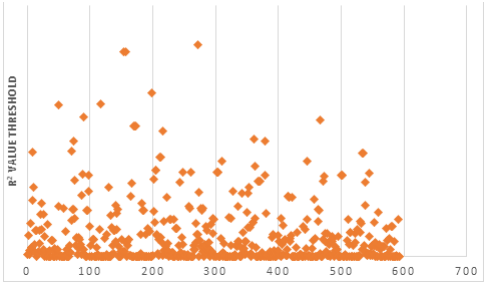

Fig. 6: A scatter plot depicting the marker-traits association

Ambreen H, S Kumar, A Kumar, M Agarwal, A Jagannath, S Goel (2018). Association mapping for important agronomic traits in safflower (*Carthamus tinctorius* L.) core collection using microsatellite markers. *Front Plant Sci* 9; Article 402

Arshad M, MA Khan, SA Jadoon, AS Mohmand (2010). Factor analysis in sunflower (*Helianthus annuus* L.) to investigate desirable hybrids. *Pak J Bot* 42:4393–4402

Darvishzadeh R (2016). Population structure, linkage disequilibrium and association mapping for morphological traits in sunflower (*Helianthus annuus* L.). *Biotechnol Biotechnol Equip* 30:236–246

Earl DA (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Rresour* 4:359–361

Emanuelli F, S Lorenzi, L Grzeskowiak, V Catalano, M Stefanini, M Troggio, S Myles, JM Martinez-Zapater, E Zyprian, FM Moreira, MS Grando (2013). Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. *BMC Plant Biol* 13; Article 39

Falush D, M Stephens, JK Pritchard (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164:1567–1587

Fick GN, JF Miller (1997). Sunflower breeding. *In: Sunflower Production and Technology*, pp:395–440. Scheiter AA (Ed.). American Society of Agronomy, Madison, Wisconsin, USA

Filippi CV, N Aguirre, JG Rivas, J Zubrzycki, A Puebla, D Cordes, MV Moreno, CM Fusari, D Alvarez, RA Heinz, HE Hopp, NB Paniego, VV Lia (2015). Population structure and genetic diversity characterization of a sunflower association mapping population using SSR and SNP markers. *BMC Plant Biol* 15; Article 52

Flint-Garcia SA, AC Thuillet, J Yu, G Pressoir, SM Romero, SE Mitchell, J Doebley, S Kresovich, MM Goodman, ES Buckler (2005). Maize association population: A high-resolution platform for quantitative trait locus dissection. *Plant J* 44:1054–1064

Fusari CM, JAD Rienzo, C Troglia, V Nishinakamasu, MV Moreno, C Maringolo, F Quiroz, D Álvarez, A Escande, E Hopp, R Heinz, VV Lia, NB Paniego (2012). Association mapping in sunflower for sclerotinia head rot resistance. *BMC Plant Biol* 12; Article 93

Hussain M, S Farooq, W Hasan, S Ul-Allah, M Tanveer, M Farooq, A Nawaz (2018). Drought stress in sunflower: Physiological effects and its management through breeding and agronomic alternatives. *Agric Water Manage* 201:152–167

Ibrar D, MA Khan, T Mahmood, M Ahmad, IA Hafiz, SA Jan, R Ahmad (2018). Determination of heterotic groups among sunflower accessions through morphological traits and total seed storage proteins. *Intl J Agric Biol* 20:2025–2031

Ilyas M, G Shabbir, SA Jan, MA Rabbani, SI Malik, M Ansar, MI Ibrahim, H Khurshid (2018). Estimation of linkage disequilibrium and population structure among *Brassica napus* genotypes for association mapping. *Fresen Environ Bull* 27:5980–5986

Jalil S, HA Sadaqat, HN Tahir (2014). Correlation studies among yield related traits for seed yield in sunflower (*Helianthus annuus* L.) under charcoal rot stress condition. *Eur Sci J* 10:391–398

Kinman M (1970). *New developments in the USDA and state experiment station sunflower breeding programs*, pp:181–183. *In: Proceedings of the 4th International Sunflower Conference*, Memphis, Tennessee, USA

Leclerq J (1969). Present day problems in monasticism. *Downside Rev* 87:135–156

Lochner TC (2011). Prediction of heterotic groups and hybrid performance in south african sunflower (*Helianthus annuus* L.) germplasm using SSR analysis. *MSc Thesis*. University of the Free State, Bloemfontein, South Africa

Mandel JR, S Nambeesan, JE Bowers, LF Marek, D Ebert, LH Rieseberg, SJ Knapp, JM Burke (2013). Association mapping and the genomic consequences of selection in sunflower. *PLoS Genetics* 9; Article e1003378

Mandel JR, JM Dechaine, LF Marek, JM Burke (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

Maulana F, H Ayalew, JD Anderson, TT Kumssa, W Huang, XF Ma (2018). Genome-wide association mapping of seedling heat tolerance in winter wheat. *Front Plant Sci* 9; Article 1272

Myles S, J Peiffer, PJ Brown, ES Ersoz, Z Zhang, DE Costich, ES Buckler (2009). Association mapping: Critical considerations shift from genotyping to experimental design. *Plant Cell* 21:2194–2202

Noble TJ, Y Tao, ES Mace, B Williams, DR Jordan, CA Douglas, SG Mundree (2018). Characterization of linkage disequilibrium and population structure in a mungbean diversity panel. *Front Plant Sci* 8; Article 2012

Rauf S, N Jamil, SA Tariq, M Khan, M Kausar, Y Kaya (2017). Progress in modification of sunflower oil to expand its industrial value. *J Sci Food Agric* 97:1997–2006

Seiler GJ, LL Qi, LF Marek (2017). Utilization of sunflower crop wild relatives for cultivated sunflower improvement. *Crop Sci* 57:1083–1101

Souza LMD, LHBD Santos, JRBF Rosa, CCD Silva, CC Mantello, ARO Conson, EJJ Scaloppi, JDF Fialho, MLTD Moraes, PDS Goncalves, GRA Margarido, AAF Garcia, VL Guen, APD Souza (2018). Linkage disequilibrium and population structure in wild and cultivated populations of rubber tree (*Hevea brasiliensis*). *Front Plant Sci* 9; Article 815

Thornsberry JM, MM Goodman, J Doebley, S Kresovich, D Nielsen, ES Buckler (2001). Dwarf8 polymorphisms associate with variation in flowering time. *Nat Genet* 28:286–289

Yu JK, J Mangor, L Thompson, KJ Edwards, MB Slabaugh, SJ Knapp (2002). Allelic diversity of simple sequence repeats among elite inbred lines of cultivated sunflower. *Genome* 45:652–660