ANALYSIS OF POPULATION STRUCTURE AND GENETIC DIVERSITY IN BREAD WHEAT GENOTYPES

ISRAR AHMAD*[1], HABIB AHMAD[2], INAMULLAH[3], SAMI ULLAH KHAN[1], ABDUL RAZZAQ[4], MUHAMMAD ISLAM[3], MUHAMMAD SHOAIB AMJAD[1]

[1] Department of Botany, Women University of AJK, Bagh, Pakistan
[2] Islamia College University Peshawar, Pakistan
[3] Department of Genetics, Hazara University Mansehra, Pakistan
[4]Centre of Plant Biodiversity. University of Peshawar, Pakistan
Corresponding email: iabotany32@gmail.com

Abstract

Bread wheat *(Triticum aestivum;* 2n=6x=42; AABBDD) having a genome size of 17 Gb is the major staple food of Pakistan. Progress in plant breeding is facilitated by accurate information about genetic structure and diversity. In the present research hundred bread wheat genotypes were evaluated. A total of 102 molecular markers (SSR) out of 150 were found to be polymorphic. The highest marker diversity (66%) was showed by *Xwmc 798, Xbarc 147, Xgwm 60, Xgwm 469, Xgwm 471* followed by *Xbarc 154* (65%), *Xgwm 372* (65%), *Xwmc 52* (65%), *VRN AF* (65%) *Xbarc 172* (64%), *Xgwm 261* (64%) while the lowest diversity (9%) was found in *Xbarc 137*. Polymorphic information content (PIC) values of the markers was also calculated in the range of 0.03 – 0.59. The highest PIC value was confirmed in *Xgwm 471*(0.59), *Xbarc 147* (0.59) and the lowest (0.03) was recorded in *Xwmc* 606. Genotypic data of 102 SSR markers was applied across the whole genome of wheat for analysis of population structure. An admixture model with correlated allele frequencies for determination of population structure was used. The analysis of population structure was accomplished using structure software. Bar plot shows all the hundred genotypes are admixed due to complex and long history of evolution. All the hundred genotypes showed 100% admix with no purity.

Key words: Evolution, Molecular markers, PIC value, Structure, Wheat,

Introduction: The allohexaploid bread wheat evolved by mixing of three grass genomes as A- genome from *T. urartu*, B-genome from *Aegilops speltoides* and D-genome from *Aegilops tauschii* (Dvorak and Zhang, 1992). *Triticum aestivum* L. (Bread wheat) is one of the most important cereal crops around the world, growing on an area of 218 million hectare in 2010 (FAO, 2012). Wheat germplasm showing a high level of trait to trait genetic diversity due to presence of superior alleles. These alleles are very useful for structure analysis, marker trait association using association mapping as well as QTL mapping (Brenchley *et al.,* 2012). A number of molecular markers have been used to dissect diversity among wheat genotypes. Profiling of RFLP, RAPD, SSR, SNP markers always recorded immense level of genotyping (Landjeva *et al.,* 2007). Among all these markers SSR or microsatellites are considered to be better for diversity studies in wheat genotypes. SSR markers are chromosomes specific and multi allelic distributed throughout chromosomes (Novoselovic et al., 2016).

The main objectives of the present research was to profile hundred bread wheat genotypes on 102 SSR markers for (1) assessing genetic diversity among bread wheat genotypes collected from different geographic regions; (2) Describing genetic population structure; (3) Providing polymorphic information content (PIC) values of SSR markers for diversity studies;

MATERIALS AND METHODS:

The plant materials consisted of hundred bread wheat genotypes collected from different region of Pakistan and CIMMYT (Table 1). All the genotypes were evaluated for genetic diversity. The wheat

genotypes were grown in the field of Hazara University Mansehra, Pakistan for three years (2011-2013). Genomic DNA was extracted using Weining and Langridge, 1992 method with some modifications. The DNA concentration was measured through Nano Drop Analyzer (model ND-1000 Spectrophotometer NanoDrop Technologies, Inc. Wilmington, USA.) at University of Leicester UK. The concentrated DNA was diluted further to the required (50 ng/ul) quantity by the following dilution formula:

$$\text{Genomic DNA in ng/ul} \times X = 50 \text{ ng/ul} \times 100$$
$$\text{So} \quad X = 50 \text{ ng/ul} \times 100 / \text{Genomic DNA in ng/ul}$$

Polymerase Chain Reaction (PCR) was carried out using protocol described by Roder et al., 1998. Each PCR was carried out in a 25µL reaction volume, containing 11.3µL double-distilled deionized H2O, 2.5µL 10X buffer, 2µL MgCl2, 2µL dNTPs, 0.2µL Taq polymerase, 1µL of each primer, and 5µL DNA.

Thermo cycling conditions were required as:

(i) Denaturation of double standard genomic DNA template at 94ºC.
(ii) Annealing of specific primer with the template DNA at specific temperature.
(iii) Extension of Primer at 72ºC and formation of new DNA strand.

The molecular data was analyzed using different computer softwares as power marker version 3.2, Mega 6, structure version 2.3.4 and structure harvester. Structure software commonly used for determination of population structure of diverse populations (Pritchard et al., 2000). A burn-in of 20000 runs and MCMC 20000 iterations were used to test the K value in the range of 2-20. The online structure harvester program was used for estimation of number of clusters (K) using logarithmic likelihood LnP(D) (Yu et al., 2006) while PIC value was calculated by using computer package power marker.

RESULTS: A total of 102 molecular markers included in the present study. These markers produced a total of 271 alleles across the hundred wheat genotypes. The number of alleles per locus ranged from 1-3 with an average of 2.63 per locus. All the markers showed relatively high polymorphism (table 2). Most of the primers have displayed a maximum of 3 and minimum of 1 allele. The highest marker diversity (66%) was showed by *Xwmc 798, Xbarc 147, Xgwm 60, Xgwm 469, Xgwm 471* followed by *Xbarc 154* (65%), *Xgwm 372* (65%), *Xwmc 52* (65%), *VRN AF* (65%) *Xbarc 172* (64%), *Xgwm 261* (64%) while the lowest diversity (9%) was found in *Xbarc 137*. The overall mean diversity among all the markers was recorded as 47%. Polymorphic information content (PIC) values of the markers was also calculated in the range of 0.03 – 0.59 . The highest PIC value was confirmed in *Xgwm 471*(0.59), *Xbarc 147* (0.59) and the lowest (0.03) was recorded in *Xwmc606*. The overall average PIC value found as 0.40. The alleles of high frequency per locus (major allele‘s frequency) ranged from 0.38 to 1 with mean of 0.62 (table 12). Our results of molecular markers polymorphism matching with results of Liu et al (2010b) for association mapping of wheat for agronomic traits and Maccaferri et al (2011) for association mapping of durum wheat.

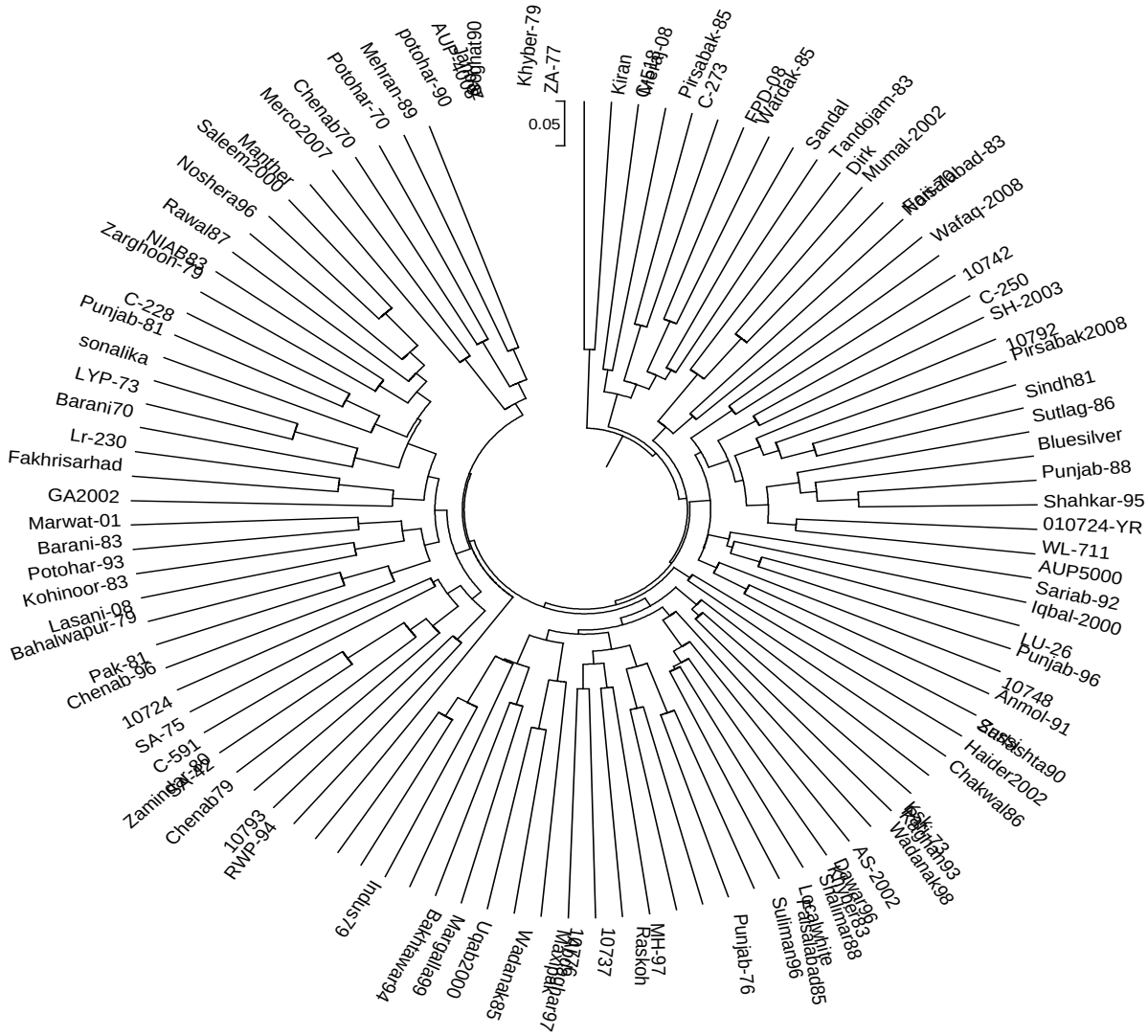
Genotypic data of 102 SSR markers was applied across the whole genome of wheat for analysis of population structure. An admixture model with correlated allele frequencies for determination of population structure was used (Falush et al., 2003). The analysis of population structure was accomplished using structure software (Pritchard et al., 2000). Burn-in of 20,000 iterations followed by 20,000 MCMC (Monte Carlo Markov Chain) replicates was used to test K values (number of subpopulations) in the range of 2-20 while performed 10 runs for K values. The suitable cluster numbers (K) was calculated using online structure harvester software (Yu et al., 2006) by applying logarithmic likelihood LnP(D) (natural log of probability data) method (figure 2). Two major peaks have been detected at K=2 and K=13 (Evanno et al., 2005).

The hundred wheat genotypes at K=2 were separated into two subgroups, G1 and G2. Group G1 comprised of local land races while G2 contain CIMMYT lines (010724, 010737, 010748, 010776 and 010792) as

Figure 1: UPGMA tree constructed using molecular markers showing diversity across hundred wheat genotypes

well as local land races. Bar plot shows all the hundred genotypes are admixed due to complex and long history of evolution (figure 2). All the hundred genotypes showed 100% admix with no purity.

All the genotypes were divided into 13 sub-groups at K=13 as G1, G6, G11, G 12 and G13 comprised of 44 (44%) genotypes consisting of both local and CIMMYT lines (figure 2). Group G2, G3, G4, G5, G7, G8, G9 and G10 include 56 (56%) admix genotypes (all were local genotypes).


Khyber-79
ZA-77
0.05
Kiran
Pirsabak-85
C-273
FPD-08
Wardak-85
Sandal
Tandojam-83
Dirk
Murtal-2002
Wafaq-2008
10742
C-250
SH-2003
10792
Pirsabak2008
Sindh81
Sutlag-86
Bluesilver
Punjab-88
Shahkar-95
010724-YR
WL-711
AUP5000
Sariab-92
Iqbal-2000
LU-26
Punjab-96
10748
Ahmol-91
Haider2002
Chakwal86
AS-2002
Punjab-76
Sulliman96
MH-97
Raskoh
10737
Indus79
Bakhtawar94
Margalla99
Uqab2000
Wadanak98
Mehran-89
Potohar-70
Chenab70
Merco2007
Manther
Saleem2000
Noshera96
Rawal87
NIAB83
Zarghoon-79
C-228
Punjab-81
sonalika
LYP-73
Barani70
Lr-230
Fakhrisarhad
GA2002
Marwat-01
Barani-83
Potohar-93
Kohinoor-83
Lasani-08
Bahalwapur-79
Pak-81
Chenab-96
10724
SA-75
C-591
Chenab79
10793
RWP-94

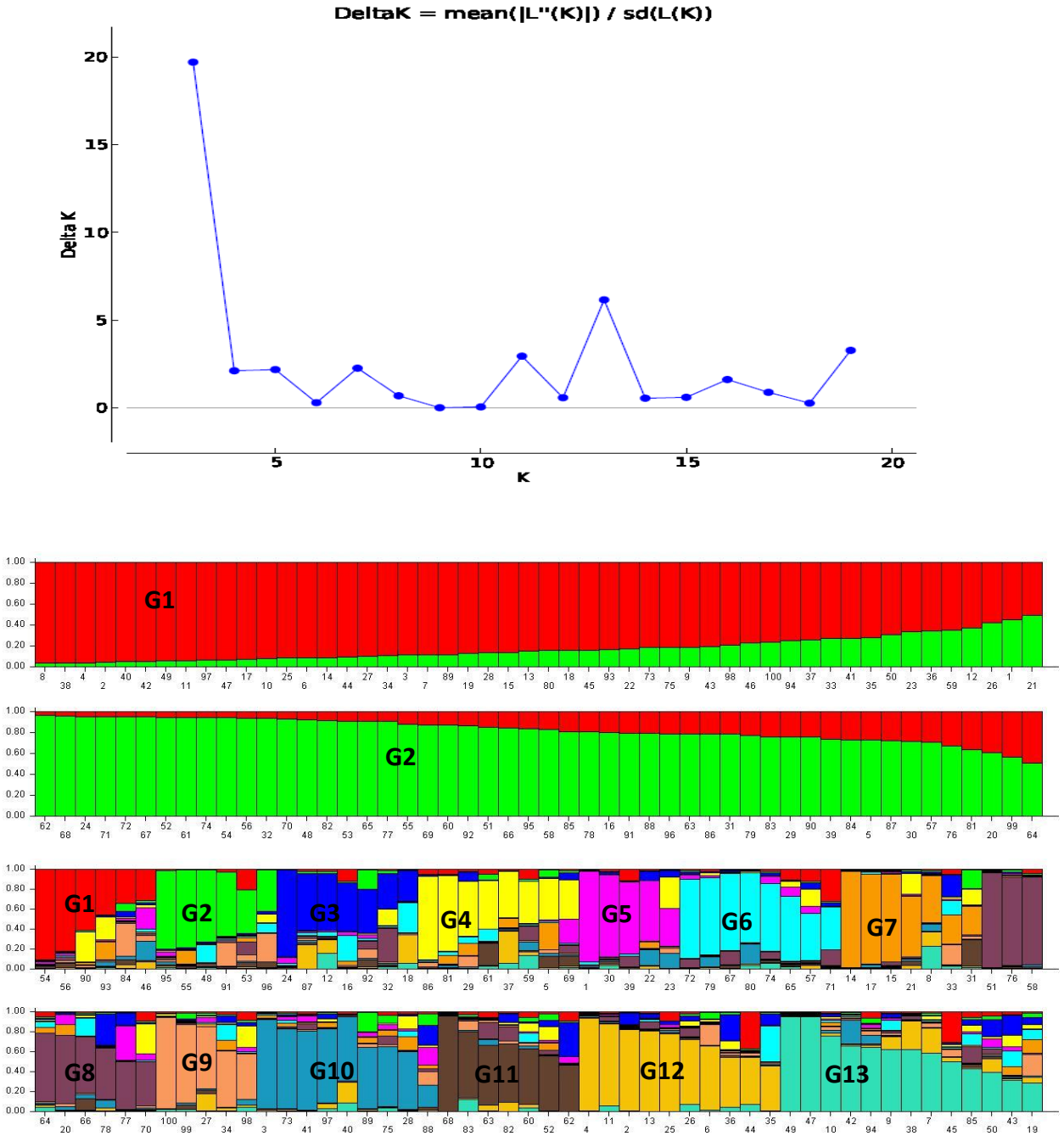


Figure 2(a,b,c): Population structure analysis of wheat genotypes based on SSR markers (a) Line graph. The X-axis shows LnP (D) value and Y-axis shows k. (b) Graphical bar plot at k=2 presenting two subgroup (G1 & G2). (c) Graphical bar plot at k=13 presenting thirteen subgroup (G1- G13). The X-axis shows accessions numbers and Y-axis shows sub group membership.

Discussion: estimation of genetic diversity through structure analysis is essential platform for crop improvement. Evaluation of the genetic diversity, population structure and relationships provide valuable information needed to broaden the narrow genetic base and to enhance breeding and conservation strategies of crops. The UPGMA tree constructed on the base of molecular markers showing great extent of diversity among CIMMYT and local genotypes. This diversity would be of core initial step for breeders in genotypes selection for developing new promising lines. SSR markers showed their efficiency in assessing the genetic diversity of different crops. The crossing among inter groups may suggest a promising lines having stable agronomic traits. The present research revealed that regional contribution of breeders could unlock the genetic make up for improving conservation strategies of wheat. Group G2, G3, G4, G5, G7, G8, G9 and G10 include 56 (56%) admix genotypes (all were local genotypes) could be crossed with the remaining groups for getting modern genotypes having good agronomic traits. Average of PIC value suggest that SSR markers could be used for diversity studies. Furthermore, the CIMMYT lines are admix with local genotypes due to extensive exploitation of breeding with local genotypes for getting promising lines.

Acknowledgments: The author is very thankful to funding agency (Higher Education Commission of Pakistan (HEC)) for successful completion of this huge project under 5000 Indigenous scholarships (Batch No 2) of PIN No. 117-1836-BM7-274. I am also thankful to NARC for providing bread wheat germplasms.

REFERENCES

Brenchley R, Spannagl M, Pfeifer M, Barker GLA, D'Amore R, et al. (2012). Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491: 705–710.

Dvorak, J., and Zhang, H. K. (1992). Reconstruction of the phylogeny of the genus Triticum from variation in repeated nucleotide sequences. Theor. Appl. Genet.84, 419–429. doi: 10.1007/BF00229502

Evanno, G., S. Regnaut and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol., 14: 2611-2620.

Falush. D., M. Stephens and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics, 164: 1567-1587.

FAO (2012) FAOSTAT production. Available:/http://faostat3.fao.org/home/index.html#DOWNLOAD. Accessed 17 March 2013.

Landjeva S, Korzun V, Börner A (2007) Molecular markers: actual and potential contributions to wheat genome characterization and breeding. Euphytica 156: 271–296.

Liu, L., L. Wang, J. Yao and Y. Zheng. 2010b. Association mapping of six agronomic traits on chromosome 4A of wheat (Triticum aestivum L.). Mol. Plant Breed., 1: 1-10.

Maccaferri, M., M. C. Sanguineti, A. Demontis, A. El-Ahmed, L G. del Moral, F. Maalouf, M. Nachit, N. Nserallah, H. Ouabbou, S. R. C. Royo, C. Royos, D. Villegas and R. Tuberosa. 2011. Association mapping in durum wheat grown across a broad range of water regimes. J. Exp. Bot., 62: 409-438.

Novoselovic, D, A.R. Bentley, R. Šimek, K. Dvojkovic, M.E. Sorrells, N. Gosman, R. Horsnell, G. Drezner and Z. Šatovic. 2016. Characterizing Croatian Wheat Germplasm Diversity and Structure in a European Context by DArT Markers. Front. Plant Sci. 7:184. doi: 10.3389/fpls.2016.00184

Pritchard, J. K., M. Stephens and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics, 155: 945-959.

Roder, M. S,. V. Korzun, K. Wendehake, J. Plaschke, et al. 1998. A microsatellite map of wheat. Genetics 149: 2007-2023.

Weining, S and P. Langridge. 1992. Identification and mapping of polymorphism in cereals base on polymerase chain reaction. Theor. Appl. Genet. 82: 209-216.

Yu, J., G. Pressoir, W. H. Briggs, B. I. Vroh, M. Yamasaki, J. Doebley, M. Mcmullen, B. Gaut, J. J. Holland, S. Kresovich and E. Buckler. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet., 38: 203-208.

Table 2: SSR markers, their chromosome position (ch pos), Major Allele frequency (MAF), allele No, genetic diversity (H) and polymorphic information content (PIC) used for profiling of hundred wheat genotypes.

Marker	Ch pos	MAF	Allele No	H	PIC	Marker	Chr pos	MAF	Allele No	H	PIC
Cfd 15	1AS,1D	0.94	3	0.11	0.11	Xbarc 154	7A	0.39	3	0.65	0.58
Cfd 18	5D	0.92	2	0.15	0.14	Xbarc 158	1AL	0.49	3	0.61	0.53
Xwmc 24	1AS	0.61	2	0.48	0.36	Xbarc 159	2BL	0.63	3	0.53	0.46
Xwmc 25	2B	0.54	3	0.60	0.53	Xbarc 163	4BS	0.77	2	0.35	0.29
Xwmc 27	2B,5B	0.69	3	0.47	0.42	Xbarc 164	3BL	0.67	3	0.50	0.45
Xwmc 43	3B,3D	0.59	2	0.48	0.37	Xbarc 165	5AL	0.48	3	0.62	0.54
Xwmc 51	7B	0.58	2	0.49	0.37	Xbarc 167	2BS	0.44	3	0.63	0.56
Xwmc 52	1B,4D	0.43	3	0.65	0.58	Xbarc 172	7DL	0.46	3	0.64	0.57
Xwmc 94	7D	0.59	2	0.48	0.37	Xbarc 173	6DS	0.45	3	0.64	0.56
Xwmc 97	5D	0.51	3	0.62	0.55	Xbarc 175	6DL	0.47	3	0.64	0.57
Xwmc 104	1A,6B	0.69	3	0.47	0.42	Xbarc 264	7AL	1.00	1	0.00	0.00
Xwmc 147	1D,3A	0.77	3	0.38	0.34	Xgwm 4	4AS	0.85	2	0.26	0.22
Xwmc 149	5B	0.66	3	0.51	0.45	Xgwm 10	2AS	0.51	3	0.62	0.55
Xwmc 153	1D,3A	0.79	3	0.35	0.32	Xgwm 33	1DS	1.00	1	0.00	0.00
Xwmc 154	2B	0.70	2	0.42	0.33	Xgwm 37	7DL	0.49	3	0.61	0.53
Xwmc 157	7D	0.79	2	0.33	0.28	Xgwm 55	2BL	0.51	3	0.58	0.49
Xwmc 161	4A	0.53	2	0.50	0.37	Xgwm 60	7AS	0.38	3	0.66	0.59
Xwmc 163	6A	0.53	2	0.50	0.37	Xgwm 71	3DS	0.59	3	0.57	0.50
Xwmc 166	7B	0.66	3	0.48	0.40	Xgwm 99	1AL	0.52	2	0.50	0.37
Xwmc 167	2D	0.66	3	0.47	0.38	Xgwm 111	7DS	0.52	2	0.50	0.37
Xwmc 168	7A	0.47	3	0.63	0.56	Xgwm 136	1A	0.69	3	0.48	0.43
Xwmc 169	3A	0.66	3	0.48	0.41	Xgwm 194	4DL	0.48	3	0.56	0.46
Xwmc 175	2B	0.66	3	0.50	0.45	Xgwm 261	2DS	0.45	3	0.64	0.57
Xwmc 177	2A	0.52	2	0.50	0.37	Xgwm 293	5AS	0.49	3	0.54	0.43
Xwmc 181	2D	0.89	2	0.20	0.18	Xgwm 299	3BL	0.53	3	0.60	0.53
Xwmc 182	7B	1.00	1	0.00	0.00	Xgwm 302	7BL	0.48	3	0.62	0.55
Xwmc 216	1D	0.57	2	0.49	0.37	Xgwm 325	6DS	0.88	2	0.21	0.19

Xwmc 219	4A	0.55	3	0.52	0.41	Xgwm 359	2AS	0.72	3	0.43	0.37
Xwmc 232	4A	0.91	2	0.16	0.15	Xgwm 372	2AL	0.39	3	0.65	0.58
Xwmc 233	5D	0.62	3	0.48	0.37	Xgwm 389	3BS	0.50	3	0.63	0.55
Xwmc 235	5BL	0.48	3	0.57	0.47	Xgwm 443	5BS	0.56	3	0.59	0.52
Xwmc 398	6BC	0.50	3	0.60	0.52	Xgwm 471	7AS	0.37	3	0.66	0.59
Xwmc 420	4AS	0.53	3	0.60	0.53	Xgwm 469	6DS	0.38	3	0.66	0.59
Xwmc 606	7BS	0.98	2	0.04	0.04	Xgwm 484	2DS	0.45	3	0.62	0.53
Xwmc 718	4AL	0.56	3	0.59	0.52	Xgwm 544	5BS	1.00	1	0.00	0.00
Xwmc 749	6DC	0.50	3	0.62	0.55	Xgwm 608	4DC	0.50	3	0.61	0.53
Xwmc 798	1BS	0.38	3	0.66	0.59	Xgwm 609	4DL	0.62	3	0.54	0.48
Xbarc 42	3DS	0.85	2	0.26	0.22	Xgwm 642	1DL	0.93	2	0.13	0.12
Xbarc 45	3AS	0.63	3	0.53	0.47	xgwm 908	2DS	1.00	1	0.00	0.00
Xbarc 76	6BS	0.56	3	0.59	0.52	Xgdm 3	5DS	0.84	2	0.27	0.23
Xbarc 101	2BL	0.60	3	0.56	0.50	Xgdm 5	2DS	0.82	3	0.31	0.29
Xbarc 127	6B	0.49	3	0.56	0.46	Xgdm 6	2DL	0.66	2	0.45	0.35
Xbarc 128	2BL	0.59	3	0.55	0.48	Xgdm 19	1DL	0.76	2	0.36	0.30
Xbarc 134	6BL	0.52	3	0.60	0.53	Xgdm 28	1BS	0.68	3	0.49	0.44
Xbarc 137	1BL	0.95	2	0.10	0.09	Xgdm 33	1DS	0.70	3	0.46	0.41
Xbarc 140	5BL	0.81	3	0.33	0.30	Xgdm 46	7DL	0.63	3	0.53	0.46
Xbarc 141	5AL	0.68	3	0.48	0.43	Xgdm 114	2BS	0.48	3	0.61	0.53
Xbarc 144	5DL	0.47	3	0.61	0.53	VRN AF	5A	0.38	3	0.66	0.58
Xbarc 147	3BS	0.36	3	0.67	0.59	VRN B1 R3	5B	0.55	2	0.50	0.37
Xbarc 148	1AS	0.62	3	0.54	0.48	PpD1 R1	2A	0.68	3	0.47	0.40
Xbarc 149	1DS	0.68	2	0.44	0.34	PpD 1 R2	4D	0.55	3	0.60	0.53
						Mean		0.62	3	0.47	0.41

Table 1: List of the genotypes evaluated in the present study for drought stress

S.No	Genotypes	S.No	Genotypes
1	sonalika	2	Shalimar 88
3	Merco 2007	4	Khyber 83
5	Manther	6	Chenab 70
7	Lr-230	8	Soghat 90
9	Ksk	10	Pari -73
11	Maxi pak	12	Chakwal 86
13	Indus 79	14	Wadanak 98
15	Bakhtawar 94	16	Nori -70
17	Wadanak 85	18	ZA- 77
19	Abdaghar 97	20	Kaghan 93
21	Margalla 99	22	Dawar 96
23	Uqab 2000	24	Suliman 96
25	Raskoh	26	AS -2002
27	Haider 2002	28	LYP -73
29	Local white	30	Noshera 96
31	MH-97	32	Sindh 81
33	Zarlashta 90	34	Fakhri sarhad
35	Punjab-76	36	10737
37	Faisalabad 85	38	10776
39	Barani 70	40	10748
41	Rawal 87	42	10724
43	NIAB 83	44	10792
45	GA 2002	46	Pirsabak 2008
47	Chenab 79	48	Punjab-96
49	Saleem 2000	50	Mumal-2002
51	Zamindar-80	52	SA-42
53	Iqbal-2000	54	Marwat-01
55	SH-2003	56	Barani-83
57	Anmol-91	58	Potohar-93
59	LU-26	60	Kohinoor-83
61	Chenab-96	62	Potohar-70
63	Faisalabad-83	64	Pak-81
65	Zarghoon-79	66	Pirsabak-85
67	C-228	68	C-273
69	Shahkar- 95	70	Tandojam-83
71	Punjab-88	72	Dirk
73	10793	74	Bahalwapur-79
75	Punjab-81	76	Lasani-08
77	C-591	78	Sussi
79	Sutlag-86	80	Khyber-79

81	C-250	82	FPD-08
83	Blue silver	84	Sandal
85	RWP-94	86	Kiran
87	Sariab-92	88	Wardak-85
89	Wafaq-2008	90	Meraj-08
91	10742	92	C-518
93	010724-YR	94	potohar-90
95	AUP 5000	96	Mehran-89
97	WL-711	98	Janbaz
99	SA-75	100	AUP-4008