Pak. J. Bot., 56(3): 989-999, 2024. DOI: http://dx.doi.org/10.30848/PJB2024-3(29)

DEVELOPMENT OF A CORE SET OF SIMPLE SEQUENCE REPEAT MARKERS FOR DNA FINGERPRINTING OF TOMATO GERMPLASM

RAHIL SHAHZAD*, SHAKRA JAMIL, SAJID UR RAHMAN, ERUM YASMEEN AND MUHAMMAD ZAFFAR IQBAL

Agricultural Biotechnology Research Institute, Ayub Agricultural Research Institute, Faisalabad, Pakistan
**Corresponding author's email: rahilshahzad91@gmail.com*

Abstract

Distinctness, uniformity and stability (DUS) testing of varieties based on morphological, physiological and biological descriptors has raised many concerns i.e., these traits are influenced by environment hence non reproducible. Therefore, DNA fingerprinting has been proposed as a necessary part of DUS testing of a plant variety for protection of PBRs and registration of a variety in PBRs registry. Present study was conducted for DNA fingerprinting and genetic diversity assessment in recently developed 13 tomato genotypes including hybrids, OPVs and inbred lines. About 212 SSR markers were used for DNA fingerprinting and genetic diversity assessment. Out of these 199 markers were amplified whereas 13 were not amplified. A total of 1314 alleles were amplified by 199 SSR markers and among these 912 alleles (70%) were polymorphic which exhibited high genetic diversity among genotypes. Cluster and structure analysis was performed to study the genetic relationship of varieties with each other. Genetic similarity coefficient between genotypes ranged from 0.76 to 0.88. Cluster and structure analysis grouped genotypes to three distinct classes. Hybrids genotypes i.e. Salar-F1, Saandal-F1, Sundar-hybrid and Ahmar-hybrid were clustered with each other, two of the OPVs i.e. Nadir and Naqeeb shared same cluster and inbred lines 8502, 8505, 9108, 13195, 8504, 68543 and 17253 were shared the same cluster. Different markers for unique identification of genotypes were also proposed i.e. XM010323394 was most informative SSR marker as it amplified unique alleles for 8 genotypes i.e. Naqeeb, Nadir, 17253, 8502, 8504, 8505, 68543 and 9108. A set of 50 most informative SSR markers was proposed for future DNA fingerprinting studies based on number of alleles, polymorphic alleles and PIC contents. Results reported in this study would be useful for future DNA fingerprinting and genetic diversity studies and protection of varieties under Plant Breeders Rights Rules.

Key words: Allelic polymorphism, Cluster analysis, Genetic diversity, Polymorphic information content, Structure analysis, Plant breeders rights.

Introduction

Tomatoes (*Solanum lycopersicum* L.) is an economically valuable fruit crop that is widely grown in temperate regions across the globe (Kulus, 2018). It is a diploid, dicot, self-pollinated vegetable crop (2n=24 chromosomes) with 950 Mb genome size and is used as model organism for *Solanaceae* family (Kaushal *et al*., 2017; Foti *et al*., 2021). Continuous breeding efforts in tomato have evolved many high yielding varieties. However, in the race of development of high yielding varieties, genetic diversity is compromised and cultivated gene pool has narrow genetic base (Iqbal *et al*., 2021a; Alcalá-Gómez *et al*., 2022). There are possibilities that identical varieties may got registered with different trademarks. Therefore, for varietal identification in breeding programs, DUS (distinctness, uniformity and stability) testing is considered core for selection and examination of varieties. Distinctness means that the variety must be distinct from any other variety of the same species; uniformity explains that the candidate variety must be unique and uniform in its characteristics and traits, and these traits must be stable and should not change even after repeated propagations (Lone *et al*., 2021; Rahman *et al*., 2022). Traditionally, DUS testing for classification and differentiation of crops was based on morphological descriptions such as color, fruit shape, height etc. that displayed variable results due to environmental fluctuations and spatio-temporal variations which lead to misidentification of varieties. These techniques were also laborious, time consuming, expensive and impractical (Iqbal *et al*., 2021b).

On the contrary, DNA fingerprinting using DNA markers is an effective tool for identification of varieties and assessment of genetic diversity nullifying the effects of environmental manipulations (Rahman *et al*., 2022). There are different types of DNA markers i.e. Non-PCR Based (RFLPs) or PCR Based markers (RAPD, AFLP, KASP and SSR). However, simple sequence repeats (SSR) markers are considered to be more reliable. These are microsatellites, tandem repeats of 2 to 4 base pairs randomly dispersed in genome which are originated due to unequal crossing over or slippage of DNA polymerase during replication process. SSR are highly reproducible, multi-allelic, genetically codominant and abundant (Jamil *et al*., 2020). These short and standardize DNA regions are referred "barcodes" are powerful tools which distinguish different cultivars at the DNA level (Jamil *et al*., 2021). In addition, they are robust, unambiguous and amplified using a single primer pair, hence, used for varietal identification, genome mapping and marker assisted selection (Iqbal *et al*., 2021b).

It has been established that SSR markers are best tool for development of DNA fingerprinting profile of tomato genotypes. However, one question is yet to be answered that how many SSR markers will be sufficient for DNA fingerprinting? One very popular SSR markers database for *Solanum lycopersicum*, Sol Genomics Network (https://solgenomics.net/markers/microsats.pl) possessed 3485 SSR markers which is a huge number. Therefore, there is a need to develop a core set of SSR markers on local germplasm to reveal genetic diversity and development of DNA fingerprinting profile in a time and cost-efficient manner (Mueller *et al*., 2005).

In context of above discussion, present study was designed for DNA fingerprinting of tomato gene pool (varieties/hybrids and inbred lines) available at Vegetable Research Institute, Ayub Agricultural Research Institute, Faisalabad. Most informative 212 SSR markers evenly distributed across the genome and evenly distributed on each chromosome were selected from Sol Genomics Network (https://solgenomics.net/markers/microsats.pl) based on their effectiveness and degree of polymorphism. Some other key objectives of the study were development of core set of SSR markers for DNA fingerprinting of local tomato gene pool based on their polymorphic information content (PIC). Moreover, development of a standardize-able reference barcodes based on DNA fingerprints to enable identification of each genotype.

Material and Methods

Plant materials: Present research work was conducted at Agricultural Biotechnology Research Institute, Ayub Agricultural Research Institute Faisalabad. A population of thirteen tomato genotypes (Table 1) comprising of inbred lines, hybrids and varieties were used for DNA fingerprinting and genetic diversity studies. All 13 genotypes were sown in pots and two weeks old young fresh seedling were collected and stored at -40ºC for DNA extraction.

Table 1. List of thirteen (13) tomato genotypes and their pedigree parentage.

Sr. #	Genotype name	Pedigree parentage
1.	Salar-F1 (Hybrid)	8502 × 8504
2.	Saandal-F1 (Hybrid)	8504 × 68543
3.	Sundar-Hybrid	8502 × 8505
4.	Ahmar-Hybrid	13195 × 9108
5.	Naqeeb (OPV)	Rio stone-2-2 (Exotic hybrid)
6.	Nadir (OPV)	TWL-33-5-2-1 (Exotic hybrid)
7.	17253 (Inbred lines)	QF-Red (Exotic hybrid)
8.	8502 (Inbred line)	TWL-3-2-3-1 (Exotic hybrid)
9.	8504 (Inbred line)	TWL-99-4-5-1 (Exotic hybrid)
10.	8505 (Inbred line)	TWL-3-2-5-1 (Exotic hybrid)
11.	68543 (Inbred line)	Lyco L. 5 (Exotic hybrid)
12.	9108 (Inbred line)	NTH-242 (Introduction)
13.	13195 (Inbred line)	Rio Grande (Introduction)

DNA extraction & PCR: DNA was extracted using modified CTAB method previously described by (Jamil *et al*., 2020a, b; Kanwal *et al.*, 2021). DNA was quantified using Nanodrop spectrophotometer (ND 2000, Thermo Scientific, U.S.A.). It was considered pure when A_{260}/A_{280} ratio ranged between 1.80 and 2.0. The quality of extracted DNA was also assessed by loading 20 ng/ µL DNA on 0.8% (w/v) agarose gel stained with ethidium bromide. After DNA extraction, most polymorphic 212 tomato SSR markers evenly distributed on whole genome were searched from different databases (https://solgenomics.net/markers/microsats.pl) and different scientific papers i.e. (Castellana *et al*., 2020; Al-Shammari *et al*., 2021) and got synthesized from Gene Link USA (https://www.genelink.com/). Primers dilutions were prepared to 100 *µM* and stored for PCR reactions.

PCR were assembled for 212 SSR markers individually in 200 µl eppendorf tube (PCR Tube) with a total reaction volume of 25 *µL* including 12 *µL* of DreamTaq Green PCR master mix 2X (K1081), 0.6 *µM* of each forward and reverse primer and 20 *ng*/*µL* genomic DNA of each genotype (Iqbal *et al*., 2019; Iqbal *et al*., 2021a). The following amplification temperatures were applied at different steps i.e., initial denaturation 95ºC for 10 min, 35 cycles of each denaturation 95ºC for 01 min, annealing at variable temperatures according to primers for 1 min (Table 2) and extension at 72ºC for 1 min. followed by final extension at 72ºC for 10 min. The products were stored at 4ºC before gel running.

Polyacrylamide gel electrophoresis (PAGE) analysis and binary data scoring: The products were visualized on vertical gel electrophoresis also called polyacrylamide gel electrophoresis (PAGE) model POWERPRO-3AMP (cleaver scientific limited) by preparing 06% PAGE. After loading of the samples gel was run on constant power mode (16W) followed by staining of gel using fixative (680 *mL* d_3H_2O acetic acid, 80 *mL* 10% acetic acid, 40 *mL* 10% ethanol solution), strainer (1.4 g silver nitrate dissolved in 800 *mL* d_3H_2O) and developer (12 g NaOH pellets dissolved in 792 *mL* d_3H_2O and 8 *mL* 37% formaldehyde solution) solutions called silver nitrate staining using already described protocol (Jamil *et al*., 2020a). Gel images were captured using the gel documentation system (GelPro, Cleaver Scientific). The data was recorded in the form of binary matrix i.e., 1 for presence of band and 0 for absence of band for certain alleles and in this way 212 SSR marker and their alleles were scored against 13 genotypes.

Statistical data analysis

To understand the genetic relationship among genotypes, binary data was analyzed using similarity/dissimilarity matrix via un-weighted pair group method of arithmetic means (UPGMA) in NTSYSpc (Yujian & Liye, 2010) software version 2.0 and cluster/dendrogram was generated. Genetic diversity was assessed using binary data of 212 SSR markers in model-based Bayesian clustering approach implemented in STRUCTURE v. 2.3.4 (Pritchard *et al*., 2000) for understanding of the population structure and geographic distribution of alleles. Following parameters were used in STRUCTURE v. 2.3.4 for inference of population structure i.e. no admission model; K ranging from 1 to 6; 10,000 Burn-in period; Reps. hypothetical populations number (k) (3), number of in-iteration burns (10,000), number of Markov chain Monte Carlo simulations (100000). Number of clusters in this population were determined by subjecting the population structure results to Evano test by plotting LnP(K) values against ΔK values (Evanno *et al*., 2005). Binary data was used for calculation of Polymorphic Information content value (PIC) for each SSR marker. Similarly number of alleles (NOA), Polymorphic alleles (PA) were also calculated for each marker and listed in (Table 2).

Table 2. Alleles Polymorphism and Polymorphic information content (PIC) of SSR markers used for DNA Fingerprinting of tomato.

Primer name	Amplification size	Annealing temperature	Total alleles	Polymorphic alleles	Polymorphic information content
AI780156	100-380	50	2	1	0.5
AT2	160-170	50	2	1	0.5
SSR188	140-150	50	2	2	0.48
SSR223	210-225	55	2	1	0.5
SSR286	100-120	48	2	2	0.49
SSR34	190-225	45	2	2	0.5
SSR4	100-175	50	2	1	0.5
SSR40	145-225	50	2	1	0.5
SSR449	490-500	50	2	1	0.5
SSR49	160-180	50	2	2	0.5
SSR595	430-730	55	2	1	0.5
SSR74	225-250	50	2	1	0.5
SSR76	195-205	50	2	1	0.49
TG202	170-190	57	2	1	0.5
TOM11-26	190-240	50	2	1	0.5
AI778183	105-115	50	3	3	0.65
AQ368062	110-125	50	3	3	0.67
AW034362	125-145	46	3	3	0.67
SCN20	100-110	58	3	3	0.64
SSR124	185-700	50	3	2	0.57
SSR231	115-185	50	3	2	0.66
SSR244	150-260	55	3	3	0.64
SSR27	150-160	50	3	3	0.63
SSR285	270-290	55	3	2	0.54
SSR300	140-225	55	3	2	0.67
SSR327	155-175	50	3	2	0.67
SSR52	190-245	50	3	1	0.53
SSR57	125-140	50	3	2	0.67
SSR594	400-600	55	3	2	0.66
SSR602	190-310	55	3	2	0.66
SSR68	150-170	55	3	1	0.67
SSR69	140-160	55	3	3	0.57
SSR72	180-240	52	3	2	0.67
SSR90	180-230	42	3	1	0.59
SSR98	140-150	50	3	2	0.67
AI895126	115-130	46	4	4	0.73
SSR136	145-165	50	4	2	0.74
SSR150	220-275	55	4	2	0.7
SSR218	115-145	55	4	1	0.75
SSR276	150-180	55	4	1	0.71
SSR304	190-225	55	4	2	0.75
SSR325	130-150	55	4	3	0.57
SSR340	200-380	50	4	2	0.75
SSR43	225-275	55	4	2	0.75
SSR5	140-240	45	4	3	0.75
SSR605	145-225	55	4	3	0.71
SSR92	190-225	50	4	2	0.71
TOM146-147	125-225	48	4	1	0.73
TOM67-68	130-145	52	4	1	0.75

Table 2. (Cont'd.).

Primer name	Amplification size	Annealing temperature	Total alleles	Polymorphic alleles	Polymorphic information content
25-C	125-300	50	5	2	0.78
AI486387	185-225	52	5	5	0.76
AI773078	120-145	52	5	2	0.76
KM094129	140-400	50	5	3	0.72
SLR13	225-300	60	5	4	0.66
SSR11	100-250	55	5	3	0.77
SSR115	145-240	50	5	4	0.75
SSR13	100-115	50	5	5	0.7
SSR155	190-240	55	5	5	0.76
SSR17	210-250	50	5	3	0.8
SSR237	100-220	55	5	5	0.63
SSR248	100-275	55	5	5	0.73
SSR293	150-270	50	5	3	0.8
SSR295	140-570	55	5	3	0.8
SSR301	125-210	55	5	4	0.8
SSR310	150-180	55	5	3	0.77
SSR335	225-280	55	5	4	0.8
SSR350	125-380	55	5	5	0.75
SSR360	275-520	50	5	4	0.79
SSR48	190-235	50	5	2	0.76
SSR66	185-240	50	5	5	0.73
SSR85	190-350	50	5	2	0.79
TG263	160-400	55	5	3	0.8
TMS9	300-550	46	5	2	0.74
TOM144-145	160-275	48	5	5	0.78
TOM39A-40A	140-275	47	5	4	0.8
TOM47-48	105-205	49	5	4	0.76
ME8-EM5	120-480	45	6	1	0.83
SCAE16	180-500	50	6	2	0.83
SCAF10	140-500	50	6	2	0.85
SSR103	110-175	50	6	2	0.82
SSR109	150-265	50	6	3	0.83
SSR110	170-250	42	6	4	0.78
SSR111	190-235	50	6	5	0.72
SSR159	370-600	50	6	6	0.83
SSR19	140-225	50	6	6	0.83
SSR261	170-195	50	6	5	0.71
SSR306	270-400	55	6	6	0.68
SSR320	100-220	55	6	2	0.83
SSR38	175-330	50	6	1	0.83
SSR50	210-275	55	6	6	0.71
SSR572	290-380	45	6	4	0.83
SSR70	130-275	50	6	6	0.83
SSR75	125-400	40	6	4	0.83
SSR80	190-270	50	6	6	0.83
SSRB56555	290-410	50	6	1	0.83
TOM41-42	140-190	49	6	4	0.8
TOM43-44	160-250	47	6	4	0.82
TOM61-62	200-350	46	6	1	0.83
TOM63-64	180-450	52	6	3	0.8

Table 2. (Cont'd.).

Primer name	Amplification size	Annealing temperature	Total alleles	Polymorphic alleles	Polymorphic information content
X13437	200-275	52	6	5	0.75
X90937	240-650	52	6	4	0.83
ZUP641	240-305	45	6	4	0.83
SLR4	165-250	56	7	3	0.81
SSR296	110-520	50	7	3	0.84
SSR326	200-440	55	7	3	0.84
SSR344	100-330	60	7	7	0.83
SSR349	165-330	55	7	3	0.85
SSR37	105-610	45	7	3	0.86
SSR47	115-240	50	7	4	0.81
SSR51	130-550	50	7	4	0.83
SSR593	200-370	55	7	5	0.86
SSR599	155-380	55	7	2	0.85
SSR65	245-300	50	7	7	0.82
SSR95	175-320	55	7	4	0.85
TOM236-273	120-250	49	7	6	0.85
TOM57-58	210-290	46	7	1	0.86
TOM65-66	170-550	48	7	3	0.81
TOM8-9	130-200	54	7	2	0.86
AW031453	200-380	46	8	8	0.86
AW037347	155-250	46	8	7	0.81
AY562123	175-270	46	8	2	0.88
LEAT014	140-280	50	8	1	0.87
SSR105694	120-300	50	8	4	0.88
SSR108	100-600	45	8	6	0.87
SSR122	125-450	45	8	3	0.8
SSR146	180-400	50	8	7	0.86
SSR308	150-410	55	8	3	0.88
SSR330	190-400	50	8	2	0.88
SSR48	120-650	50	8	4	0.86
SSR526	100-320	60	8	7	0.87
SSR578	300-410	55	8	5	0.85
SSR580	240-380	41	8	3	0.88
SSR67	145-700	50	8	5	0.86
SSR73	255-390	45	8	8	0.85
SSRB102358	250-430	50	8	1	0.85
SSRB60800	180-490	50	8	4	0.86
LEATA004	115-500	45	9	8	0.88
MI23	120-560	45	9	7	0.89
SCN13	180-480	50	9	4	0.89
SLR3	115-200	56	9	9	0.87
SSR104	150-550	45	9	8	0.9
SSR128	110-175	50	9	8	0.82
SSR134	175-580	50	9	8	0.9
SSR139	150-215	45	9	8	0.84
SSR14	130-275	55	9	3	0.97
SSR15	110-230	50	9	7	0.86
SSR162	140-275	50	9	8	0.89
SSR192	110-410	50	9	8	0.9
SSR22	140-300	55	9	8	0.86

Table 2. (Cont'd.).

Primer name	Amplification size	Annealing temperature	Total alleles	Polymorphic alleles	Polymorphic information content
SSR29	115-170	50	9	8	0.82
SSR318	100-385	55	9	5	0.87
SSR32	175-215	50	9	9	0.79
SSR44	150-700	45	9	9	0.8
SSR450	260-420	55	9	8	0.8
SSR601	170-200	60	9	9	0.8
SSR62	140-580	40	9	9	0.87
TFSUR1	300-620	47	9	8	0.87
TM548	160-650	48	9	5	0.85
TOM152-153	125-450	47	9	5	0.86
TOM49-50	115-600	47	9	8	0.87
TOM95-96	225-560	52	9	5	0.88
U81986	175-270	54	9	9	0.88
LETTC002	120-500	45	10	10	0.92
NM110278976	100-250	50	10	10	0.88
SLR20	125-500	58	10	1	0.91
SLR21	115-700	58	10	6	0.9
SLR28	145-700	45	10	8	0.89
SSR105	105-245	52	10	9	0.92
SSR201	100-750	45	10	10	0.92
SSR214	100-800	50	10	10	0.88
SSR28	140-500	40	10	10	0.93
SSR333	140-360	50	10	10	0.9
SSR356	100-750	55	10	7	0.9
SSR555	200-400	41	10	10	0.88
SSR565	100-550	55	10	10	0.93
SSR63	100-300	55	10	4	0.9
SSR86	200-420	50	10	9	0.92
TG479	145-510	55	10	8	0.86
TM-533	110-450	58	10	4	0.91
TOM55-56	160-550	52	10	8	0.9
TOM59-60	120-225	46	10	10	0.91
X90770	105-470	52	10	7	0.9
SLR10	110-600	58	11	9	0.94
SLR26	105-450	60	11	10	0.92
SSR09	110-700	55	11	7	0.93
SSR20	110-320	50	11	11	0.93
SSR345	160-900	60	11	9	0.93
SSR46	100-800	50	11	8	0.93
SSR479	100-400	52	11	7	0.93
SSR603	100-800	50	11	7	0.95
SSR94	100-550	50	11	10	0.95
SSRB18031	100-390	50	11	1	0.93
SSRB50753	125-550	50	11	6	0.93
TGO302	100-660	55	11	11	0.96
TOM160-161	100-500	47	11	7	0.92
TOM31A-32A	125-375	50	11	6	0.93
XM010323394	110-550	55	11	9	0.88
Y09371	125-480	52	11	5	0.93
Z1063	120-550	45	11	9	0.9

Table 3. Core set of 50 SSR markers for DNA fingerprinting and genetic diversity studies of tomato varieties.

Marker name	Forward primer	Reverse primer
TGO302	TGGCTCATCCTGAAGCTGATAGCGC	AGTGTACATCCTTGCCATTGACT
SLR26	AACGGTGGAAACTATTGAAAGG	CACCACCAAACCCATCGTC
SSR94	AATCAGATCCTTGCCCTTGA	AGCTGAGAAAGAGCAGCCAT
XM010323394	GACCATTATGTTG TTTGGTGCCG	AGAGGTCCAACTTC TGGATCGCAT
SSR345	AAGCCAAGCTCGAACCTGTA	AAGCCAAGCTCGAACCTGTA
SLR10	AGAATTTTTTCATGAAATTGTCC	TATTGCGTTCCACTCCCTCT
SSR20	TTCGGTTTATTCTGCCAACC	GCCTGTAGGATTTTCGCCTA
LETTC002	TTCTCACACCTGCAACACC	AGCGGGATGATTACAGAAATG
SSR214	AAATTCCCAACACTTGCCAC	CCCACCACTATCCAAACCC
SSR46	CCGAGGCGAATCTTGAATAC	GCACCATCTCTTGTGCCTCT
SSR565	GAGGATGATGAGAACTCGCC	TCAGAGGCTTCTGGGTCAGT
SSR201	AAGACAGAAAGTGCACGTCAGA	TGGATGAGAAGAGGGAATCCT
SSR603	GAAGGGACAATTCACAGAGTTTG	CCTTCAACTTCACCACCACC
TOM59-60	TAACACATGAACATTAGTTTGA	CACGTAAAATAAAGAAGGAAT
NM110278976	GCAACTGCTGTCTT CAGCACTGTAT	GAACTCTGCAAAATC ACTTCACCCT
SSR555	TTGATATTAACCATGGCAGCAG	TTGATGGGATTGCACAGAAA
Z1063	ATTTGAAAGCGTGGTATGC	CTTAAACTCACCATTAAATC
SSR134	CCCTCTTGCCTAAACATCCA	CGTTGCGAATTCAGATTAGTTG
SSR22	GATCGGCAGTAGGTGCTCTC	CAAGAAACACCCATATCCGC
SSR28	ACCAAATGGAAATGGGTCAA	CCCTAAGACTAACGACAACCAA
SSR333	GTTCCCGCTTGAGAAACAAC	CCAATGCTGGGACAGAAGAT
TOM160-161	TGCTGAAGAATACAATGTTACC	ATTGTTGGATGCTCAGTTTG
TOM31A-32A	AATGTCCTTCGTATCCTTTCGT	CTCGGTTTTAATTTTTGTGTCT
SLR3	GCACGAGCACATATAGAAGAGAATCA	CCATTTCATCATATCTCTCAGCTTGC
SSR105	GAGCGGCTTCGAATTCATC	CATTTGAGCAGAAGCGAACA
SSR32	TGGAAAGAAGCAGTAGCATTG	CAACGAACATCCTCCGTTCT
SSR44	TCATCTGCAATTCATGGCTC	AGGTCAAGGATGTGCTTCCC
SSR62	TGCAAATGAATGTCCAGGAT	TCAGCAGAGTTATGCCATGC
SSR86	AGGGCAACAAATCCCTCTTT	GGAGACGAGGCTGCTTACAC
U81986	AGGTTGATGAAAGCTAAATCTGGC	CAACCACCAATGTTCATTACAAGAC
AW031453	GCCGTTCTTGGTGGATTAG	CCTCCTTTCGTGTCTTTGTC
LEATA004	CAACTGGATAGGTCGATG	GATGTGGATGAAACGGATG
SSR28	ACCAAATGGAAATGGGTCAA	CCCTAAGACTAACGACAACCAA
SSR104	TTCCATTTGAATTCCAACCC	CCCACTGCACATCAACTGAC
SSR128	GGTCCAGTTCAATCAACCGA	TGAAGTCGTCTCATGGTTCG
SSR139	TGGGTATGGGATTTACACCAA	AAACGAAGGCAACAACGAAG
SSR192	ACAACATGGGAAGCACTTGA	ATTAAATTGGGCCATGGTGA
SSR162	GCTCTCTACAAGTGGAACTTTCTC	CAACAGCCAGGAACAAGGAT
SSR450	AATGAAGAACCATTCCGCAC	ACATGAGCCCAATGAACCTC
SSR73	TGGGAAGATCCTGATGATGG	TTCCCTTTCCTCTGGACTCA
TFSUR1	CTGAAACTCTCCGTATTTC	CGAAGAGTGATTGGAGAT
TG479	GGTGATTATGGGTGATCCTATG	CCAAGTGAGTCCAACAGTTCC
TOM49-50	AAGAAACTTTTTGAATGTTGC	ATTACAATTTAGAGGTCAAGG
TOM55-56	ATTTCTGTAACTCCTTGTTTC	TGACTTCAACCCGACCCCTCTT
AW037347	GCCACGTAGTCATGATATACATAG	GCCTCGGACAATGAATTG
MI23	TGGAAAAATGTTGAATTTCTTTTG	GCATACTATATGGCTTGTTTACCC
SSR09	CCCTTTGCAAGTTCTTCTTCA	TTCATGAGCCAACATAGGAGG
SSR146	TATGGCCATGGCTGAACC	CGAACGCCACCACTATACCT
SSR15	CACTTGCCATCTTCTAGCCC	ATGGATGCCCAAATTGAAGA
SSR344	TGTTGCTCGAACTCTCCAAA	CATAGGAGAGGTAACCCGCA

Results

SSR polymorphism and selection of core set of markers: Genetic diversity and DNA fingerprinting studies were carried out using two hundred and twelve (212) SSR markers which were evenly distributed across the genome (Table 2). Thirteen (13) of these SSR markers i.e. 43DF1R1, CF-5, NM002187774, SSR112, SSR125, SSR241, SSR266, SSR287, SSR290, SSR388, SSR557, SSRB105694 and TFSUR2 were not amplified even at different annealing temperature (45-60^{o}C). A total of 1314 alleles were amplified by remaining 199 SSR markers generating an average of 6.6 alleles per loci. Out these 912 alleles (70%) was polymorphic generating an average of 4.58 polymorphic alleles per loci exhibiting great extent of genetic diversity in the gene pool under study. Lowest number of alleles (TA) two was amplified by fifteen SSR markers i.e. AI780156, AT2, SSR188, SSR223, SSR286, SSR34, SSR4, SSR40, SSR449, SSR49, SSR595, SSR74, SSR76, TG202 and TOM11-26. Highest number of alleles (11) was amplified by seventeen SSR markers i.e. SLR10, SLR26, SSR09, SSR20, SSR345, SSR46, SSR479, SSR603, SSR94, SSRB18031, SSRB50753, TGO302, TOM160-161, TOM31A-32A, XM010323394, Y09371 and Z1063. Highest number of polymorphic alleles (PA) 11 was amplified by two SSR markers i.e. SSR20 and TGO302 whereas lowest number of polymorphic allele (one) was amplified by 27 markers as listed in Table 2.

Polymorphic information content (PIC) value of 199 SSR markers was also computed which ranged from 0.48 to 0.96. Highest PIC value was recorded for SSR14 (0.97) whereas lowest (0.48) was observed for SSR188. The allele's size in all markers ranged from 100-800 bp whereas bands below 100 bp and above 800 bp were ignored. Allele size for most of the markers ranged from 150-500. On the basis of TA, PA and PIC value, a core set of 50 SSR markers was reported for DNA fingerprinting and genetic diversity studies in tomato for future (Table 3).

Genetic diversity and population structure estimation: Binary data of 199 SSR markers was used to study the genetic relationship and population structure of tomato gene pool. The genetic similarity coefficients between genotypes were studied based on binary data of 1314 TA using Unweighted Pair Group Method with Arithmetic Average (UPGMA). Genetic similarity coefficient between genotypes varied from 0.73 (between genotype Salar-F1 and 17253) to 0.88 (between genotype Nadir and Naqeeb). Genetic similarity coefficient between hybrid genotypes i.e. Salar-F1 to Saandal-F1 (0.87), Salar-F1 to Sundar-hybrid (0.84), Salar-F1 to Ahmar-hybrid (0.81), Saandal-F1 to Sundar-hybrid (0.82), Saandal-F1 to Ahmar-hybrid (0.83), Sundar-hyrbid to Ahmar-hybrid (0.84) was higher than OPVs and inbred lines. Similar trend was observed between OPVs and inbred lines whereas both OPVs Naqeeb and Nadir depicted highest similarity (0.88) as compared to other genotypes (Fig. 1).

On the basis of cluster analysis genotypes were broadly classified into 2 clusters. Salar-F1, Saandal-F1 and Ahmar-hybrid were lying together in cluster I (CI). On the other hand, Ahmar-hybrid, Naqeeb, Nadir, 8502, 8505, 9108, 13195, 8504 and 68543 formed cluster II (C-II) and 17253 was not part of any cluster (Fig. 1). Population structure analysis was also performed to understand the genetic constitution of genotypes under study. Model-based cluster analysis using a Bayesian approach was carried out to infer population structure of 12 potato varieties using data of 199 SSR markers. The LnP(D) scores for number of populations (K) was increased up to 3 and showed inflation point at K3 which divides population into three groups. Similarly, ΔK-value also showed a peak at K = 3, indicating that genotypes comprised of 3 sub-populations. Population 1 (P1) contained 3 genotypes i.e. Salar-F1, Saandal-F1 and Sundar-hybrid whereas Population 2 (P2) also contained 3 genotypes (Naqeeb, Ahmar-hyrbid and Nadir) and Population 3 (P3) comprised of 7 genotypes i.e. 17253, 8502, 8504, 8505, 68543, 9108 and 13195 (Fig. 2).

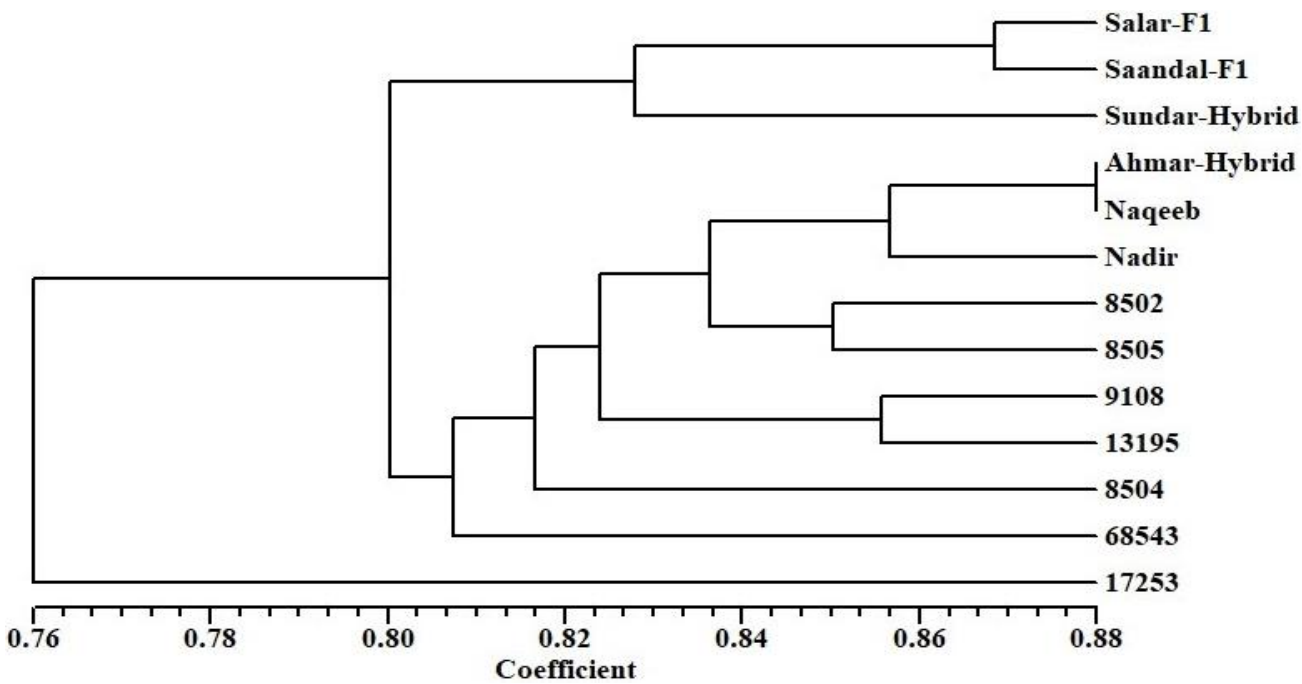
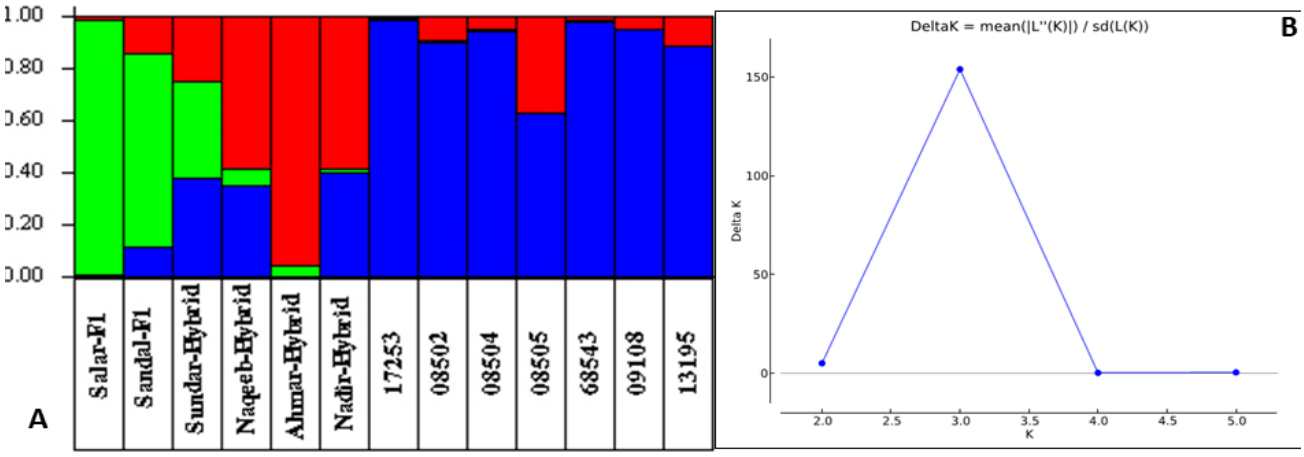


Fig. 1. Cluster analysis of thirteen tomato genotypes using un-weighted Pair Group Method with Arithmetic Average (UPGMA).



C

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	2	-5799.550000	1.767767	—	—	—
2	2	-6495.350000	156.341309	-695.800000	788.200000	5.041534
3	2	-6402.950000	54.376511	92.400000	8369.550000	153.918480
4	2	-14680.100000	5151.131480	-8277.150000	1040.850000	0.202062
5	2	-21916.400000	21484.308176	-7236.300000	8466.100000	0.394060
6	2	-37618.800000	42440.407585	-15702.400000	—	—

Fig. 2. Population structure analysis of thirteen tomato genotypes estimated by using binary data of 212 SSR markers: parameters: no admission model; K = 03; 10,000 Burn-in period; 100000 Rep

Table 4. List of variety specific SSR markers for unique genotypic identification.

Genotype name	SSR marker for unique identification
Salar-F1	A1773078, SLR3, SSR139, SSR134, SSR214, SSR237, SSR52, SSR66, Tom47-48, Z1063
Sandal-F1	SLR3, SSR111, SSR46, Z1063
Sundar-hybrid	SLR26, TG302
Ahmar-hybrid	AW037347, NM110278976, SSR306, SSR450, TOM160-161
Naqeeb	Z1063, XM010323394
Nadir	Z1063, XM010323394
17253	SLR23, SLR10, SSR09, SSR115, SSR128, SSR13, SSR110, SSR237, SSR26, SSR32, SSR325, SSR44, SSR50, SSR603, SSR65, SSR94, TOM160-161, X13437, XM010323394
8502	M123, XM010323394
8504	SLR4, SSR44, TOM31A-32A, XM010323394
8505	TOM41-42, XM010323394
68543	SIR23, SSR48, XM010323394
9108	XM010323394
13195	NM110278976, Z1063

DNA fingerprinting: DNA markers which identified unique alleles with each genotype are listed in Table 4 for unique identification of each genotype. XM010323394 was most informative SSR marker as it amplified unique alleles for 8 genotypes i.e., Naqeeb, Nadir, 17253, 8502, 8504, 8505, 68543 and 9108. Similarly Z1063 was also one of the informative SSR markers which amplified unique alleles for 4 genotypes namingly Salar-F1, Saandal-F1, Nadir and 13195. Four SSR markers identified unique alleles for 2 genotypes i.e. SLR23 (Salar-F1, Saandal-F1), SSR26 (17253, 8502), and TOM160-161 (Ahmar-hybrid, 17253) as detailed in (Table 4).

Discussion

Plant Breeders Act 2016 has been approved and being implemented consequently Plant Breeders Rights Rules (PBRs) has been framed which allows breeders to protect their varieties and prevent malfunctioning of the seed business (Jamil *et al*., 2020a). New plant varieties are registered on the basis of Distinctiveness, Uniformity and Stability (DUS) testing. However, DUS testing based on morphological, phenological and physiological testing (Kanwal *et al*., 2019) is unreliable, as these plant attributes are highly influenced by the environment and non-reproducible. Therefore, now apart from conventional DUS testing, DNA profiling/DNA fingerprinting of varieties is mandatory for registration in Plant Breeders Rights Registry (Iqbal *et al*., 2021a). DNA Fingerprinting can be done through different markers systems i.e., RFLPs, AFLPs, RAPD, SSR and SNPs (Iqbal *et al*., 2021b). However, SSR markers have advantage over other PCR based markers due to uniform genome coverage, reproducibility, and codominance, ease of genotyping and high level of polymorphism (Jamil *et al*., 2020b; Rahman *et al*., 2022).

In the present study 212 SSR markers were used for DNA fingerprinting and genetic diversity studies which had amplified 1314 total alleles and 912 polymorphic alleles

(70% alleles) which exhibited high level of genetic diversity in our plant material in comparison to previous studies (Castellana *et al*., 2020; Al-Shammari *et al*., 2021). Further, PIC in present study ranged from 0.48 (SSR188) to 0.97 (SSR14) and with an average of 0.78 (Table 2) which is very high compared to other studies 0.38 (Castellana *et al*., 2020) and 0.36 (Al-Shammari *et al*., 2021). These evidences have suggested that our genotypes have diverse genetic makeup. This is also proved from cluster analysis which depicted 0.76 to 0.88 genetic similarity coefficients between genotypes (Fig. 1). Cluster and structure analysis results explained that different genotypes i.e. hybrids, OPVs and inbred lines tend to be more similar to each other compared to other types (Fig. 1). The hybrid genotypes were clustered together in both structure and cluster analysis and similar trend was observed for OPVs and inbred lines (Fig. 2). The pedigree parentage of genotypes used in present study also provided strong evidence about extent of genetic diversity between them. Maximum genetic similarity (88%) was observed between Ahmar-hybrid and Naqeeb OPV which might be due to common origin of their parents i.e. Naqeeb OPV was derived from Rio stone-2-2 (Exotic hybrid) whereas Ahmar-hyrbid was constituted from two inbred lines one of which was derived Rio Grande. Salar-F1 and Saandal-F1 shared 87% similar genetic regions which were due to one common parent (8504 inbred line). Similarly Salar-F1 and Sundar-hybrid possessed 84% genetic similarity due to one common parent (8502 inbred line), similar trend was also reported previously (Rahman *et al*., 2022).

One major obstacle for DNA fingerprinting of tomato was choice of SSR markers. There are different genomic databases Sol Genomics Network (https://solgenomics.net/markers/microsats.pl) and KaTomicsDB: Kazusa Tomato Genomics Database (https://www.kazusa.or.jp/tomato/). Number of SSR markers in Sol Genomics Network alone is 4K; one cannot invest so much capital on DNA fingerprinting with 4K SSR markers. Therefore, we proposed a set of 50 SSR markers which possessed maximum number of alleles, polymorphic alleles and PIC value among 13 tomato genotypes (Table 3) for future DNA fingerprinting studies. Further, DNA profile of 13 tomato genotypes developed in this study will be useful for variety registration and its protection under PBRs in PBR registry (Jamil *et al*., 2020; Jamil *et al*., 2021). DNA fingerprinting and genetic diversity studies will also be useful for breeders to decide crossing plan among genetically diverse genotypes to ensure genetic diversity among cultivated varieties in the field to avoid any future epidemic outbreak of biotic or abiotic stress (Jamil *et al*., 2020; Iqbal *et al*., 2021a).

Conclusion

Present study was carried out for DNA fingerprinting and genetic diversity studies among 13 tomato genotypes using 212 SSR markers. A total of 1314 alleles were amplified by 199 SSR markers, 70% of which (912 alleles) were polymorphic. Genotypes were broadly classified to 3 groups based on the results of structure and cluster analysis. Genetic similarity coefficient between genotypes varied from 0.76-0.88. Each genotype was uniquely identified by different SSR markers and a set of 50 most polymorphic SSR markers was proposed for future DNA fingerprinting and genetic diversity studies.

Acknowledgement

The authors are very grateful to the fellow scientist for moral support and technical guidance from time to time. Acknowledgement is also due towards the breeders of Vegetable Research Institute, Faisalabad for provision of plant materials for carrying out this study. The efforts of Mr. Ahmad Shehzad Lab Assistant, Sobia Jabeen, Lab Assistant and Saira Sattar (Research Associate) are also acknowledged for providing physical support for carrying out this study.

References

Alcalá-Gómez, G., J. Pérez-Alquicira, D. Cabrera-Toledo, M. Cortés-Cruz, M. del Pilar Zamora-Tavares and O. Vargas-Ponce. 2022. Genetic diversity and structure in husk tomato (*Physalis philadelphica* Lam.) based on SNPs: a case of diffuse domestication. *Genet. Resour. Crop Evol*., 69: 443-459.

Al-Shammari, A.M.A., G.J. Hamdi, M.A.S. Al-Mahdawi and N.K. Mohammed. 2021. Genetic diversity analysis and DNA fingerprinting of tomato breeding lines using SSR markers. *J. Agri. Sci*., 32: 1-7.

Castellana, S., L. Ranzino, I. Beritognolo, M. Cherubini, R. Luneia, F. Villani and C. Mattioni. 2020. Genetic characterization and molecular fingerprint of traditional Umbrian tomato (*Solanum lycopersicum* L.) landraces through SSR markers and application for varietal identification. *Genet. Resour. Crop Evol*., 67: 1807-1820.

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.

Foti, V.T., A. Scuderi, C. Bellia and G. Timpanaro. 2021. Biofortification of tomatoes in Italy: Status and level of knowledge. *Agric. Econ*., 67: 227-235.

Iqbal, M.Z., S. Jamil, A. Mehmood and R. Shahzad. 2019. Identification of seven olive varieties using rapd molecular markers. *J. Agric. Res.*, 57: 07-14.

Iqbal, M.Z., S. Jamil, R. Shahzad and S.U. Rahman. 2021b. DNA fingerprinting and cultivar identification of olive (*Olea europaea* L.) using SSR markers. *Adv. Life Sci*., 8: 143-148.

Iqbal, M.Z., S. Jamil, R. Shahzad, K. Bilal, R. Qaisar, A. Nisar, S. Kanwal and M.K. Bhatti. 2021a. DNA Fingerprinting of crops and its applications in the field of Plant Breeding *J. Agri. Res*., 59: 13-28.

Jamil, S., R. Shahzad, E. Yasmeen, S.U. Rahman, M. Younas and M.Z. Iqbal. 2020b. DNA fingerprinting of Pakistani maize hybrids and parental lines using simple sequence repeat markers. *Pak. J. Bot*., 52: 2133-2145.

Jamil, S., R. Shahzad, M.Z. Iqbal, E. Yasmeen and S.U. Rahman. 2021. DNA fingerprinting and genetic diversity assessment of GM cotton genotypes for protection of plant breeders rights. *Int. J. Agric. Biol*., 25(4): 768-776.

Jamil, S., R. Shahzad, S. Kanwal, E. Yasmeen, S.U. Rahman and M.Z. Iqbal. 2020a. DNA fingerprinting and population structure of date palm varieties grown in Punjab Pakistan using simple sequence repeat markers. *Int. J. Agric. Biol*., 23: 943-950.

Kanwal, A., Z. Ali, R. Shahzad, M. Makhdoom, I. Ghafoor, S. Saleem, A. Bakhsh, M. Zulkiffal, N. Parveen and A. ur Rehman. 2019. Genetic diversity for grain size and its association with yield components in bread wheat. *Int. J. Biosci*., 14: 112-122.

Kanwal, S., S. Jamil, R. Shahzad, S.U. Rahman and M.Z. Iqbal. 2021. Standardization of different protocols for genomic DNA isolation from *Phoenix dactylifera* L. *Pak. J. Bot.*, 53(5): 1665-1668.

Kaushal, A., A. Singh and A.S. Jeena. 2017. Genetic diversity in tomato (*Solanum lycopersicum* L.) genotypes revealed by simple sequence repeats (SSR) markers. *J. App. Nat. Sci*., 9: 966-973.

Kulus, D. 2018. Genetic resources and selected conservation methods of tomato. *J. Appl. Bot. Food Qual*., 91: 135-144.

Lone, S., K. Hussain, K.Z. Masoodi, S. Narayan, S. Mazahir, M.R. Hussain and H. Kumar. 2021. Distinctness, uniformity and stability testing of various cherry tomato accessions. *J. Pharm. Phyto*., 10: 2776-2788.

Mueller, L.A., T.H. Solow, N. Taylor, B. Skwarecki, R. Buels, J. Binns, C. Lin, M.H. Wright, R. Ahrens and Y. Wang. 2005. The SOL Genomics Network. A comparative resource for Solanaceae biology and beyond. *Plant Physiol*., 138: 1310-1317.

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

Rahman, S., S. Jamil, R. Shahzad, E. Yasmeen, S. Sattar and M. Iqbal. 2022. Genetic diversity and DNA fingerprinting of potato varieties using simple sequence repeat (SSR) *J. Ani. Plant Sci*., 32: 775-783.

Yujian, L. and X. Liye. 2010. Unweighted multiple group method with arithmetic mean. In: 2010 IEEE Fifth International Conference on Bio-Inspired Computing: *Theories and Applications* (BIC-TA). IEEE: pp: 830-834.

(Received for publication 08 December 2022)