Australian Journal of Crop Science AJCS

AJCS 13(04):578-587 (2019) ISSN:1835-2707
doi: 10.21475/ajcs.19.13.04.p1581

Phenotypic diversity and marker-trait association studies under heat stress in tomato (*Solanum lycopersicum* L.)

Muhammed Alsamir[1,3*], Nabil Ahmad[1], Vivi Arief[2], Tariq Mahmood[1], Richard Trethowan[1]

[1]Plant Breeding Institute, Faculty of Agriculture and Environment, University of Sydney, 107 Cobbitty Road, Cobbitty, NSW 2570, Australia
[2]School of Agriculture and Food Sciences, Faculty of Science, The University of Queensland, St Lucia, Queensland 4072, Australia
[3]The Date Palm Research Center, University of Basrah, Basrah, Iraq

***Corresponding author: alsamir.hameed@sydney.edu.au; malsamir@hotmail.com**

Abstract

Tomato is a mild season crop and high temperature stress impacts productivity negatively. However, the development of cultivars with improved heat tolerance is possible as genetic variability has been consistently reported. This study aimed to identify candidate genes that impact various traits under heat stress. Genome-wide association studies (GWAS) were conducted on a diverse set of 144 tomato genotypes collected from various germplasm centers and breeding programs. The genotypes were grown under control and heat stress in poly tunnels having mean temperatures of 30°C and 45°C for two seasons and phenotypic data were collected on seven agro-physiological traits. All individuals were genotyped withthe80K DArTseq platform using 31237 SNP markers. Data were analysed using a mixed model based on restricted maximum likelihood (REML). Pattern analysis of the phenotypic data showed five primary clusters each with genotypes from multiple origins. Based on the genotypic data, three wild tomato genotypes showed a degree of un-relatedness with the other materials as they were distantly located from the rest of the genotypes in the scatter plot. Control treatment data were used to ascertain markers that are exclusively important under high temperature stress. A large number of markers were significantly associated with various traits under heat stress. These included strong marker associations for number of inflorescence/plant (IPP), number of flowers/inflorescence (FPI), fresh fruit weight (FFrW), and electrolyte leakage (EL). High association with EL was found due to two SNPs 7858523|F|0-25:G>A-25:G>A and 4705224|F|0-60:C>G-60:C>G located on Chr 6. Other less pronounced marker-trait associations were observed for plant dry weight (PDW), and number of fruit/plant (FrPP).

Keywords: Genetic diversity, genome-wide association studies, high temperature stress, plant breeding, tomato germplasm.
Abbreviations: IPP_Inflorescence/plant, FPI_Number of flowers/inflorescence, FFrW_Fresh fruit weight, EL_Electrolyte leakage.PDW_Plant dry weight, FrPP_Number of fruit/plant.

Introduction

Tomato (*Solanum lycopersicum* L.) is an important vegetable crop in most regions of the world both for field and greenhouse production. Tomato is also the second largest of the major vegetable commodities produced in Australia (ABS, 2014), however the size of the harvest fluctuates across years. The national production in 2013-14 was 326189 tons, which was a substantial reduction (-28%) compared to the previous year. This was largely due to a reduced production area (-18%) and dry and hot conditions, particularly in New South Wales, Victoria and Queensland (ABS, 2014). Tomatoes ranked 16th in quantity and value for Australian agricultural production in 2010, and 20th in 2011; a year that Australia did not rank within the top 20 tomato producing countries. China, India and the U.S. are the largest producers of tomatoes with China surpassing the U.S. in 1995 and maintaining that position. Countries such as Italy, Egypt, Iran and Turkey produce substantial quantities and global production is increasing (FAOSTAT, 2013).

Tomato production is often exposed to extreme temperatures and the frequency of high temperatures is projected to increase with climate change. Climate change is likely to increase the earth's surface temperature between 1.5 and 11°C by 2100 and this will pose serious problems for plant reproduction (Stainforth et al., 2005; Reddy and Kakani, 2007).When day/night temperatures exceed 26/20°C, tomato fruit set is interrupted leading to significant reductions in yield (Stevens and Rudich, 1978; El Ahmadi and Stevens, 1979; Bartsur et al., 1985; Lohar and Peat, 1998).

Large scale genomic resources can provide insight into the genetics of complex abiotic stresses such as heat stress. Selection for heat tolerance under field conditions provides breeders with general germplasm performance data (Blum, 1988). However, with the rising global temperatures the

need for heat tolerant varieties has increased and new methods of selection should be explored.

Genetic variation in tomato fruit set under high temperature stress is vital for selection under heat stress (Alsamir et al., 2017; Mansour et al., 2009; Giorno et al., 2010). Wild relatives of tomato have been exploited as sources of tolerance to abiotic stresses and diseases. However, it is often challenging to enrich elite lines with genes from wild species while maintaining their agronomical advantage. The complications arise from the polygenic nature of these complex traits (Villalta et al., 2007).Traditional breeding techniques provide inadequate information on the chromosomal regions controlling the polygenic traits (Semel et al., 2006). Selection based only on phenotypic analyses under conditions of large genotype–environment interactions is complex.

Genome-wide association studies (GWAS) are used to map loci responsible for natural variation in a target phenotype (Saidou et al., 2014; Matsuda et al., 2015) and represent an alternative to bi-parental linkage mapping for determining the genetic basis of trait variation. GWAS is based on the identification of significantly associated genetic polymorphisms in large populations (Brachi et al., 2011) and can therefore be integrated with the phenotypic and genotypic data routinely obtained from plant breeding programs. Thus the genotype-to-phenotype relationships, especially for complex multi-genic traits such as tolerance to abiotic stresses, can be determined. While genetic markers have been found for major tomato fruit quality traits (Ruggieri et al., 2014; Zhang et al., 2015; Zhang et al., 2016) and disease resistance (Arens et al., 2010), limited information on markers for heat stress tolerance is so far available(Lin et al., 2006; Xu et al., 2017).

DArT markers have been extensively used for genotyping plant populations in various plant species (Wenzl et al., 2008 Iorizzo et al., 2014; Van Schalkwyk et al., 2012). DArTseq™ represents a combination of DArT complexity reduction methods and next generation sequencing platforms (Kilian et al., 2012; Courtois et al., 2013; Cruz et al., 2013; Raman et al., 2014). Similar to DArT methods based on array hybridizations, the DArTseq™ technology is optimized for each organism and applied by selecting the most appropriate complexity reduction method (both the size of the representation and the fraction of a genome selected for assays).

The objective of this study was to assess genetic diversity in a broad range of tomato materials to identify markers and candidate genes responsible for genetic variation in heat tolerance in tomatoes. This is the first study of its kind which has focused on a large number of tomato traits using SNP markers under heat stress production environments. After validation the marker-trait associations found here may go a long way in tomato breeding to develop materials for heat stress conditions.

Results

Pattern analysis

Phenotypic data

The estimate of genetic variance (V_G) for the number of fruits/plant (FrPP), fresh fruit weight (FFrW), and electrolyte leakage (EL) were small and not significantly different from zero; less than double of its standard error (Table 2).Whereas, V_G for number of inflorescences/plant (IPP), number of flower/inflorescence (FPI), and plant dry weight (PDW) were large and significantly different from zero. There were small and non-significant estimate of variance component for genotype-by-season (V_{GS}) for all traits, except for PDW. These results indicate the lack of seasonal effects for these traits. Whereas, the estimated variance for genotype-by-heat stress (V_{GH}) were significantly different from zero for all traits, except PDW. The values of variance components for genotype-by-season-by-heat stress (V_{GSH}) were smaller than V_{GH} for all traits (Table 2).

Due the small value of V_G, FrPP, FFrW and EL had low line-mean heritability. For these traits, most of the variability was explained by V_{GH} (Table 2). The coefficient of variability (CV) of PDW and EL were larger than the rest (Table 2). These values indicated large variability around the mean.

The lack of seasonal effects were also observed from the results of pattern analysis (Figure 2, a and b). The traits under the same heat stress treatment grouped together across seasons. The effect of heat stress in discriminating genotypes was evident from the results of pattern analysis as traits measured under control and heat stress tended to be in separate groups.

The pattern analysis among traits summarized their correlations (Figures 2, a and b). It seem that the traits could be divided into two groups: one contains fresh fruit weight (FFrW), fruit per plant (FrPP), and number of inflorescences/plant (IPP) and the other contains electrolyte leakage (EL), number of flowers/inflorescence (FPI) and plant dry weight (PDW). Electrolyte leakage under heat stress was more correlated to plant dry weight (PDW), while EL under control was more correlated to FPI.

The biplots showed a group of genotypes with higher values for IPP, FrPP, and FPI (Figure 2b) whereas, another group of genotypes with higher values for FFrW and IPP. There was no obvious grouping based on the origin of the genotypes as the genotypes were seen randomly distributed in the plot.

Markers data

Pattern analysis of the genotype conducted using all markers showed that the wild genotypes USA1 (LA0373), USA2 (LA0716) and USA3 (LA1930) were clearly at a distance from the rest of the materials (Figure 3a). After removing these three genotypes, the majority of the genotypes grouped into two main groups with a smaller number found scattered in the plot (Figures 3, a and b). However, again these sub-groups did not clearly reflect the grouping based on their origins.

Association analysis

Association analysis for each trait was conducted using the phenotypic and marker data of 144 genotypes. The analyses were resulted using 3,625 markers with non-missing associations (Fig 4). These marker-trait associations can be assigned to three categories; i) significant under control only, ii) significant under control and heat stress, and iii) significant under heat stress only. The third category of markers-trait association were mostly identified for number of fruit/plant (FrPP), fresh fruit weight (FFrW), and electrolyte leakage (EL) (Figure 4). Whereas category one and two were mostly identified for number of inflorescent/plant (IPP), number of flower/inflorescence (FPI), and plant dry weight (PDW). These results were in line

Table 1. The wild and cultivated tomato accessions and their origins.

Acc. code	Acc. name	Origin	Acc. code	Acc. name	Origin	Acc. code	Acc. name	Origin
TGRC*			AV 5	VI005672	Australia	PK 14	C chaus	-
USA 1	LA 0373	Peru	AV 6	VI005673	Australia	PK 15	Legend	USA
USA 2	LA 0716	Peru	AV 7	VI005856	Australia	PK 16	Alaskan Fancy	USA
USA 3	LA 1930	Peru	AV 10	VI005897	India	PK 17	Raad Red	USA
USA 4	LA 2375	IL[a]	AV 21	VI006578	Australia	PK 18	Early Wonder	USA
USA 5	LA 2661	IL	AV 23	VI006604	India	PK 19	Polar Beauty	USA
USA 6	LA 3320	IL	AV 24	VI006605	India	PK 20	Zhezha	USA
USA 7	LA 3344	IL	AV 25	VI006606	India	PK 21	Campbell's 1327	USA
USA 8	LA 3345	IL	AV 26	VI006607	India	PK 22	Bonita	USA
USA 9	LA 3847	IL	AV 27	VI006608	India	PK 23	Rio Grande	-
USA 10	LA 3866	IL	AV 28	VI006610	India	PK 24	New Yorker	USA
USA 11	LA 3867	IL	AV 31	VI006613	India	PK 25	Beef Steak	-
USA 12	LA 3869	IL	AV 33	VI006617	India	PK 26	Leeper	-
USA 13	LA 3870	IL	AV 34	VI006618	India	PK 27	LA 2010	USA
USA 14	LA 3871	IL	AV 35	VI006619	India	PK 28	Grus Chovka	USA
USA 15	LA 3874	IL	AV 37	VI006622	India	PK 29	Napoli	USA
USA 16	LA 3875	IL	AV 38	VI006628	India	PK 30	Dona	USA
USA 17	LA3876	IL	AV 41	VI006706	India	PK 31	Pres Cott	USA
USA 18	LA 3878	IL	AV 42	VI006748	India	PK 32	Tai-1042	Syngenta
USA 19	LA 3879	IL	AV 43	VI006749	India	PK 33	Bush Beef Steak	USA
USA 20	LA 3882	IL	AV 44	VI006750	India	PK 34	Cold Set	-
USA 21	LA 3883	IL	AV 45	VI006777	India	PK 35	Naqeeb	Pakistan
USA 22	LA 3886	IL	AV 46	VI006778	India	PK 36	Kaldera	Pakistan
USA 23	LA 3889	IL	AV 48	VI007532	India	PK 37	Caro Rich Tomato	USA
USA 24	LA 3892	IL	AV 49	VI007533	India	PK 38	Forme De Coeur	USA
USA 25	LA 3893	IL	AV 50	VI007534	India	PK 39	NTH-671	Pakistan
USA 26	LA 3906	IL	AV 52	VI007536	India	PK 40	Spekled Siberian	USA
USA 27	LA 4230	IL	AV 53	VI007537	India	PK 41	Northern Delight	USA
USA 28	LA 4231	IL	AV 54	VI007538	India	PK 42	Anahu	USA
USA 29	LA 4232	IL	AV 55	VI008101	India	PK 43	Taxi	USA
USA 30	LA 4233	IL	AV58	VI006777	India	PK 44	Forme De Coeur	-
USA 31	LA 4234	IL	AV59	VI006778	India	PK 45	NTH-671	Pakistan
USA 32	LA 4235	IL	AV60	VI006779	India	PK 46	Spekled Siberian	USA
USA 33	LA 4236	IL	AV81	VI007532	India	PK 47	Northern Delight	USA
USA 34	LA 4237	IL	AV82	VI007533	India	PK 48	Anahu	USA
USA 35	LA 4247	IL	AV83	VI007534	India	PK 49	Taxi	USA
USA 36	LA 4248	IL	AV87	VI007538	India	PK 50	Nagina	Pakistan
USA 37	LA 4249	IL	AV89	VI008108	India	DIGGERS CLUB		
USA 38	LA 4252	IL	AV90	VI008132	India	DIG 1	Costoluto Genovese	-
USA 39	LA 4256	IL	PAKISTAN			DIG 2	Amish Paste	-
USA 40	LA 4257	IL	PK 1	Sasha Altai		DIG 3	Black Cherry	-
USA 41	LA 4272	IL	PK 5	Jaguar	USA	DIG 5	Lemon Drop	-
USA 42	LA 4273	IL	PK 6	Iles Yellow Latvian	Pakistan	DIG 6	Big Rainbow	-
USA 43	LA 4283	IL	PK 7	Zarnitza	USA	DIG 7	Wild Sweetie	-
USA 44	LA 4284	IL	PK 8	Pakit	USA	DIG 8	Violet Jasper	-
AVRDC**			PK 9	UC-134	Pakistan	DIG 10	Sweet Sue	-
AV 1	VI005503	India	PK 10	Bradley	USA	DIG 11	Green Grape	-
AV 2	VI005504	India	PK 11	Long Keeper	USA	DIG 16	Jaune Flamme	-
AV 3	VI005595	Sri Lanka	PK 12	Parter Improved	USA			
AV 4	VI005670	Australia	PK 13	Roma	Pakistan			

*Tomato Genetic Resource Center, University of California, Davis, USA. **World Vegetable Center (Previously known as the Asian Vegetable Research and Development Center), Taiwan

[a] IL: Introgression line.

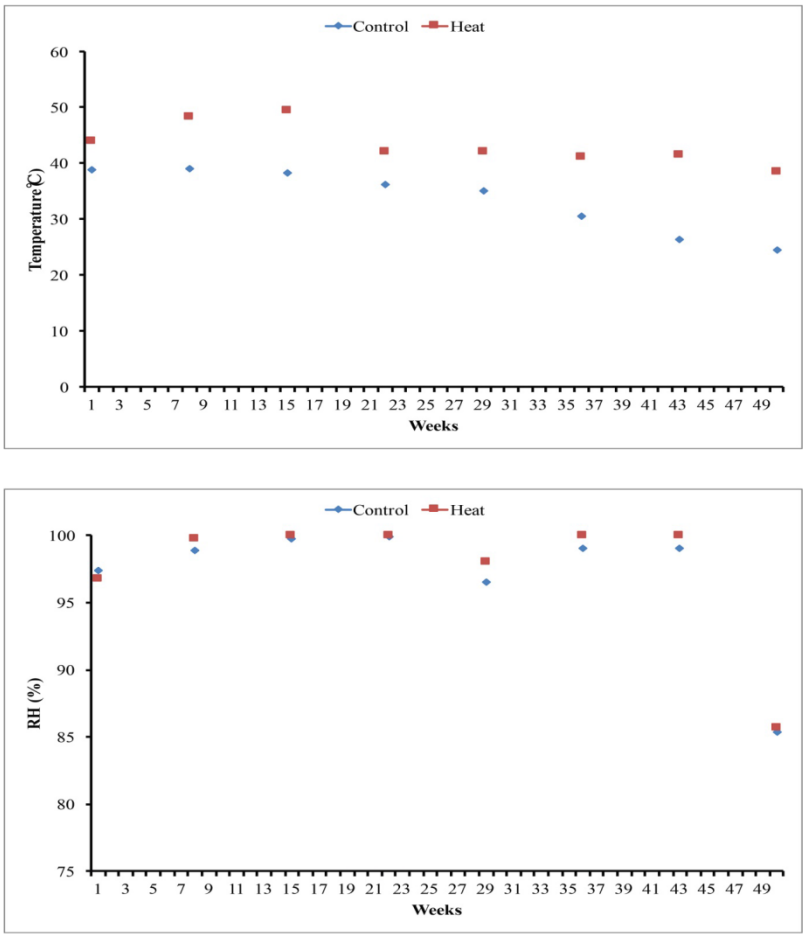

Fig 1. Weekly mean temperature and relative humidity (RH) recorded inside the green house during experimentation.

Table 2. Variance components and standard errors, coefficient of variability (CV), line mean heritability (H^2), and summary statistics for the six traits from the combined analysis across seasons and heat stress treatments.

Source	IPP	FPI	FrPP	FFrW	PDW	EL
Season	0	0	0	0	197.8 (292.1)	0
Heat stress	20.4 (29.4)	0.01 (0.02)	692.1 (983.6)	229, 366.8 (326, 928.3)	0	13.7 (20.4)
Season x Heat stress	0.4 (0.5)	0.01 (0.01)	2.3 (2.4)	195.8 (206.2)	0	0.4 (0.7)
Genotype (V_G)	79.7 (10.4)	1.5 (0.2)	38.0 (31.1)	6, 260.5 (21, 105.5)	3, 144.0 (457.0)	10.4 (6.8)
Genotype x Season (V_{GS})	0	0	0	0	491.5 (156.8)	0
Genotype x Heat stress (V_{GH})	12.1 (2.2)	0.2 (0.05)	322.7 (39.2)	244, 174.0 (29, 069.7)	55.9 (109.4)	50.8 (8.2)
Genotype x Season x Heat stress (V_{GSH})	8.6 (1.0)	0.3 (0.03)	10.8 (1.0)	0	0	19.5 (2.7)
Residue (V_E)	6.4 (0.4)	0.1 (0.004)	2.8 (0.2)	2979.9 (143.8)	3, 119.9 (165.2)	23.6 (1.4)
CV (%)	4.2	6.5	7.7	9.5	39.2	16.6
H2	0.9	0.89	0.19	0.05	0.83	0.24
Range	1.0-88.0	4.0-9.0	1.0-252	0.5-3820	50.0-316	10.2-92.0
Mean	19.06	4.83	21.74	574.52	142.42	29.22
Correlation between Season1 & 2 data	0.96	0.99	0.79	0.87	0.96	0.81

IPP - number of inflorescence/plant; FPI - number of flowers/inflorescence; FrPP - number of fruits/plant; FFrW – fresh fruit weight (g); PDW - plant dry weight (g); EL - electrolyte leakage (%)

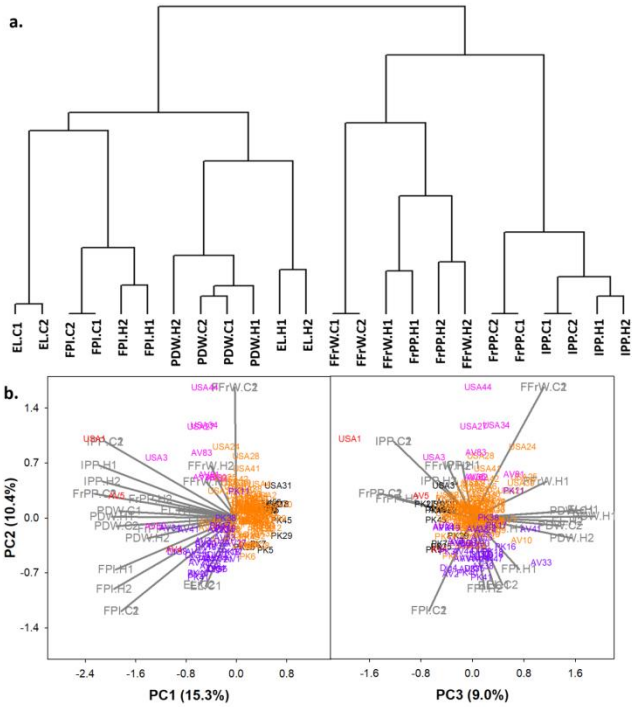

Fig 2. Pattern analysis based on column standardised BLUP of accession by season and heat stress combinations. Pattern analysis was done using squared Euclidean distance as the dissimilarity measure and incremental sum of squares as the clustering strategy. Ordination was done using principal component analysis based on singular value decomposition. (a) optimized dendrogram for season and heat stress combinations; (b) biplots with the origin of the accessions.

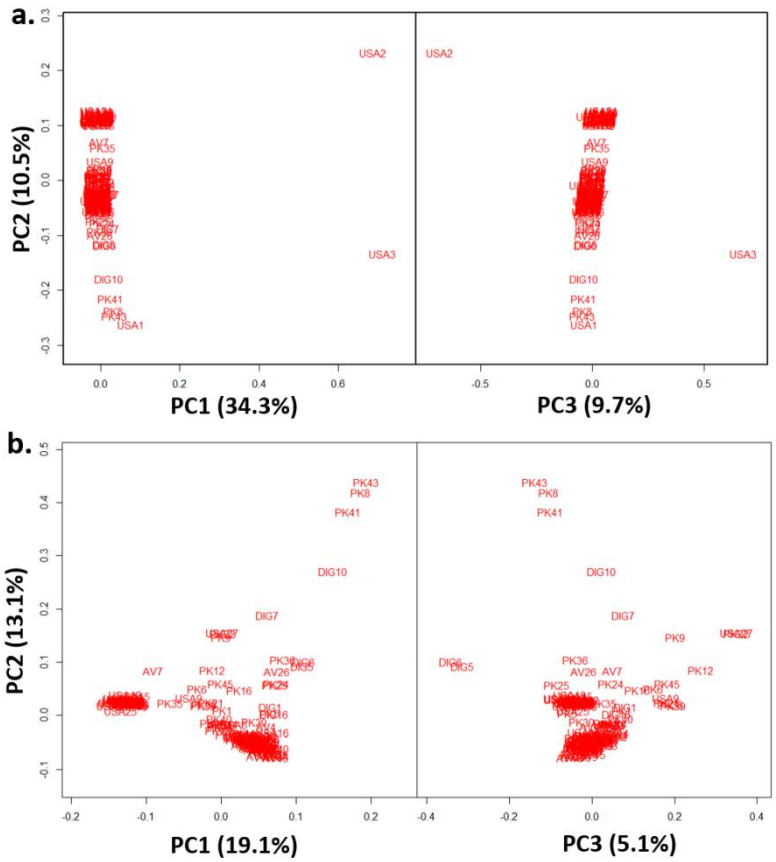

Fig 3. Pattern analysis based on marker data. Pattern analysis was done using complementary dissimilarity measure of simple matching coefficient as the dissimilarity measures and group average as the clustering strategy. Ordination was done using principal coordinate analysis on the simple matching coefficient based on eigen value decomposition. (a) ordination results for all accessions; (b) ordination results after removing three accession (USA1, USA2, and USA3).

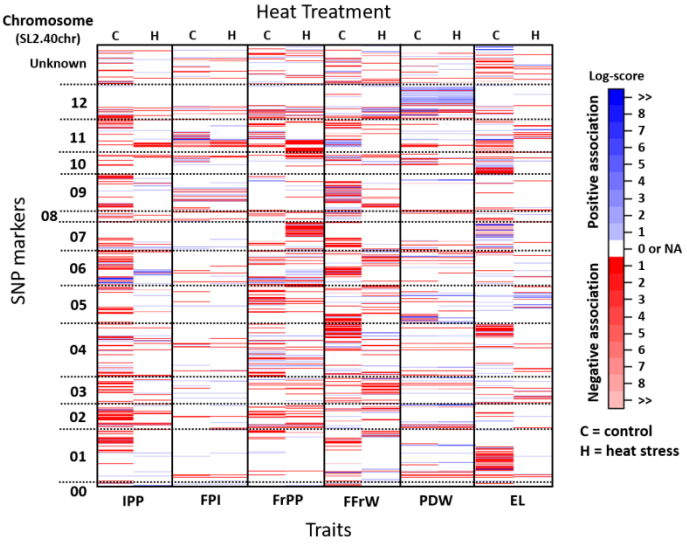

Fig 4. Heatmap for the results of the association analysis for each trait and each temperature treatment. This heatmap was constructed for the 3,625 SNP markers with no missing association. The intensity of the colours indicate the strength of the association (LOG score). Blue indicates that the positive association (i.e. the presence of the marker associated with higher mean of phenotypic value) and red indicates negative associations (i.e. the presence of the marker associated with lower mean of phenotypic value).

with the variance components and pattern analysis on phenotypic data which indicated that these three traits were the ones that discriminated the genotypes in control and heat stress conditions.

Two markers on chromosome 5 (8030576|F|0-25:A>G-25:A>G; 4694305|F|0-8:A>C-8:A>C), three on chromosome 6 (7986869|F|0-58:C>G-58:C>G; 4705011|F|0-8:T>G-8:T>G; 4695679|F|0-14:A>G-14:A>G), and seven on chromosome 12 (4704161|F|0-63: A>G-63:A>G; 7850821|F|0-41:A>G-41:A>G; 7850384|F|0-32:T>C-32:T>C; 7851754|F|0-23:T>C-23:T>C; 7850449|F|0-39:G>A-39:G>A; 7836102|F|0-63:A>G-63:A>G; 4696217|F|0-15:A>T-15:A>T)were positively associated with two traits, PDW and FPI. The markers having high but negative association with EL were found on chromosome 1 and 6 (Supplementary Table 1). The highest marker-trait associations under heat stress treatment were depicted in FFrW, IPP, and FPI whereas the lowest associations were found in FrPP.

Discussion

The plant materials used in this study originated from Southeast Asia, South and North America, and Australia (Table 1). The materials were grown under control and high temperature stress during the growing season in the greenhouse where temperatures frequently exceeded 45^{o}C during the growing season (Figure 1). FFrW, FrPP, and EL were the three traits that differentiate the response of genotypes under different heat treatments (Figures 2, a and b). These three traits showed significant variability for genotype-by-heat treatment (Table 2). Low line-mean heritability for these three traits indicated that larger number of replications may help get the precise information on these traits.

Pattern analysis of the phenotypic data did not group genotypes based on their origin (Figure 2b), indicating that the materials selected from different geographical origins were phenotypically diverse, and may reflect the relatively recent exchange of genetic variability globally in the 20th century. This observation is supported by others (Blanca et al., 2012; Reza and Amri 2013).

As expected, pattern analysis on the marker data showed that the wild genotypes (USA1, USA2, and USA3) were substantially different to cultivated forms (Figure 3). Pattern analysis on phenotypic data also indicated that USA1 and USA3 have high mean value for FrPP and FFrW under heat stress (Figure 2b). Therefore, they may represent an important source of new trait diversity.

The positive marker-trait associations (Supplementary Table 1) were high for FrFW, IPP and FPI whereas relatively low associations were observed for PDW and EL. The association of FrPP with markers was poorly exhibited in these data. The high number of markers that could not be assigned to any chromosome diminished the value of the data. Interestingly, of those unassigned markers a few showed high association with IPP, FPI, and EL (Supplementary Table1). There is a little published evidence of marker distribution under heat stress in tomato (Lin et al., 2006; Xu et al., 2017), however more evidence is available in other species (Levy and Veilleux, 2007; Ye et al., 2015). Marker-trait associations reported here are specific to heat stress environment (Supplementary Table 1, Figure 4) and relevant information on these associations in tomato is not currently available in literature.

While pattern analysis based on marker data indicate the presence of sub-groups (Fig 3). However these sub-groups did not clearly correspond to the origin of the accessions. Further study is required to identify the cause of these sub-groups and how they affect the results of association analysis. While it is important to account for population structure in association analysis, given the size of these data, it would have reduced the power of the analysis (Ranc et al., 2012). The marker-trait associations reported here are indicative of a complex relationship between phenotypic appearance and the genetic markers (Figure 4). Their

potential use in tomato breeding for heat stress tolerance is likely. However, the QTLs being growth condition specific (Bac-Molenaar et al., 2015), these marker-trait associations must be validated in a wider set of materials and environments.

Materials and methods

Plant materials

One hundred and forty four tomato accessions collected from around the world were genotyped and assessed for their heat stress response (Table 1). These materials included one accession each of the wild species *S. pimpinellifolium*, *S. Pennellii* and *S. chilense* and 141 accessions of *S. lycopersicum*, including 11 heirloom varieties. Seeds of 44 accessions were obtained from the Tomato Genetic Resource Center (UC Davis, USA), 43 accessions from the World Vegetable Center (AVRDC, Taiwan), 47 from the Vegetable Research Institute, Faisalabad, Pakistan and 10 heirloom varieties from the Diggers Club, Australia. To maintain genotypic uniformity, the experiments were established using cuttings from the source plants. The materials were grown in a hydroponic greenhouse at The University of Sydney Plant Breeding Institute (Latitude: −34.02, Longitude: 150.67, Altitude: 87m). The experiments were conducted in two greenhouses, one each for control and high temperature stress, for two seasons (normal and late planting) during summer 2014-15. Randomized complete block design (RCBD) with two replications was used in this study.

DNA extraction and quantification

Fresh young leaves of the 144 accessions were collected for DNA extraction (Sahu et al., 2012). The DNA was extracted from 200 mg of fresh leaves using the plant DNA isolation Mini Kit (Bioline, Australia) following the manufacturer's protocols. Quality and quantity of DNA was assessed using 2.0% agarose gel electrophoresis. All samples were checked using a Nanodrop spectrophotometer (NanoDrop® ND-1000 Spectrophotometer, NanoDrop Technologies Inc., and USA) to calculate the ratio of absorbance at 260nm and 280nm. Samples with a ratio of approximately 1.8 were accepted (Desjardins and Deborah 2010).

Genotyping

Genotyping was conducted by Diversity Arrays Technology Pty Ltd. (Yarralumla, Australia) onthe80KDArTseq platform and 31237 SNPs were used. Four methods of complexity reduction were tested in tomato (data not presented) by DArT and the *PstI-MseI* method was selected DNA samples were processed in digestion/ligation reactions principally as per Kilian et al., (2012); however a single PstI-compatible adaptor was replaced with two different adaptors corresponding to two different Restriction Enzyme (RE) overhangs.

The PstI-compatible adapter was designed to include an Illumina flow cell attachment sequence, sequencing primer sequence and a "staggered" and varied length barcode region, similar to the sequence reported by Elshire et al., (2011). The reverse adapter contained a flow cell attachment region and MseI-compatible overhang sequence.

Only "mixed fragments" (PstI-MseI) were effectively amplified. The PCR programme consisted of a denaturation step of 94°C/1 min, followed by 30 cycles of 94°C/20 s, 58°C/30 s and 72°C/45 s, and a final incubation step of 72°C/7 min.

After PCR equimolar amounts of amplification products from each sample of the 96-well microliter plate were bulked and applied to c-Bot (Illumina) bridge PCR followed by sequencing on Illumina Hiseq2500. The sequencing (single read) was run for 77 cycles.

Sequences generated from each lane were processed using proprietary DArT analytical pipelines. In the primary pipeline the fastq files were first processed to filter away poor quality sequences, applying more stringent selection criteria to the barcode region compared to the rest of the sequence. In that way the assignments of the sequences to specific samples carried in the "barcode split" step were very reliable. Approximately 2,500,000 sequences per barcode/sample were identified and used in marker calling. Finally, identical sequences were collapsed into "fastqcoll files". The fastqcoll files were "groomed" using DArT PL's proprietary algorithm which corrects low quality base from singleton tag into a correct base using collapsed tags with multiple members as a template. The "groomed" fastqcoll files were used in the secondary pipeline for DArT PL's proprietary SNP (presence/absence of restriction fragments in representation) calling algorithms (DArTsoft14).

All tags from all libraries included in theDArTsoft14 analysis are clustered using DArT PL's C++ algorithm at the threshold distance of 3, followed by parsing of the clusters into separate SNP loci using a range of technical parameters, especially the balance of read counts for the allelic pairs. Additional selection criteria were added to the algorithm based on analysis of approximately 1,000 controlled cross populations. The Mendelian distribution of alleles was tested in these populations to facilitate the selection of true allelic variants from paralogous sequences. In addition multiple samples were processed from DNA to allelic calls as technical replicates and scoring consistency was used as the main selection criteria for high quality/low error rate markers. Calling quality was assured by high average read depth per locus (the average across all markers was > 30 reads/locus).

Phenotyping

A set of 146 tomato accessions was phenotyped during summer 2014-15 in normal and late planting experiments. The plants were grown inside a hydroponic greenhouse using 10L Coco peat bags as a substrate and fertigated with commercial grade fertilizer recipe. The accessions were evaluated for seven traits including: number of inflorescences/plant (IPP), flowers/inflorescence (FPI), fruits/plant (FrPP), fruit fresh weight (FrFW), plant dry weight (PDW), and electrolyte leakage (EL).

Analysis of variance

Data for each trait were analyzed using mixed model analysis based on Restricted Maximum Likelihood (REML) (Patterson and Thompson, 1971) method implemented in ASREML

(Gilmour et al., 2009). The model fitted to obtain variance components was:

$$y_{ijkl} = \mu + S_j + H_k + (SH)_{jk} + G_l + (SG)_{jl} + (HG)_{kl} + (SHG)_{jkl} + \varepsilon_{ijkl},$$

where y_{ijkl} was the observation; μ was the grand mean; S_j was the effect of season j, where j=1,2; H_k was the effect of heat stress k, where k=1,2; $(SH)_{jk}$ was the interaction effect of season j and heat stress k; G_l was the effect of entry l, where l=1, 149; $(SG)_{jl}$ was the interaction effect of season j and entry l; $(HG)_{kl}$ was the interaction effect of heat stress k and entry l; (*SHG*) jkl was the interaction effect of season j and heat stress k and entry l; ε_{ijkl} was the residual effect. All terms except for μ were fitted as random effects. The analysis was done for each trait.

To evaluate the quality of the experiments, two criteria were used: line mean heritability (*H*) and coefficient of variability (CV). The estimate of variance components were used to calculate line-mean heritability. Line-mean heritability was calculated as the proportion of line-mean phenotypic variance due to genetic variance (Fehr, 1987):

$$H = \frac{V_G}{V_G + \frac{V_{GS}}{s} + \frac{V_{GH}}{h} + \frac{V_{GHS}}{sh} + \frac{V\varepsilon}{rsh}}$$

Where; V_G was the estimated variance due to genotype; V_{GS} was the estimated variance due to the interaction between genotype and season; V_{GH} is the estimated variance due to the interaction between genotype and heat stress; V_{GSH} was the estimated variance due to the interaction between genotype and season and heat stress; V_ε was the estimated residual variance; s was the number of season; h was the number of heat stress; and r was the number of replications. Coefficient of variability (%) was also calculated for each trait.

Pattern analysis protocol

Pattern analysis (Williams, 1976; Cooper and DeLacy, 1994) using phenotypic data was conducted to study the relationship among genotype based on the measures phenotypic values and to study the relationship among the traits within season and heat stress in the genotype discrimination.

A two-way table of genotype by season and heat stress combinations were obtained from REML analysis using the following model:

$$y_{ijkl} = \mu + B_i \,|\, (SH)_{jk} + (SH)_{jk} + (SHG)_{jkl} + \varepsilon_{ijkl},$$

where y_{ijkl} was the observation; μ was the effect of grand mean; $B_i|(SH)_{jk}$ was the effect of block i within planting date j and heat stress k , where i=1,2; $(SH)_{jk}$ was the effect of planting date j and heat stress k; $(SHG)_{jkl}$ was the effect of season j and heat stress k and entry l, where l=1,2,…,144; and ε_{ijkl} was the residual effect. All terms, except $(SHG)_{jkl}$, were fitted as fixed effects. The model was applied for each trait and the results were combined across traits to produce a two-way table of genotype by trait × season × heat treatment.

The resulting two-way table of genotype by trait × season × heat treatment from REML analysis was column standardized and used to calculate squared Euclidean distance (SED) among genotype and among trait × season × heat stress combinations. The SED were used as dissimilarity measure in hierarchical cluster analysis with incremental sum of squares (Ward's method) (Ward, 1963) as the clustering strategy. Ordination was conducted using Principal Component Analysis based on Singular Value Decomposition of the standardised table. The clustering results were displayed using an optimized dendrogram (Ari al., 2012) and the ordination results were displayed using biplot with symmetrical scaling (Kroo ef et nenberg, 1997).

Markers data

To evaluate the relationship among genotypes pattern analysis was also conducted for 31,237SNP markers data. Prior to analysis heterozygous values were set as missing. Clustering was conducted using a complementary dissimilarity measure of a simple matching coefficient. This dissimilarity matrix was calculated as one minus the simple matching coefficient (Hubalek, 1982). Average linkage or UPGMA (Sokal and Michener, 1958; McQuitty, 1967) was used as clustering strategy. Ordination was conducted using Principal Coordinate Analysis based on Eigen decomposition on the simple matching coefficient matrix. Clustering results were displayed using an optimized dendrogram and the ordination results were displayed using a scatter plot.

Association analysis

A two-stage approach was used for association analysis (Stich et al., 2006). Best Linear Unbiased Predictors (BLUPs) were firstly calculated for each trait and then used to calculate the log-score of each marker-trait association. The BLUPs of genotype by heat stress for the association analysis were obtained from REML analysis using the following model:

$$y_{ijkl} = \mu + B_i \,|\, (SH)_{jk} + (SH)_{jk} + (HG)_{kl} + (SHG)_{jkl} + \varepsilon_{ijkl},$$

where y_{ijkl} was the observation; μ was the effect of grand mean; $B_i|(SH)_{jk}$ was the effect of block i within planting date j and heat stress k , where i=1,2; $(SH)_{jk}$ was the effect of planting date j and heat stress k; $(HG)_{kl}$ was the interaction effect of heat stress k and entry l, where l=1,2,…,149; $(SHG)_{jkl}$ was the effect of season j and heat stress k and entry l; and ε_{ijkl} was the residual effect. All terms, except $(SHG)_{jkl}$, were fitted as fixed effects. This REML analysis was run for each trait.

The log-score was calculated for each marker class based on a t-test for BLUP mean difference between the two marker classes. A log score of three was used as a threshold to declare an association significant (Newell et al., 2012). The marker was considered to be positively associated with the trait if presence of the marker contributed to better phenotypic performance and vice versa for negative associations.

Weather data

Temperature, relative humidity and photosynthetically active radiation (PAR) in the green house were recorded using a CR200X Data Logger (CAMPBELL SCIENTIFIC, INC., Australia) as shown in Figure 1.

Conclusion

Genetic variability for heat stress tolerance exists in the tomato germplasm and the variation is distributed well across the continents. Genetic markers under heat stress

conditions are associated with various traits although the strength of their association varies among the traits. Fresh fruit weight, a highly important trait under heat stress conditions, is strongly associated with the genetic markers. Further studies may help identify specific markers associated with fresh fruit weight under those environments.

Acknowledgements

Authors acknowledge the contributions made by the staff and students of the Plant Breeding Institute, The University of Sydney for their help during these studies. Authors also thankfully acknowledge the contributions made by AVRDC, TGRC at UC Davis, and VRI Faisalabad Pakistan by providing the tomato germplasm, and ACIAR vegetables project HORT/2012/002 for supporting the heat stress research on tomatoes.

Supplementary material

Supplementary Table 1. Chromosomal distribution of markers with the highest LOG scores (threshold >3) associated with heat tolerance in tomato.

References

ABS (2014) "Agricultural Commodities, Australia 2013-14 (cat. no. 7121.0) [Online]", in: http://www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/7121.0Main%20Features32013-14?opendocument&tabname=Summary&prodno=7121.0&issue=2013-14&num=&view=. 29/5/2015 ed. (Canberra, Australia: Australian Bureau of Statistics).

Alsamir M, Ahmad NM, Keitel C, Mahmood T, and Trethowan R (2017) Identification of high temperature tolerant and agronomically viable tomato (S. lycopersicum) genotypes from a diverse germplasm collection. Adv Crop Sci Tech. 5: 299.

Arens P, Mansilla C, Deinum D, Cavellini L, Moretti A, and Rolland S (2010)Development and evaluation of robust molecular markers linked to disease resistance in tomato for distinctness, uniformity and stability testing. TheorAppl Genet120(3): 655-664.

Arief VN (2010) Methodology development to integrate phenotypic and genotypic data from a plant breeding program through association analysis with a case study using data from the CIMMYT International Spring Wheat Field Trials. PhD, School of Agriculture and Food Sciences, The University of Queensland.

Arief VN, DeLacy IH, Basford K, and Dieters M (2012) "Optimized dendrogram: extracting population information using seriation to enahance hierarchical clustering", in: 4th ICQG. (Edinburgh).

Bac-Molenaar JA, Fradin EF, Becker FFM, Rienstra JA, van der Schoot J, Vreugdenhil D, Keurentjes JJB (2015) Genome-wide association mapping of fertility reduction upon heat stress reveals developmental stage-specific QTLs in *Arabidopsis thaliana*. Plant Cell 27:1857–1874.

Bartsur A, Rudich J, and Bravdo B (1985) High-temperature effects on co2 gas-exchange in heat-tolerant and sensitive tomatoes. J Am Soc Hortic Sci., 110(4) 582-586.

Blanca J, Can˜izares J, Cordero L, Pascual L and Diez MJ (2012) Variation Revealed by SNP Genotyping and Morphology Provides Insight into the Origin of the Tomato. Plos One 7(10), e48198.

BlumA (1988) Plant breeding for stress environments.

Brachi B, Morris GP, and Borevitz JO (2011) Genome-wide association studies in plants: the missing heritability is in the field. Genome Biol 12(10).

Cooper M, and DeLacy IH (1994) Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. Theor Appl Genet. 88: 561-572.

Courtois B, Audebert A, Dardou A, Roques S, Ghneim-Herrera T, and Droc G (2013) Genome-Wide Association Mapping of Root Traits in a Japonica Rice Panel. Plos One 8(11).

Cruz VMV, Kilian A, and Dierig DA (2013) Development of DArT Marker Platforms and Genetic Diversity Assessment of the US Collection of the New Oilseed Crop Lesquerella and Related Species. Plos One 8(5).

Desjardins P, and Conklin D (2010) NanoDrop microvolume quantitation of nucleic acids. JoVE45. e2565-e2565.

El Ahmadi AB, and Stevens MA (1979) Reproductive responses of heat-tolerant tomatoes to high temperatures. J Am Soc Hortic Sci.(USA).

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, and Buckler ES (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. Plos One 6(5). doi: 10.1371/journal.pone.0019379.

FAOSTAT (2013) "Production statistics [Online]". ([Accessed 19/9/2015]: Food and Agriculture Organisation of the United Nations, Rome, Italy. Avialable: http://faostat.fao.org/site/339/default.aspx).

Fehr WR, Ed (1987)Principles of Cultivars Development. Vol.1. USA, McGraw – Hill, Inc.

GilmourAR, Gogel BJ, Cullis BR, Welham SJ, and Thompson R (2009) ASReml User Guide Release 3.0. Hemel Hempstead, HP1 1ES, UK: VSN International Ltd.

Giorno F, Wolters-Arts M, Grillo S, Scharf K-D, Vriezen WH, and MarianiC (2010) Developmental and heat stress-regulated expression of HsfA2 and small heat shock proteins in tomato anthers. J Exp Bot. 61(2), 453-462.

Hubalek Z (1982) Coefficients of association and similarity, based on binary (presence-absence) data: an evaluation. Biological Review. 57: 669-689.

Iorizzo M, Gao L, Mann H, Traini A, Chiusano ML, Kilian A, Aversano R, Carputo D, and Bradeen JM (2014) A DArT marker-based linkage map for wild potato Solanum bulbocastanum facilitates structural comparisons between Solanum A and B genomes. BMC Genetics. 15 (1), 123.

Kilian B, Ozkan H, Shaaf S, Huebner S, Pasam RK, and Sharma R (2012) Comparing Genetic Diversity within a Crop and Its Wild Progenitor: A Case Study for Barley. Agrobiodiversity Conservation. Securing the Diversity of Crop Wild Relatives and Landraces. 186-192.

Kroonenberg PM (1997) "Introduction to Biplots for G × E Tables". (Brisbane, Australia: Centre for Statistics, The University of Queensland).

Levy D and Veilleux RE (2007) Adaptation of potato to high temperatures and salinity-a review. Am J Pot Res. 84(6), 487-506.

Lin KH, Lo HF, Lee SP, Kuo CG, Chen JT and Yeh WL (2006) RAPD markers for the identification of yield traits in tomatoes under heat stress via bulked segregant analysis. Hereditas. 143, 142-154.

Lohar D, and Peat W (1998) Floral characteristics of heat-tolerant and heat-sensitive tomato (< i> Lycopersicon esculentum</i> Mill.) cultivars at high temperature. Sci Hort. 73(1), 53-60.

Mansour A, Ismail HM, Ramadan MF, and Gyulai G (2009) Variations in tomato (*Lycopersicon esculentum*) cultivars grown under heat stress. Journal für Verbraucherschutz und Lebensmittelsicherheit. 4(2), 118-127.

Matsuda F, Nakabayashi R, Yang Z, Okazaki Y, Yonemaru Ji., and Ebana K (2015) Metabolome-genome-wide association study dissects genetic architecture for generating natural variation in rice secondary metabolism. Plant J. 81(1): 13-23.

McQuitty LL (1967) Expansion of similarity analysis by reciprocal pairs for discrete and continous data. Edu Psycolog Manage. 27: 253-255.

Newell MA, Asoro FG, Scott MP, White PJ, Beavis WD, and Jannink JL (2012) Genome-wide association study for oat (*Avena sativa* L.) beta-glucan concentration using germplasm of worldwide origin. Theor Appl Genet. 125(8): 1687-1696.

Patterson HD, and Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biometrika. 58(3)545-554.

Raman H, Raman, R, Kilian A, Detering F, Carling J, and Coombes N (2014) Genome-Wide Delineation of Natural Variation for Pod Shatter Resistance in Brassica napus. Plos One 9(7). doi: 10.1371/journal.pone.0101673.

Reddy, K.R., and Kakani, V.G. (2007). Screening Capsicum species of different origins for high temperature tolerance by in vitro pollen germination and pollen tube length. Sci Hort. 112(2), 130-135.

Ranc N, Munos S, Xu J, Le Paslier MC, Chauveau A, Bounon R, Rolland S, Bouchet JP, Brunel D, and Causse M (2012) Genome-wide association mapping in tomato (Solanum lycopersicum) is possible using genome admixture of Solanum lycopersicum var. cerasiforme. G3 (Bethesda) 2: 853-64.

Reza M and Amri A (2013) Phenotypic diversity and relationships among a worldwide durum wheat (*Triticum turgidum* L. var. durum) germplasm collection under rainfed conditions of Iran. Crop Pasture Sci. 64(2), 87-99.

Ruggieri V, Francese G, Sacco A, D'Alessandro A, Rigano MM, and Parisi M (2014) An association mapping approach to identify favourable alleles for tomato fruit quality breeding. BMC P Biol. 14(337).

Sahu SK, Muthusamy T, Kandasamy K (2012) DNA extraction protocol for plants with high levels of secondary metabolites and polysaccharides without using liquid nitrogen and phenol. ISRN Molecular Biol. Article ID 205049

Saidou AA, Thuillet AC, Couderc M, Mariac C, and Vigouroux Y (2014) Association studies including genotype by environment interactions: prospects and limits. BMC Genetics. 15(3).

Semel Y, Nissenbaum J, Menda N, Zinder M, Krieger U, Issman N, Pleban T, Lippman Z, Gur A, and Zamir D (2006) Overdominant quantitative trait loci for yield and fitness in tomato. Proc Natl Acad Sci. USA 103, 12981–12986.

Sharanappa KP and Mogali SC (2014) Studies on genetic variability, heritability and genetic advance for yield and yield components in F2 segregating population of tomato (Solanum lycopersicon L.). J Agric Sci. 27 (4), 524-525.

Sokal RR, and Michener CD (1958) A statistical method for evaluating systematic relationships. University of Kansas Science Bulletin. 38, 1409-1438.

Stainforth DA, Aina T, Christensen C, Collins M, Faull N, and Frame DJ (2005)Uncertainty in predictions of the climate response to rising levels of greenhouse gases. Nature 433(7024): 403-406.

Stevens M, and Rudich J (1978) Genetic potential for overcoming physiological limitations on adaptability, yield, and quality in the tomato. HortScience.

Stich B, Melchinger AE, Piepho HP, Heckenberger M, Maurer HP, and Reif JC (2006) A new tes for family-based association mapping with inbred lines from plant breeding programs. Theor Appl Genet. 113, 1121-1130.

Van Schalkwyk A, Wenzl P, Smit S, Cobollo RL, Kilian A, Bishop G, Hefer C, and Berger DK (2012) Bin mapping of tomato diversity array (DArT) markers to genomic regions of Solanum lycopersicum× Solanum pennellii introgression lines." Theor Appl Genet. 124 (5), 947-956.

Villalta I, Bernet G P, Carbonell E A, and Asins M J (2007) Comparative QTL analysis of salinity tolerance in terms of fruit yield using two Solanum populations of F7 lines. Theor Appl Genet. 114(6), 1001-1017.

Ward JH (1963) Hierarchical grouping to optimise an objective function. J Am Stat Assoc. 58, 236-244.

Wenzl P, Huttner E, Carling J, Xia L, Blois H, Caig Kasia V, Uszynska H, Jaccoud D, and Hopper C (2008) Diversity Arrays Technology (DArT): A generic high-density genotyping platform. Safflower: unexploited potential and world adaptability. 7th International Safflower Conference, Wagga Wagga, New South Wales, Australia, 3-6 November. Agri-MC Marketing and Communication.

Williams WT (1976) Pattern Analysis in Agricultural Science. Amsterdam: Elsevier Scientific Publishing Company.

Xu J, Driedonks N, Rutten M J M, Vriezen W H, de Boer G-J, and Rieu I (2017) Mapping quantitative trait loci for heat tolerance of reproductive traits in tomato (*Solanum lycopersicum*). Mol Breed.37(5): 58.

Ye C, Tenorio FA, Argayoso MA, Laza MA, Koh HJ, Redoña ED, Jagadish KS and Gregorio GB (2015) Identifying and confirming quantitative trait loci associated with heat tolerance at flowering stage in different rice populations. BMC Genetics. 16(1), 41.

Zhang J, Zhao J, Liang Y, and Zou Z (2016) Genome-wide association-mapping for fruit quality traits in tomato. Euphytica. 207(2): 439-451.

Zhang J, Zhao J, Xu Y, Liang J, Chang P, and Yan F (2015) Genome-Wide association mapping for tomato volatiles positively contributing to tomato flavor. Front Plant Sci. 6. doi: 10.3389/fpls.2015.01042.