

# Exploring morpho-physiological profiles of a collection of tomato (*Solanum lycopersicum*) germplasm using multivariate statistics

Ifigeneia Mellidou<sup>1\*</sup> , Konstantinos Krommydas<sup>1</sup>, Irini Nianiou-Obeidat<sup>2</sup>, Georgia Ouzounidou<sup>3</sup>, Apostolos Kalivas<sup>1</sup> and Ioannis Ganopoulos<sup>1</sup>

<sup>1</sup>Institute of Plant Breeding and Genetic Resources, HAO ELGO-DEMETER, Themi, Thessaloniki, GR-57001, Greece, <sup>2</sup>Laboratory of Genetics and Plant Breeding, School of Agriculture, Forestry and Natural Environment, Aristotle University of Thessaloniki, Thessaloniki, GR-54124, Greece and <sup>3</sup>Institute of Food Technology HAO-DEMETER, Lycovrissi Attikis, Greece

Received 28 November 2019; Accepted 16 February 2020 – First published online 16 March 2020

## Abstract

A selection of tomato (*Solanum lycopersicum* L.) genotypes with diverse origin and breeding history including 33 landraces, eight modern varieties and two commercial hybrids has been characterized using a set of 25 qualitative descriptors and six quantitative traits. A wide range of variation was evident for the majority of traits, highlighting their utility for characterizing tomato germplasm collections. A plethora of qualitative traits including type of leaf blade, depression and ribbing at peduncle end, fruit shape at blossom end, number of locules and flowering time, as well as measured traits with economic importance such as fruit fresh weight, firmness and total yield per plant, were found to be highly variable within the collection, with a diversity index greater than 0.8. Strong correlations were detected among several traits related to fruit yield and quality. Two-dimensional principal component analysis as well as the unsupervised hierarchical clustering grouped genotypes according to their phenotypic resemblance and morphological characteristics to a great extent. Landraces from different origins were scattered across the whole variation spectrum of PC1 and PC2. A set of six qualitative traits could efficiently discriminate cultivars in PCA (explaining 75% of total variation), suggesting that it can serve as a valuable breeding tool for the germplasm characterization. The evaluation of the phenotypic diversity in the collection as well as the identification of traits that contribute most to heterogeneity have important implications for establishing core collections with high diversity, as well as designing breeding schemes across the Mediterranean basin.

**Keywords:** characterization, descriptors, diversity index, landraces, principal component analysis

## Introduction

The protection of plant biodiversity has become a major priority worldwide, as the emerging necessity to protect genetic resources is closely associated with adaptation to climate change. The drastic reduction of the genetic

diversity of cultivated tomato (*Solanum lycopersicum* L.) has required the conservation and utilization of all the existing genetic resources, including wild relatives and heterogeneous landrace populations (Terzopoulos and Bebeli, 2010; Foolad and Panthee, 2012). Favourable traits such as resistance to pathogens and high productivity have systematically replaced other important traits including tolerance to abiotic stresses and enhanced nutritional value (Gonias *et al.*, 2019). In this context, to create new tomato cultivars with commercial value and adaptability to adverse

\* Corresponding author. E-mail: [ifimellidou@gmail.com](mailto:ifimellidou@gmail.com); [imellidou@ipgrb.gr](mailto:imellidou@ipgrb.gr)

growth conditions, it is imperative that breeders exploit wider germplasm collections (Jin *et al.*, 2019).

A wide selection of such local germplasm is preserved at National and International Seed Banks around the world, which are responsible for the preservation of such – largely untapped – genetic diversity. The ‘in depth’ exploitation and characterization of such underutilized germplasm may serve as a valuable tool to enrich the genetic pool of modern tomato cultivars especially at low-input agricultural systems (Gur and Zamir, 2004; Casañas *et al.*, 2017). The geographic pattern of diversity of tomato landraces in the Mediterranean region has been widely documented (Terzopoulos and Bebeli, 2010; Cebolla-Cornejo *et al.*, 2013; Parisi *et al.*, 2016; Figàs *et al.*, 2018; Lázaro, 2018; Gonias *et al.*, 2019). Typically, landraces serve as a valuable reservoir of genes for fruit quality and abiotic stress tolerance, while they lack pathogen resistance genes (Sacco *et al.*, 2015; Csambalik *et al.*, 2019; Gonias *et al.*, 2019).

Standardized descriptor lists published by either IPGRI (1996), or UPOV (2017) can facilitate seed banks and breeders to compare their morphological characterization datasets across different environments, as many of them are considered as highly informative, with high heritability and low environmental input (Rao *et al.*, 2006; Mazzucato *et al.*, 2008; Figàs *et al.*, 2018). The characterization of phenotypic diversity is a valuable method to discover the phenotypic traits that contribute to the total diversity in a germplasm collection. A powerful statistical technique for analysing genetic relations from morpho-physiological traits is multivariate data analysis (de Oliveira *et al.*, 2012; Mehmood *et al.*, 2014).

Approximately 95 tomato varieties are registered at the Greek National Register of Varieties, but none of them is considered as ‘traditional variety’. By contrast, the National Greek Genebank possess more than 300 accessions of local tomato germplasm from all over the country, representing different ‘gene pools’ adapted to a plethora of diverse climates. In the light of these considerations, a set of 25 morphological descriptors were used to evaluate a collection of tomato germplasm consisted of 33 landraces, eight modern cultivars and two hybrids. Some quantitative traits such as total yield, fruit firmness, fresh weight and dry weight were also evaluated. The assessment of the phenotypic diversity in the collection, as well as the identification of traits that contribute significantly at most to heterogeneity, will be relevant for further germplasm characterization and breeding schemes across the Mediterranean basin.

## Materials and methods

### Plant material

The 33 landraces studied herein originated either from the Greek National Gene Bank, or from the Community Seed

Bank (Peliti) and local producers (online Supplementary Table S1 and Fig. S1). The origin of these landraces was distributed across different districts in Greece (Crete, Aegean islands and region of central Macedonia). Modern varieties were obtained from collaborative Institutes within HAO DEMETER. Each accession/genotype was represented by 15 plants in the collection, with 80 cm distance from row-to-row, and 50 cm from plant-to-plant within the row, organized in a completely randomized design. Plants were grown in the experimental field at the Institute of Plant Breeding and Genetic Resources which is located at 40°32′13″ N latitude, and 23°00′01″ E longitude, at an altitude of 12 m during spring-summer 2018. Conventional horticultural practises including fertilizer application, spraying for common diseases and irrigation were applied (Avdikos *et al.*, 2011; Koutsika-Sotiriou *et al.*, 2016). Weeds were removed manually when necessary.

### Analysis of morpho-physiological traits

To evaluate phenotypic diversity, 25 morphological descriptors were selected based on the protocol for tests on distinctness, uniformity and stability (CPVO-TP/044/4, 2016) (Table 1, online Supplementary Tables S1 and S2). These included leaf attitude (LA), leaf length (LL), leaf width (LW), type of leaf blade (LB), size of leaflets (LS), intensity of leaf colour (ILC), attitude of petiole of leaflet (AP), type of inflorescence (IT), peduncle abscission layer (PA), fruit green shoulder (FGSh), intensity of fruit green colour (FIG), fruit green stripes (FGSt), fruit size (FS), fruit shape in longitudinal section (FSL), fruit ribbing at peduncle end (FR), fruit depression at peduncle end (FD), size of fruit peduncle scar (FSPS), size of fruit blossom scar (FSBS), shape at blossom end (FSB), thickness of fruit pericarp (FTP), number of locules (FNL), fruit colour at maturity (FC), colour of fruit flesh (FCF), time of flowering (FL) and time of fruit maturity (MAT).

As a further step, quantitative traits that are important for tomato breeding including total yield per plant (YD), polar diameter (DP) and equatorial diameter (DE), firmness (FM), fruit fresh (FW) and dry (DW) weight, were also evaluated. Tomato fruits were picked and weighed when individual fruits reached maturity over the entire growing season. Total yield was calculated as the sum of the weights of all fruits for each plant. For the other traits, twenty fruits were randomly harvested at the stage of red ripe for each genotype. Firmness was measured in two opposite points in the equatorial part of each fruit after removal of a thin skin layer using a penetrometer (Maturity Meter 100, ISOELECTRIC Apparacchiature Elettroniche) with an 8-mm tip probe. Dry weight (%) of homogenized fruit tissues was assessed after 4 d dehydration in an oven at 72°C.

**Table 1.** Descriptive statistics, CV %, as well as the estimated phenotypic diversity index ( $H'$ ), for the 25 qualitative traits (CPVO morphological descriptors) evaluated in the 43 tomato genotypes

Trait/CPVO descriptors	Mean	Min	Max	CV %	$H'$
LA	4.89 ± 0.80	3	7	16.5	0.34
LL	4.86 ± 1.16	3	7	23.7	0.79
LW	4.76 ± 1.10	3	7	22.4	0.74
LB	1.59 ± 0.49	1	2	30.5	0.98
LS	4.71 ± 1.37	1	7	28.7	0.64
ILC	4.58 ± 0.82	3	5	18.0	0.47
AP	3.83 ± 0.99	3	5	25.8	0.62
IT	1.27 ± 0.57	1	3	44.7	0.43
PA	8.82 ± 1.19	1	9	13.7	0.16
FGSh	7.15 ± 3.38	1	9	48.0	0.78
FIG	3.65 ± 1.17	1	7	32.4	0.55
FGSt	1.58 ± 1.95	1	9	23.7	0.35
FS	4.07 ± 1.52	1	9	37.8	0.58
FSL	3.70 ± 3.08	1	10	82.0	0.73
FR	5.74 ± 2.80	1	9	49.4	0.96
FD	3.36 ± 1.98	1	7	58.7	0.82
FSPS	5.01 ± 1.74	1	9	35.0	0.76
FSBS	4.78 ± 1.23	1	7	25.3	0.57
FSB	2.54 ± 1.28	1	5	49.4	0.94
FTP	4.77 ± 1.29	3	7	27.4	0.65
FNL	3.31 ± 1.44	1	5	43.8	0.95
FC	4.33 ± 1.01	1	5	23.6	0.47
FCF	4.33 ± 0.81	1	5	19.0	0.52
FL	4.65 ± 1.25	3	7	26.8	0.85
MAT	4.87 ± 1.45	3	7	29.6	0.65

Mean values and standard deviation were calculated on the basis of the 15 individual plants. Full details of the CPVO descriptors can be found in online Supplementary Table S2.

### Data scoring and statistical analysis

Coefficient of variation (CV) was calculated based on the formula  $CV (\%) = (\text{standard deviation}/\text{mean values}) \times 100$ . Mean values and standard deviation were calculated on the basis of the 15 individual plants (qualitative traits) or of the 20 randomly harvested fruits (quantitative traits). For the calculation of the Shannon–Weaver diversity index, all continuous traits were transformed into ordinal. In brief, three discrete ranks were formed by dividing the range of continuous trait into three equal parts, essentially as previously described (Bechere *et al.*, 1996; Terzopoulos and Bebeli, 2010). The Shannon–Weaver diversity index ( $H'$ ) used to characterize the phenotypic frequencies was defined as:

$$H = \sum_{i=1}^n p_i \ln p_i$$

where  $n$  is the number of phenotypic classes for a trait, and  $p_i$  is the proportion of the total number of entries in the  $i$ th class (Mengistu *et al.*, 2015). The relative phenotypic diversity index  $H'$  was calculated by dividing each value of  $H$  by  $H_{\max} = \ln(n)$  in order to express the values of  $H'$  in the range of 0–1. The diversity index was considered as low ( $0.10 \leq H' \leq 0.40$ ), intermediate ( $0.40 \leq H' \leq 0.60$ ) or high ( $H' \geq 0.60$ ) according to Eticha *et al.* (2005).

The classification of landraces and modern varieties on the basis of both qualitative and quantitative traits was performed using multivariate principal component analysis (PCA) in XLSTAT software (2019). Factor loadings greater than 0.55 were regarded as significant. Non-parametric Spearman correlations were used to analyse qualitative traits, while parametric Pearson correlations were used to analyse quantitative traits (Mehmood *et al.*, 2014). The construction of two-dimensional (2-D) plots was based on the first two principal components (PCs). Agglomerative hierarchical clustering (AHC) analysis was performed using Euclidean distance, and Ward's method, to systematically analyse and visualize the genetic dissimilarity for the full set of data.

## Results

### Morphological characterization of the collection based on qualitative traits

In total, a collection of 33 diverse local and eight modern varieties from different districts of Greece, as well as two hybrids, were used in the present study (online Supplementary Table S1 and Fig. S1). The 25 qualitative traits (based on CPVO descriptors) and the descriptive statistics of average, minimum/maximum values, as well as CV are presented in Table 1. Within this tomato collection, the majority of cultivars was characterized as medium-large fruited (51%) with indeterminate growth (77%) (online Supplementary Table S1). Additionally, most of the selected CPVO descriptors were polymorphic and showed more than two phenotypes, apart from LB (pinnate and bipinnate), PA, FGSh and FGSts (absent or present) (Table 1, online Supplementary Table S3). Several traits had high CVs, including FSL (82.0%), FD (58.7%), FR (49.4%), FSBS (49.4%), FGSh (48.0%), IT (44.7%) and FNL (43.8%). These results are indicative of the extensive morphological variability within the collection of Greek tomato landraces. By contrast, qualitative traits such as PA (13.7%), LA (16.5%), ILC (18.0%) and FCF (19.0%) displayed the lowest CV, suggesting that these traits are less informative within the population studied herein.

The estimated diversity ( $H'$ ) for individual traits ranged from 0.16 (PA) to 0.98 (LB), with an average of 0.65 (Table 1), indicating a high phenotypic diversity. Among the qualitative traits, traits showing high levels of polymorphism and noticeable diversity included FR ( $H' = 0.96$ ), FNL ( $H' = 0.95$ ), FSB ( $H' = 0.94$ ), FL ( $H' = 0.85$ ) and FD ( $H' = 0.82$ ). Most of the accessions were equally dispersed between pinnate (40.9%) and bipinnate (59.1%) LB. A similar trend was observed for FR, FNR and FSB, for which the collection has been nearly proportionally classified in five groups. Furthermore, most of the accessions had medium LL (65.9%), and mainly uniparous IT (78.9%). Lower levels of diversity were observed for LA ( $H' = 0.34$ ) and FGSt ( $H' = 0.35$ ). Indeed, the majority of germplasm had no PA (97.7%), a horizontal LA (83.9%) and no FGSt (93.4%). Half of the accessions had pink (41.8%) or red (48.0%) FCF. Regarding the earliness of flowering and fruit maturity, most of the landraces were characterized as medium (58.2 and 47.3%, respectively).

Strong positive correlations ( $P < 0.01$ ) were observed among several qualitative traits (online Supplementary Table S4). In particular, high positive correlations were detected between LW and LS (0.823), FR and FNL (0.829), FR and FD (0.819) as well as FD and FNR (0.806). Additionally, significant correlations were observed between FSPS and FSBS (0.727), FSL and FSB (0.726), LL and LW (0.678), LL and LS (0.662), FS and FNR (0.649) and FS and FR (0.629). On the contrary, high negative correlations were spotted between FSL and FD (−0.822), or FR (−0.707) or FNR (−0.686). Other significant negative correlations were observed between FD and FSB (−0.686), FTP and FNR (−0.663) or FR (−0.649). It should be noted that most of the correlations between traits were found to be significant at  $P < 0.05$ .

### **Morphological characterization of the collection based on quantitative traits**

In this study, six quantitative traits were also measured to characterize the tomato collection, and their descriptive statistics are shown in Table 1. Traits such as FM, FW and YD displayed high CV (65.7, 56.1 and 47.1%, respectively), revealing extensive morphological diversity within the collection. On the contrary, DW had the lowest CV (12.6%), ranging from 4.0% (GRC 488/04) to 7.3% (GRC 442/04). High phenotypic diversity was observed for FM, with a range from 0.1 kg/cm<sup>2</sup> (K14, also known as 'Kerasotomata', a cherry-type tomato landrace) to 1.8 kg/cm<sup>2</sup> (K5, also known as Pink tomato, a large-fruited landrace). Notably, with regard to YD, it was evident that some landraces had comparable fruit production with the hybrids. Specifically, GRC 492/04, HL 229/07 and

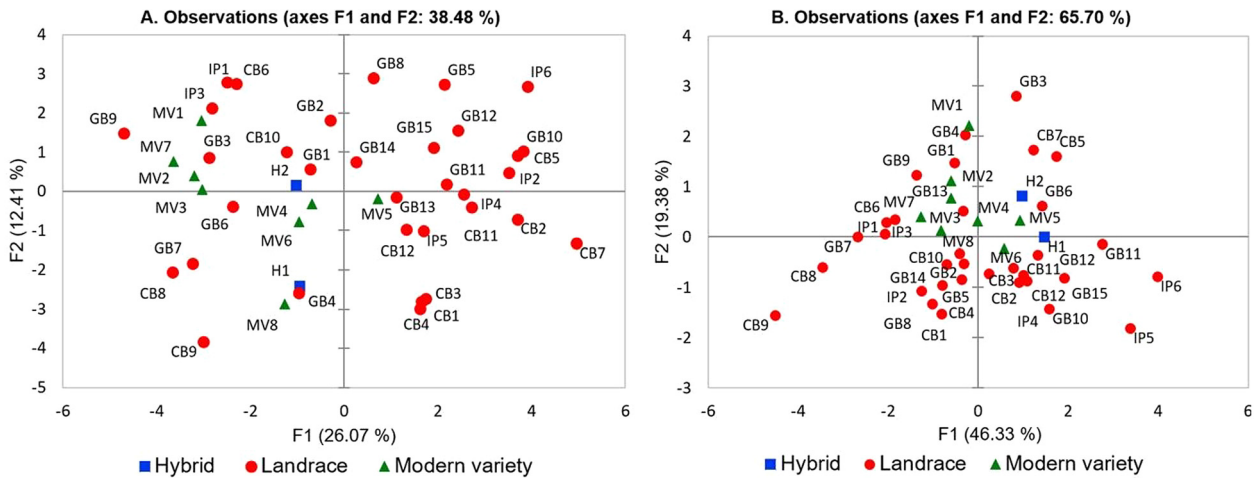
'Psomotomata' (marmande type) had on average total fruit yield greater than 4 kg per plant, similar to the hybrids, at least under our experimental and environmental conditions.

The estimated diversity ( $H'$ ) for individual traits after normalization was relatively high ( $H' > 0.60$ ) for all the quantitative traits measured in this study, with an average of 0.74 (Table 1). Total YD and DE showed the highest levels of polymorphism with  $H'$  values of 0.86 and 0.84, respectively. Moderate levels of diversity were observed for DW ( $H' = 0.62$ ) and FM ( $H' = 0.63$ ). The majority of accessions had medium DP and DE (71.8 and 62.3%, respectively), small FW (64.3%), medium DW (77.5%), low FM (75.2%) and medium YD (52.1%). In online Supplementary Table S5, correlation coefficients among quantitative traits within the whole collection of tomato landraces are displayed. Strong positive correlations were detected between fruit DE and FW (0.921), and to a lesser extent between YD and DE (0.678) or FW (0.604).

### **Principal component analysis of qualitative parameters**

In order to identify the most important traits included in our developed database we used PCA analysis, which allows the multivariate statistical analysis of a dataset. The methodology of Ganopoulos *et al.* (2015) was used because of the large number of variables. The distribution of varieties based on the PC-1 and PC-2 reveals part of the underlying phenotypic variation (35.7%) among the varieties used and showed the magnitude of their dispersion along both axes (Fig. 1a). The Kaiser's criterion ('Eigenvalue' >1) (Kaiser, 1958) was satisfied by seven components (total variation explained 74.7%; Table 2). Specifically, the first component, which accounted for 26.1% of total variation, included many fruit-related traits such as FR and FD, FNL, FSL, FS, FTP and FSB. Leaf-related traits such as LW and LL, as well as LS were the most important traits contributing to the second principal component. The third component explaining 9.67% of total variation included FGSh, FSPS and FGSt. In principal component four, which described 7.94% of total variation, FL and LA showed large contributions, whereas PA and AP accounted for much of the total variance in principal component five. The last two significant principal components, sixth and seven, explained 6.36 and 5.58% of total variation, respectively. The sixth component was mostly determined by traits such as LB and FSPS, while the seventh component by FCF and LB.

It was of interest to conduct a multivariate analysis to determine the existence of clusters that discriminate landraces from modern varieties and hybrids. A 2-D plot based on the first two components grouped landraces, modern varieties



**Fig. 1.** 2-D PCA plot of the first two components of the 43 tomato genotypes based on the 25 qualitative traits (a) or the six quantitative traits (b).

**Table 2.** Descriptive statistics, CV %, as well as the estimated phenotypic diversity index ( $H'$ ), for the six fruit-related quantitative traits evaluated in the 43 tomato genotypes

Trait	Units	Mean	Min	Max	CV %	$H'$
DM	cm	5.73 ± 1.34	2.3	9.4	23.5	0.71
DE	cm	6.45 ± 1.80	2.3	10.3	27.4	0.84
FW	g	160.8 ± 107.9	8.6	458	56.1	0.77
DW	%	5.78 ± 0.73	4.0	7.3	12.6	0.62
FM	kg/cm <sup>2</sup>	0.67 ± 0.38	0.1	1.8	65.7	0.63
YD	kg	2.56 ± 1.21	0.2	5.5	47.1	0.86

Mean values and standard deviation were calculated on the basis of the 20 randomly harvested fruits.

and hybrids according to phenotypic resemblance and qualitative characteristics (Fig. 1a). Landraces from different origins (Greek National Gene Bank, Community Seed Bank, Independent Producers) were generally scattered covering almost the whole variation spectrum along PC1 and PC2, and could not form a distinct cluster that can be separated from the other groups. Several outliers were identified including GB9 (GRC 490/04), IP6 (Katerini), CB7 (K9) and CB9 (K14). Landrace ‘Katerini’ produces extremely large fruits, marmande-type, while K14 or also known ‘Kerasotomata’, has extremely small (cherry) fruits.

Notably, when we filtered our variables according to the principle of Kaiser criterion (eigenvalue >1) retaining the six ones with the highest values of diversity index ( $H' > 0.8$ ) (according to Table 3; LB, FR, FD, FSB, FNL and FL), the first two PCs sufficed to explain as much as 74.5% of variation (online Supplementary Table S6). The discrimination of germplasm based on the evaluation of only six parameters can serve as a valuable pre-breeding tool to characterize tomato germplasm.

### Principal component analysis of quantitative parameters

The dispersal of accessions based on the first two principal components represents the existing phenotypic variation within the collection and shows how widely they are spread across the axes (Fig. 1b). By using eigenvalue greater than 1, it was possible to reduce the dimension of quantitative traits to three components, explaining 83.7% of total variation (Table 4). The first PC, which accounted for 46.3% of total variation, was strongly correlated with DE and FW. Traits such as FW and DP were larger contributors to the second component, explaining 19.4% of total variance, while fruit DW was the major contributor to the third component accounting for 18.0% of total variance within the collection. A 2-D PCA plot was constructed based on the first two components contributing to 65.7% of total variation (Fig. 1b). In general, landraces belonging either to National Gene Bank or Community Seed Bank and Independent Producers were less dispersed compared to

**Table 3.** First seven components from the PCA analysis of 25 qualitative traits studied in the 43 tomato genotypes

Trait/CPVO descriptors	F1	F2	F3	F4	F5	F6	F7
LA	-0.219	0.116	-0.058	-0.605	0.317	0.243	-0.204
LL	0.109	0.779	-0.219	-0.036	-0.158	-0.140	-0.181
LW	0.071	0.866	-0.160	0.150	-0.109	0.197	-0.175
LB	-0.258	0.315	-0.141	0.051	0.291	0.452	0.535
LS	0.317	0.789	-0.153	-0.048	-0.121	0.174	-0.278
ILC	-0.525	0.177	-0.190	0.222	-0.185	0.343	0.065
AP	-0.050	-0.002	-0.117	-0.345	0.527	-0.376	-0.163
IT	0.587	0.179	0.353	-0.264	0.099	-0.185	-0.228
PA	0.058	0.227	0.489	0.213	0.600	0.098	0.035
FGSh	0.305	-0.127	0.646	0.312	0.008	0.170	-0.404
FIG	-0.108	-0.314	0.064	-0.461	-0.451	0.181	-0.057
FGSt	-0.102	-0.172	-0.520	-0.169	0.454	0.044	-0.346
FS	0.711	0.369	-0.125	0.177	0.120	0.020	0.085
FSL	-0.782	0.192	0.127	-0.005	-0.117	-0.362	0.087
FR	0.909	-0.065	-0.046	-0.050	-0.077	0.123	-0.115
FD	0.869	-0.298	-0.034	-0.097	0.005	0.188	0.003
FSPS	0.564	0.124	-0.547	0.014	-0.056	-0.451	0.136
FSBS	0.641	-0.138	-0.466	-0.088	-0.196	-0.300	0.207
FSB	-0.708	0.049	-0.067	-0.050	0.262	-0.330	-0.149
FTP	-0.711	0.038	-0.369	0.368	0.101	-0.012	-0.068
FNL	0.868	-0.110	-0.071	-0.086	0.188	0.148	0.047
FC	-0.052	-0.588	-0.409	0.342	0.105	0.190	-0.095
FCF	-0.053	-0.294	-0.502	0.347	-0.024	0.217	-0.534
FL	0.263	-0.038	0.261	0.643	-0.061	-0.426	-0.188
MAT	0.531	0.034	-0.090	0.348	0.377	-0.019	0.330
Eigenvalue	6.517	3.102	2.419	1.986	1.676	1.590	1.396
Variability (%)	26.07	12.41	9.67	7.94	6.70	6.36	5.58
Cumulative %	26.07	38.48	48.15	56.10	62.80	69.16	74.74

the PCA plot that was based on qualitative data (Fig. 1a). Yet, some outliers were also identified in this case, including IP6 (Katerini) and CB9 (K14), similar to the

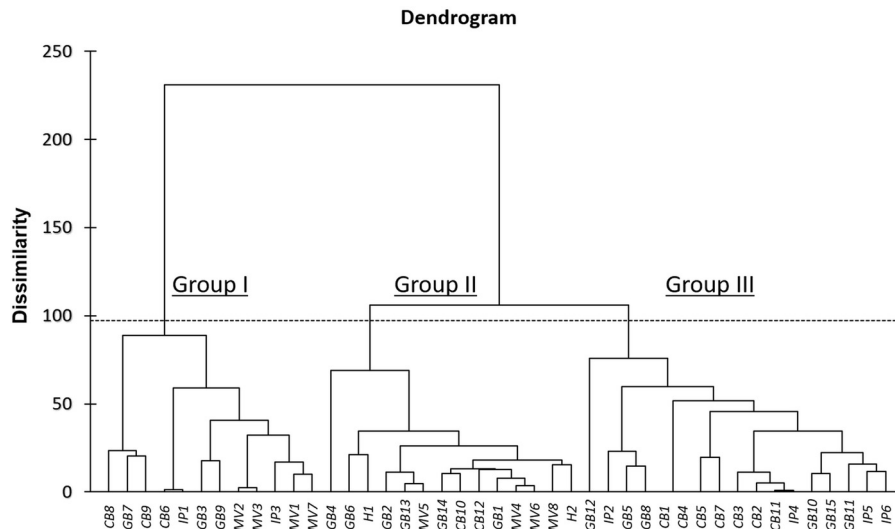
**Table 4.** First three components from the PCA analysis of six quantitative traits studied in the 43 tomato genotypes

Trait	F1	F2	F3
DM	0.610	0.610	0.111
DE	0.915	-0.279	-0.032
FW	0.271	0.706	-0.553
DW	0.922	-0.187	0.032
FM	0.217	0.329	0.862
YD	0.773	-0.267	-0.136
Eigenvalue	2.780	1.163	1.082
Variability (%)	46.326	19.377	18.026
Cumulative %	46.326	65.704	83.730

PCA plot on qualitative traits, that both have extreme fruit sizes. In the same context, GB3 (GRC 442/04) and IP5 ('Psomotomata') have extreme fruit shapes (cylindric/parallel and flattened, respectively). Interestingly, the majority of modern varieties and both hybrids were placed closed to the cross-section of the axes, indicating that these genotypes had the lowest influence on total variation within the collection.

### Cluster analysis

As a further step to investigate phenotypic diversity within out tomato collection, AHC analysis on combined qualitative and quantitative traits was used to enable grouping of the 43 accessions into clusters of increasing genetic dissimilarity, using the Euclidean distance and Ward's method for agglomeration. The resulting dendrogram revealed three distinct groups: Group I is comprised 12 genotypes,



**Fig. 2.** Dendrogram of the 43 tomato genotypes based on both the 25 qualitative and the six quantitative traits.

Group II of 14 genotypes and Group III of 17 genotypes (Fig. 2). Notably, Group III only contains landraces, whereas most of the modern varieties and the two hybrids were clustered in Group II. In accordance with PCA results, there were no specific clusters based on the source of origin or locality. Another interesting note, GB4 (GRC443/04) and GB12 (HL 123/7), were largely phenotypically dissimilar from the other genotypes within the same cluster. These two landraces originated from Chania and Iraklion, respectively, indicating potentially valuable genetic pools of phenotypic diversity.

## Discussion

The current global agenda poses new challenges for both producers and researchers to develop high-quality agricultural products, with lower inputs (agrochemicals and low carbon footprint) (Parisi *et al.*, 2016; Bai *et al.*, 2018). However, it is a common phenomenon that seed companies have no commercial interest in preserving such largely untapped germplasm including local cultivars/landraces. In Greece, for example, there are no landraces registered at the National Catalogue of Varieties Vegetables. In this context, landraces that are widely cultivated by local producers have mainly survived due to their exceptional flavour and/or taste, besides their usually peculiar morphological or organoleptic characteristics. The genetic pool of such largely unexploited germplasm can be of utmost importance in developing breeding strategies, as modern varieties and hybrids have considerable less flavour when compared to 'so-called' old cultivars. Consequently, local producers play a vital role not only in preservation of such

germplasm, but also in landrace improvement, and incorporation into breeding schemes (Campanelli *et al.*, 2019).

The aim of the present work was to provide information for the phenotypic diversity and potential utilization of a collection of tomato landraces grown in Greece. The collection, consisted of 33 landraces, eight modern varieties and two commercial hybrids, was investigated by using several economically important traits, in particular, those related to fruit yield, fruit morphology and fruit quality, to determine if such characteristics could be useful to detect valuable materials for breeders and growers (Ortiz Ríos, 2015). Results support the notion that some of these traits can be efficiently used in evaluating and discriminating cultivars, highlighting the extensive diversity across the collection. Overall, the present work confirmed findings on several features from other tomato collections of Southern Europe (Terzopoulos *et al.*, 2009; Terzopoulos and Bebeli, 2010), or other local cultivars around the world (Mazzucato *et al.*, 2008; Figàs *et al.*, 2015, 2018; Parisi *et al.*, 2016).

Tomatoes in Greece are usually cultivated in small open-field systems requiring plant support and pruning by tutors. In this respect, the studied collection herein mainly displayed indeterminate (according to CPVO), or indeterminate to semi-determinate growth habit (according to IPGRI, 1996), similar to previous reports (Terzopoulos and Bebeli, 2010; Lázaro, 2018). Furthermore, the majority of the selected accessions, which in principal originated from different parts of Crete, showed a wide range of fruit shapes and colour (online Supplementary Fig. S1), probably indicating local preferences and adaptability (Lázaro, 2018). For example, fruit shape was distributed into 9 out of 11 groups (according to CPVO classification system), with flattened, oblate, circular and pyriform shapes being the dominant

ones across the collection. Another trait which is associated with fruit quality is the size of blossom end scar. Typically, large scar can reduce the market value of the fruit, as a result of pathogen attack or juice leakage (Elkind *et al.*, 1990). Consistently to other tomato landrace-related studies (Terzopoulos and Bebeli, 2010), the majority of accessions (79%) had medium to large size of blossom end scar, indicating the lower commercial value of the product, which could probably be balanced, however, with the better flavour that consumers require. Other traits common to landrace populations, already described by others (Terzopoulos and Bebeli, 2010; Lázaro, 2018), included the presence of green shoulder, and the occurrence of slightly flattened or pear-shaped tomatoes. It is evident that local farmers around the Mediterranean basin are interested in variability in fruit shapes, colours and sizes (Mazzucato *et al.*, 2010; Terzopoulos and Bebeli, 2010), rather than total yield.

Among the studied quantitative characteristics, fruit weight and yield are also of high commercial value (Mehmood *et al.*, 2014). In this study, fruit weight values had a range from 8.6 to 458 g, which is higher compared to other collections (Cebolla-Cornejo *et al.*, 2013; Parisi *et al.*, 2016; Figàs *et al.*, 2018). Landraces such as 'Katerini' (IP6), HL 073/07 (GB11) and 'Psomotomata' (IP5) had the highest fruit weight (>380 g), whilst others such as K8, K9, K10 and GRC 451/04 had the lowest ones (<50 g). Fruit dimensions are also important in packaging and transportation (Padilla-Ramirez *et al.*, 2012). Our results revealed significant phenotypic diversity, as indicated by the diversity index for both fruit length and diameter (0.71 and 0.84, respectively). Furthermore, the majority of accessions within the collection had low FM (0.67 kg/cm<sup>2</sup> on average) compared to other reports (Figàs *et al.*, 2018), although some landraces (GRC 443/04, K7 and K9) had higher FM compared to modern varieties or hybrids. Notably, all of the 'firm' cultivars have determinate type of growth, while K7 and K9 also have pink colour of fruits. This finding highlights the potential use of such cultivars in processing industry that requires plants with determinate growth type and fruits that tolerate mechanical damage. High temperature can affect fruit colour (lycopene synthesis), vitamin C accumulation and biosynthesis (Mellidou and Kanellis, 2017) and total yield through barriers in pollen viability and fruit set (Paupière *et al.*, 2017). Indeed, the high summer temperature in Greece for the entire growing period may account for the large percentage of light-red fruit colour in the collection (online Supplementary Fig. S1), as lycopene biosynthesis is strongly inhibited by high temperatures (Steven and Rick, 1986).

Correlations among quantitative or qualitative traits revealed interesting relationships between characteristics. For instance, very strong correlations ( $P < 0.01$ ) were detected among several traits related to fruit yield and quality,

including FSK and FD, FNR and FR or FD, FW and DE or DP, YD and DE or FW. Overall, these results agree with those of previous reports for many traits (Cebolla-Cornejo *et al.*, 2013; Cortés-Olmos *et al.*, 2015; Parisi *et al.*, 2016). The strong correlations between traits contributing to fruit yield and quality could be efficiently used as a selection parameter for the rapid assessment of tomato accessions that have, for instance, many locules, or high yield, and might be helpful for plant improvement.

Phenotypic diversity arises not only from different genetic backgrounds, but also due to environmental conditions, especially for traits related to nutritional value (Csambalik *et al.*, 2019). In principal, heterogeneity of tomato landraces related to fruit yield is expected to be wide between different climates, soil fertilization and irrigation regimes. Indeed, total yield per plant was significantly lower compared to other studies (Lázaro, 2018). Presumably, these differences are due to both genotypic differences and adaptability to local conditions and cultivation systems. Although the evaluation of our collection was only performed during 1 year, and thus the environment  $\times$  genotype interaction could not be assessed, the analysis of these morphological traits was still valuable to characterize our collection, as it gave us an estimation of the wide variability present, pointing out at genotypic effects between landraces. Indeed, low and high yield genotypes have been identified, and in fact, some landraces had even higher yield than hybrids and modern varieties.

Undoubtedly, some of the local accessions represent a valuable material either to be incorporated into breeding schemes as the starting material, or to further be characterized in order to be registered as landraces at the National Catalogue of Vegetables. In order to efficiently identify the most promising germplasm, we have previously carried out the molecular screening of such local germplasm for resistance to *Fusarium oxysporum* f.sp. *radicis lycopersici* and to *Phytophthora infestans* by employing molecular markers coupled with high resolution melting analysis (Gonias *et al.*, 2019). Although almost all of them were found to be heterozygous resistant for both diseases, a landrace from the National Gene Bank (GRC 488/4) was found to be homozygous resistant to *Phytophthora infestans*. This genotype produced small fruits with oblate shape and relative low values of FM and DW. Therefore it could only be used as a donor for resistance genes in mating schemes.

Unravelling inter- and intra-population diversity, as well as identifying the key components that contribute the most to such diversity, is of utmost importance when evaluating landrace heterogeneity (Terzopoulos and Bebeli, 2010). Morphological traits such as FSL, FD and FR, FSB, FGSh and FNR were identified as key traits in discriminating tomato landraces. This is consistent with other reports



concerning phenotypic diversity in tomato populations (Cortés-Olmos *et al.*, 2015; Lázaro, 2018). Nevertheless, unlike to other reports indicating that vegetative traits contribute less to phenotypic diversity (Terzopoulos and Bebeli, 2010), in our set of cultivars, traits with the lowest value of the diversity index included the fruit-related traits peduncle abscission layer and fruit green stripes (most fruits – over 90% – had abscission layer and absence of green stripes).

The PCA results show that the most appropriate traits for grouping the cultivated tomato germplasm are FR and FD, FNL, FSL, FS, FTP, FSBS and FW. Landraces from different origins were scattered across the whole variation spectrum of PC1 and PC2. Morphological and genetic characterization of eggplant germplasm originated from Greece revealed that the accessions from mainland were differentiated from those from islands, underlying the importance of the environment, the geographic location and the agro-ecological conditions (Gramazio *et al.*, 2019). No similar correlations were identified in this study, probably due to the small number of landraces originated from the mainland. Notably, by evaluating the six traits with the highest values of the diversity index ( $H' > 0.8$ ), the first two principal components could efficiently discriminate cultivars, suggesting that it can serve as a valuable breeding tool for the characterization of a collection.

For qualitative traits, the first three axes explained 48.2% of total variation. This is in agreement with previous studies reporting relatively low percentages of variation that could be attributed, presumably, to the high intra-population heterogeneity (Terzopoulos and Bebeli, 2010). Nevertheless, cluster analysis clearly separated tomato germplasm into three groups, with the third cluster containing only landraces, and the second one mostly modern varieties and hybrids. Results from both PCA and cluster analysis revealed several genotypes as potential optimal parents for the development of heterotic patterns in mating schemes for tomato breeding (Jin *et al.*, 2019).

By analysing a broad range of qualitative and quantitative traits, the present study demonstrates a considerable phenotypic diversity for a collection of tomato landraces originated from a semi-arid region in the Mediterranean basin. Results point at valuable germplasm that could be used as the starting material for the selective breeding of particular traits of interest, such as yield, firmness and alternative fruit shape/colour. Furthermore, these findings indicate that genotypes with high morpho-physiological diversity could be employed to develop a Multi-parent Advanced Generation Intercross tomato population to tap local phenotypic diversity, and create new cultivars well-adapted to low-input agricultural systems.

### Supplementary material

The supplementary material for this article can be found at <https://doi.org/10.1017/S1479262120000088>.

## Acknowledgements

The authors wish to thank Dr Andreas Ntoulis (HAO DEMETER) for providing the seed collection, as well as the field technicians of IPBGR, Nikolaos Ligdas and Theodoros Gagalnas, for their assistance in field preparation and help with cultivation practices. Special thanks to Dimitrios Drakos and Margarita Trikoukaki for their technical assistance. The authors received no specific funding for this work.

## References

- Avdikos ID, Tselivika N, Gallidou A, Koutsika-Sotiriou M and Traka-Mavrona E (2011) Exploitation of heterosis through recurrent selection scheme applied in segregating generations of a tomato breeding program. *Scientia Horticulturae* 130: 701–707.
- Bai Y, Kissoudis C, Yan Z, Visser RGF and van der Linden G (2018) Plant behaviour under combined stress: tomato responses to combined salinity and pathogen stress. *The Plant Journal* 93: 781–793.
- Bechere E, Belay G, Mitiku D and Merker A (1996) Phenotypic diversity of tetraploid wheat landraces from northern and north-central regions of Ethiopia. *Hereditas* 124: 165–172.
- Campanelli G, Sestili S, Acciarri N, Montemurro F, Palma D, Leteo F and Beretta M (2019) Multi-parental advances generation inter-cross population, to develop organic tomato genotypes by participatory plant breeding. *Agronomy* 9: 119–132.
- Casañas F, Simó J, Casals J and Prohens J (2017) Toward an evolved concept of landrace. *Frontiers in Plant Science* 08: 1–7.
- Cebolla-Cornejo J, Roselló S and Nuez F (2013) Phenotypic and genetic diversity of Spanish tomato landraces. *Scientia Horticulturae* 162: 150–164.
- Cortés-Olmos C, Valcárcel JV, Roselló J, Díez, MJ, and Cebolla-Cornejo J (2015) Traditional Eastern Spanish varieties of tomato. *Scientia Agricola* 72: 420–431.
- Csambalik L, Gál I, Sipos L, Gere A, Koren D, Bíró B and Divéky-Értsey A (2019) Evaluation of processing type tomato plant genetic resources (*Solanum lycopersicum* L.) for their nutritional properties in different environments. *Plant Genetic Resources: Characterization and Utilization* 17: 488–498.
- de Oliveira EJ, Dias NLP and Dantas JLL (2012) Selection of morpho-agronomic descriptors for characterization of papaya cultivars. *Euphytica* 185: 253–265.
- Elkind Y, Galper OB-O, Scott JW and Kedar N (1990) Genotype by environment interaction of tomato blossom-end scar size. *Euphytica* 50: 91–95.
- Eticha F, Bekele E, Belay G and Börner A (2005) Phenotypic diversity in tetraploid wheats collected from Bale and Wello regions of Ethiopia. *Plant Genetic Resources* 3: 35–43.
- Figàs MR, Prohens J, Raigón MD, Fernández-de-Córdova P, Fita A and Soler S (2015) Characterization of a collection of local varieties of tomato (*Solanum lycopersicum* L.) using conventional descriptors and the high-throughput phenomics tool Tomato Analyzer. *Genet Resour Crop Evol* 62: 189–204.
- Figàs MR, Prohens J, Casanova C, Fernández-de-Córdova P and Soler S (2018) Variation of morphological descriptors for the evaluation of tomato germplasm and their stability across

- different growing conditions. *Scientia Horticulturae* 238: 107–115.
- Foolad MR and Panthee DR (2012) Marker-assisted selection in tomato breeding. *Critical Reviews in Plant Sciences* 31: 93–123.
- Ganopoulos I, Xanthopoulou A, Molassiotis A, Karagiannis E, Moysiadis T, Katsaris P, Aravanopoulos F, Tsafaris A, Kalivas A and Madesis P (2015) Mediterranean Basin *Ficus carica* L.: from genetic diversity and structure to authentication of a protected designation of origin cultivar using microsatellite markers. *Trees – Structure and Function* 29: 1959–1971.
- Gonias ED, Ganopoulos I, Mellidou I, Bibi AC, Kalivas A, Mylona PV, Osanthanunkul M, Tsafaris A, Madesis P and Foulis AG (2019) Exploring genetic diversity of tomato (*Solanum lycopersicum* L.) germplasm of genebank collection employing SSR and SCAR markers. *Genetic Resources and Crop Evolution* 66: 1295–1309.
- Gramazio P, Chatziefstratiou E, Petropoulos C, Chioti V, Mylona P, Kapotis G, Vilanova S, Prohens J and Papatotopoulos V (2019) Multi-level characterization of eggplant accessions from Greek Islands and the mainland contributes to the enhancement and conservation of this germplasm and reveals a large diversity and signatures of differentiation between both origins. *Agronomy* 9: 887.
- Gur A and Zamir D (2004) Unused natural variation can lift yield barriers in plant breeding. *PLoS Biology* 2: e245.
- IPGRI (1996) *Descriptors for Tomato (Lycopersicon Spp.)*. Rome: International Plant Genetic Resources Institute.
- Jin L, Zhao L, Wang Y, Zhou R, Song L, Xu L, Cui X, Li R, Yu W and Zhao T (2019) Genetic diversity of 324 cultivated tomato germplasm resources using agronomic traits and InDel markers. *Euphytica* 215: 1–16.
- Kaiser HF (1958) The varimax criterion for analytic rotation in factor analysis. *Psychometrika* 23: 187–200.
- Koutsika-Sotiriou M, Mylonas I, Tsivelikas A and Traka-Mavrona E (2016) Compensation studies on the tomato landrace “Tomataki Santorinis.” *Scientia Horticulturae* 198: 78–85.
- Lázaro A (2018) Tomato landraces: an analysis of diversity and preferences. *Plant Genetic Resources: Characterization and Utilization* 16: 315–324.
- Mazzucato A, Papa R, Bitocchi E, Mosconi P, Nanni L, Negri V, Picarella ME, Siligato F, Soressi GP, Tiranti B and Veronesi F (2008) Genetic diversity, structure and marker-trait associations in a collection of Italian tomato (*Solanum lycopersicum* L.) landraces. *Theoretical and Applied Genetics* 116: 657–669.
- Mazzucato A, Ficcadenti N, Caioni M, Mosconi P, Piccinini E, Sanampudi VRR, Sestili S and Ferrari V (2010) Genetic diversity and distinctiveness in tomato (*Solanum lycopersicum* L.) landraces: the Italian case study of “A pera Abruzzese.” *Scientia Horticulturae* 125: 55–62.
- Mehmood A, Jaskani MJ, Khan IA, Ahmad S, Ahmad R, Luo S and Ahmad NM (2014) Genetic diversity of Pakistani guava (*Psidium guajava* L.) germplasm and its implications for conservation and breeding. *Scientia Horticulturae* 172: 221–232.
- Mellidou I and Kanellis AK (2017) Genetic control of ascorbic acid biosynthesis and recycling in horticultural crops. *Frontiers in Chemistry* 5: 50.
- Mengistu DK, Kiros AY and Pè ME (2015) Phenotypic diversity in Ethiopian durum wheat (*Triticum turgidum* var. durum) landraces. *Crop Journal* 3: 190–199.
- Ortiz Ríos R (2015) Plant genetic resources for food and agriculture. *Plant Breeding in the Omics Era*. Cham: Springer International Publishing, pp. 19–39.
- Padilla-Ramirez JS, Gonzalez-Gaona E and Ambriz-Aguilar J (2012) International market of fresh and processed guava: challenges and perspectives for the Mexican case. *Acta Horticulturae*. 959, 15–22.
- Parisi M, Aversano R, Graziani G, Ruggieri V, Sigillo L, Senape V and Barone A (2016) Phenotypic and molecular diversity in a collection of ‘Pomodoro di Sorrento’ Italian tomato landrace. *Scientia Horticulturae* 203: 143–151.
- Paupière MJ, van Haperen P, Rieu I, Visser RGF, Tikunov YM and Bovy AG (2017) Screening for pollen tolerance to high temperatures in tomato. *Euphytica* 213: 130.
- Rao R, Corrado G, Bianchi M and Di Mauro A (2006) (GATA)4 DNA fingerprinting identifies morphologically characterized “San Marzano” tomato plants. *Plant Breeding* 125: 173–176.
- Sacco A, Ruggieri V, Parisi M, Festa G, Rigano MM, Picarella ME, Mazzucato A and Barone A (2015) Exploring a tomato landraces collection for fruit-related traits by the aid of a high-throughput genomic platform. *PLoS ONE* 10: 1–20.
- Steven MA and Rick CM (1986) Genetics and breeding. In: Atherton JG and Rudich J (eds) *The Tomato Crop*. London, Great Britain: Chapman and Hall Ltd, pp. 35–109.
- Terzopoulos PJ and Bebeli PJ (2010) Phenotypic diversity in Greek tomato (*Solanum lycopersicum* L.) landraces. *Scientia Horticulturae* 126: 138–144.
- Terzopoulos PJ, Walters SA and Bebeli PJ (2009) Evaluation of Greek tomato landrace populations for heterogeneity of horticultural traits. *European Journal of Horticultural Science* 74: 24–29.
- UPOV (2017) Test Guidelines. Available at [http://www.upov.int/en/publications/tg-rom/tg\\_indexhtm](http://www.upov.int/en/publications/tg-rom/tg_indexhtm).