Theor Appl Genet (2008) 116:657–669
DOI 10.1007/s00122-007-0699-6

ORIGINAL PAPER

Genetic diversity, structure and marker-trait associations in a collection of Italian tomato (*Solanum lycopersicum* L.) landraces

Andrea Mazzucato · Roberto Papa · Elena Bitocchi · Pietro Mosconi · Laura Nanni · Valeria Negri · Maurizio Enea Picarella · Francesca Siligato · Gian Piero Soressi · Barbara Tiranti · Fabio Veronesi

Received: 15 January 2007 / Accepted: 6 December 2007 / Published online: 9 January 2008

Abstract The study of phenotypic and genetic diversity in landrace collections is important for germplasm conservation. In addition, the characterisation of very diversified materials with molecular markers offers a unique opportunity to define significant marker-trait associations of biological and agronomic interest. Here, 50 tomato landraces (mainly collected in central Italy), nine vintage and modern cultivars, and two wild outgroups were grown at two locations in central Italy and characterised for 15 morpho-physiological traits and 29 simple sequence repeat (SSR) loci. The markers were selected to include a group of loci in regions harbouring reported quantitative trait loci (QTLs) that affect fruit size and/or shape (Q-SSRs) and a group of markers that have not been mapped or shown to have a priori known linkage (NQ-SSRs). As revealed by univariate and multivariate analyses of morphological data, the landraces grouped according to vegetative and reproductive traits, with emphasis on fruit size, shape and final destination of the product. Compared to the low molecular polymorphism reported in tomato modern cultivars, our data reveal a high level of molecular diversity in landraces. Such diversity has allowed the inference of the existence of a genetic structure that was factored into the association analysis. As the proportion of significant associations is higher between the Q-SSR subset of markers and the subset of traits related to fruit size and shape than for all of the other combinations, we conclude that this approach is valid for establishing true-positive marker-trait relationships in tomato.

Communicated by I. L. Goldman.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-007-0699-6) contains supplementary material, which is available to authorized users.

A. Mazzucato (✉) · P. Mosconi · M. E. Picarella · F. Siligato · G. P. Soressi
Dipartimento di Agrobiologia e Agrochimica,
Università degli Studi della Tuscia, 01100 Viterbo, Italy
e-mail: mazz@unitus.it

R. Papa · E. Bitocchi · L. Nanni
Dipartimento di Scienze degli Alimenti,
Università Politecnica delle Marche, 60100 Ancona, Italy

V. Negri · B. Tiranti · F. Veronesi
Dipartimento di Biologia Vegetale e Biotecnologie
Agro-ambientali e Zootecniche, Università degli Studi di Perugia,
06100 Perugia, Italy

Introduction

Understanding the molecular genetic control of phenotypic variation is a major task in the study of natural and cultivated plant populations. This challenging objective becomes much more complicated when traits controlled by quantitative loci are considered, as for most of the features underlying crop yield and quality. Along with the classical analysis of quantitative trait loci (QTLs) through the genotyping of suitable segregating populations, new approaches have recently been proposed, where the identification of QTLs for important traits is addressed through association genetics after the phenotypic and genotypic characterization of collections of diverse materials (Lynch and Walsh 1997; Flint-Garcia et al. 2005). Since its first application on inbred maize lines (Thornsberry et al. 2001), association mapping has proven to be a reliable tool to highlight marker-trait associations in a number of plant species (Kraakman et al. 2001; Zhang et al. 2005; Aranzana et al. 2005; Breseghello and Sorrels 2006; Herrmann et al. 2006).

Among the vegetables, the tomato (*Solanum lycopersicum* L.) has been the species of choice for the study of the

genetic control of quantitative variation, especially for those traits determining fruit size, shape and quality (Paterson et al. 1991; Eshed and Zamir 1995; Grandillo et al. 1996, 1999; Frary et al. 2000; Van der Knaap and Tanksley 2001, 2003; Fulton et al. 2002; Van der Knaap et al. 2002; Barrero and Tanksley 2004; Causse et al. 2004). The importance of tomato as a crop, as well as the reason for its wide adoption in QTL studies can be seen from its history, which shows that the large variability that exists is due to a limited number of "domestication" genes, that mainly control fruit dimensional, structural, aesthetic and quality attributes (Grandillo et al. 1999; Tanksley 2004).

First introduced into Europe from Central and Southern America at the beginning of the sixteenth century, the tomato was initially considered a botanical curiosity, and its potential as a foodstuff was hindered by the suspicion of the presence of alkaloids in the fruit. It was only in the seventeenth century that the species began to be appreciated as an edible product and that its cultivation rapidly diffused through the Old World. This introduction represented a bottleneck, narrowing the genetic diversity of the cultivated tomato germplasm (Rick 1976).

In Europe, the tomato found success mainly in the Mediterranean countries, including Spain and Italy (Soressi 1969; Esquinas-Alcazar and Nuez 1995; García-Martínez et al. 2006), from where the species was reintroduced into North America in the eighteenth century (Rick 1976). Due to its success in cultivation, in North America, Italy and Spain, *S. lycopersicum* found a secondary centre for diversification (Bailey et al. 1960; García-Martínez et al. 2006). However, in North America, the standard tomato soon became the nearly globular, solid and smooth fruit; the other forms of tomatoes were grown for curiosity as "heirlooms" or "garden" varieties (Bailey et al. 1960; Noble 1994). In contrast, in Italy and in other European countries, the flat angled and ribbed tomatoes, and the pear-shaped, heart-shaped, extremely elongated, and cherry and plum forms were appreciated and cultivated. All these types finally gave rise to landraces that have been adopted for centuries and are still common in the local markets (Soressi 1969; Ruiz et al. 2005).

The initial narrow genetic basis of the tomato was further restricted by the development of vintage and modern cultivars, when much of the diversity within the cultivated *S. lycopersicum* was lost (Rick 1976; Miller and Tanksley 1990; Williams and St. Clair 1993). For this reason, since molecular analyses have become available for the estimation of genetic diversity, the level of intraspecific polymorphism reported in the cultivated tomato gene pool has generally been very low compared to that revealed in other self-compatible, autogamous species (Miller and Tanksley 1990; Williams and St. Clair 1993 and refs therein; Broun and Tanksley 1996; Noli et al. 1999; Archak et al. 2002; García-Gusano et al. 2004). Relatively to other classes of molecular markers (Miller and Tanksley 1990; Williams and St. Clair 1993; Noli et al. 1999; Archak et al. 2002; Park et al. 2004; Frary et al. 2005; Tam et al. 2005), simple sequence repeats (SSRs) have shown higher level of intraspecific polymorphism in cultivated *S. lycopersicum* (Bredemeijer et al. 2002; Suliman-Pollatschek et al. 2002; He et al. 2003; Tam et al. 2005; García-Martínez et al. 2006). In addition, because these markers also offered the advantage of genetic co-dominance, high reproducibility, easy detection, and multiallelic variation, SSRs were widely developed for genetic studies in tomato (Smulders et al. 1997; Areshchenkova 2000; He et al. 2003; Frary et al. 2005; Tam et al. 2005; http://www.sgn.cornell.edu/).

As association analysis has not been implemented in tomato to date, the present study aimed to address this procedure by using markers and plant materials that provide the best probability of high intraspecific variation. Therefore, through morphological descriptors and SSR markers we characterised a collection of Italian tomato landraces and suitable control genotypes. Both morphological and molecular data were used to infer the existence of a genetic structure in the collection studied. To gain a direct validation of the reliability of the associations revealed, a portion of the markers was selected as located in proximity to known QTLs that control tomato fruit size and shape.

Materials and methods

Plant material

Morphological and molecular genetic diversity was assessed in 61 accessions of cultivated tomato (*Solanum lycopersicum* L.) or wild-related species. The collection included 40 landraces (36 from central and southern Italy, and four from Latin America), ten accessions of commercially available landraces, nine vintage or modern cultivars, and one accession each of *S. lycopersicum* var. *cerasiforme* and of *S. pimpinellifolium* as wild outgroups. Details on accession category, name, fruit-shape typology, acronym and origin are given in Supplementary Table S1. Seed stocks of the cultivars, the wild species and the four accessions from Latin America were obtained from either seed markets or from the C.M. Rick Tomato Genetics Resource Center (TGRC), University of California, Davis, USA (Supplementary Table S1). The Italian landraces were either gathered from the farmers or from the local seed markets. To briefly explain the acronyms used, the first two letters refer to the category (IL, Italian landrace; AL, Latin America landrace; CV, cultivar; WS, wild species). For ALs, CVs, and WSs, the category is followed by the five letters of the accession name. For ILs, the first letter after

the category indicates the geographical region of provenance (A, Abruzzo; L, Lazio; M, Marche; S, South Italy; T, Tuscany; U, Umbria) or the commercial origin (C). The two letters that follow indicate the fruit typology declared at the time of collection (FR, flattened-ribbed; PE, pear-shaped; HE, heart-shaped; EL, elongated; SM, small round or oval; OT, others). Finally, the last two numbers refer to the local code.

The original accessions were regrown to evaluate the presence of morphologically heterogeneous stocks. Eight out of 59 accessions of cultivated tomato give rise to heterogeneous progeny. In uniform accessions, a single healthy plant was selected and used for seed production and DNA extraction; in the heterogeneous ones, the material was obtained from a single plant selected as belonging to the modal class. A field trial was established with the above-described seed stocks at two locations in central Italy: Viterbo (42° 26′N, 12° 04′E) and Perugia (43° 05′N, 12° 05′E), in the same growing season (2004). The accessions were arranged in a randomised block design with two replicates and four plants per elementary experimental unit. Plants were grown in open fields with standard agronomic practices.

Phenotypic data

On a single plant basis, 22 morpho-physiological traits were scored or calculated as detailed in Supplementary Table S2. Briefly, the traits recorded were growth habit (GH, score), plant height (PH, cm), flowering date (FD, days), inflorescence type (IT, score), number of flowers per inflorescence (NF), stigma exertion (SE, score), green shoulder (GS, score), ripening date (RD, days), fruit colour (FC, score), number of harvested fruits (NHF), weight of harvested fruits (WHF, g), fruit polar (PD, mm) and equatorial (ED, mm) diameters, stem-end shape (SES, score), blossom-end shape (BES, score), number of fruit locules (LN), pericarp thickness (PT, mm), and puffiness (PUF, score). The calculated variables were fruit-shape index [FS, (PD/PE)], mean fruit weight [FW, (WHF/NHF), g], flowering-ripening interval [RT, (RD-FD), days] and pericarp-thickness index [PI, (PT/((PD + PE)/2))]. These descriptors largely conform to the guidelines of the International Plant Genetic Resources Institute (IPGRI; http://www.ipgri.cgiar.org/publications/pdf/286.pdf). Those variables used to calculate other variables (i.e. FD, RD, NHF, WHF, PD, ED, and PT) were excluded from the further analyses. Among the genotypes, the two wild species were also excluded from the analysis of morphological data. For the remaining 15 variables and 59 genotypes, differences and interactions were estimated through an analysis of variance adopting the General Linear Model (GLM) with the adjustments for experiments combined over more locations (McIntosh 1983). The Spearman rank correlation coefficients between variables were calculated on the basis of the mean values of the two environments. All univariate analyses were carried out with the Statistical Analysis System software (SAS Institute Inc. 2002). Correlation coefficients were displayed graphically after multidimensional scaling performed with the PERMAP 11.3 software (http://www.ucs.louisiana.edu/~rbh8900/permap.html).

Although all of the variables, with the exception of NF and PI, showed significant genotype × environment interactions, the examination of the data revealed that this significance was due to the behaviour of a few genotypes for each trait. Therefore, allowance was made for these interactions and, after standardisation, the arithmetic means over locations were used to perform multivariate analyses in order to delineate clusters of morphologically similar individuals. Principal component analysis (PCA) and hierarchical cluster analysis (HCA) were carried out with the SYNTAX Package 5.0 (Podani 1993). Clustering was based on the chord distance coefficient (Orloci 1972) and the average linkage method was used as fusion criterion.

Genotypic data

The DNA extracted from each genotype selected for seed harvest was analysed with 29 SSR markers that were selected from the literature to obtain an approximate coverage of the tomato genome (at least two markers per chromosome). Briefly, the loci scored were LE21085, LEMDDNa, LELEUZIP, LELE25, LE20592, LEACS4A, LESSRPSPGb (Smulders et al. 1997), TMS63, TMS58, TMS54, EST258529, EST253712, TMS60, TMS59, TMS42, TMS52, EST245053 (Areshchenkova and Ganal 2002), Tom178-179, Tom59-60, Tom300-301, Tom67-68, Tom8-9, Tom41-42, Tom198-199, Tom162-163, Tom11-28, Tom47-48, Tom292-293, and Tom236-237 (Suliman-Pollatschek et al. 2002). Details on the marker dataset are given in Table 1 and in Supplementary Table S3. The SSRs were selected to include a group of markers located in regions harbouring reported QTLs that affect fruit size and/or shape (Q-SSRs, according to the criteria detailed below) and a group of markers either not mapped or having no a priori known linkage with genes affecting fruit traits (NQ-SSRs). Out of the 29 analysed loci, 16 were Q-SSRs and 13 NQ-SSRs. To establish linkage relationships, marker (Areshchenkova and Ganal 2002; Suliman-Pollatschek et al. 2002) and QTL (Paterson et al. 1991; Grandillo et al. 1999; Van der Knaap and Tanksley 2001, 2003; Van der Knaap et al. 2002) positions were obtained from the literature. Because the marker and/or QTL positions were reported frequently as confidence intervals flanked by markers, the position was approximated to the mid point of such bin. Q-SSRs were thus defined as markers having an

Table 1 Characteristics of the polymorphic SSR markers used

Locus name	First described[a]	Chromosome	Linkage to reported QTLs controlling fruit weight and/or shape[a]	No. of alleles[a,b] (literature)	No. of alleles (present study)		Allele size range (bp)[c]
					All genotypes[c]	*Solanum lycopersicum*[d]	
TMS63	(7)	1	*fw1.1*(3), *sblk1.1, hrt1.1*, *nsf1.1*, *nfl1.1* (12)	3 (7)	3	3	130–150
TMS58	(7)	3	*fw3.1* (3), *nsf3.1*, *lcn3.1* (12)	2 (7)	4	3	223–226
Tom59-60	(9)	3	*fw3.1* (3) *ljfs3* (10), *sblk3.1, hrt3.1, fw3.2, nsf3.1*, *lcn3.1*, *nfl3.1* (12)	nd[e]	7	7	113–122
EST258529	(7)	5	*nfl5.1* (12)	2(7)	3	2	121–129
EST253712	(7)	6	*MF6a* (1)	4(7)	5	5	130–148
TMS60	(7)	7	*ljfs7* (10)	2(7)	5	5	232–246
TMS59	(7)	8	*fs8.1* (3), *bell8.1*, *bpi8.1* (12)	2(7)	3	2	100–106
Tom236-237	(9)	9	*nsf9.1*, *nfl9.3* (12)	3(9)	10	8	167–189
TMS42	(4,7)	11	*ljfs11* (10)	4(4,11)	5	5	272–283
TMS52	(7)	12	*sblk12.1*, *nsf12.1, lcn12.1* (12)	3(7)	9	9	152–174
EST245053	(7)	1		2 (7)	2	2	228–230
Tom162-163	(9)	1		nd[e]	2	2	148–156
Tom47-48	(9)	3		nd[e]	9	8	169–212
LE21085	(2)	4		3–12(2,5,8,11,13)	3	3	121–177
LEMDDNa	(2)	5		4–7(2,8,11,13)	3	3	211–227
LELEUZIP	(2)	8		1–5(2,5,8,11,13)	5	4	101–107
LELE25	(2)	10		3–4(2,8,11,13)	3	3	225–229
LE20592	(2)	11		4–7(2,8,13)	6	6	181–195
LEACS4A	(2)	nd		2(11)	8	8	144–286
LESSRPSPGb	(2)	nd		3–5(2,13)	8	8	295–352

[a] (1) Paterson et al. 1991, (2) Smulders et al. 1997, (3) Grandillo et al. 1999 and refs therein, (4) Areshchenkova 2000, (5) Alvarez et al. 2001, (6) Van der Knaap and Tanksley 2001, (7) Areshchenkova and Ganal 2002, (8) Bredemeijer et al. 2002, (9) Suliman-Pollatschek et al. 2002, (10) Van der Knaap et al. 2002, (11) He et al. 2003, (12) Van der Knaap and Tanksley 2003, (13) García-Martínez et al. 2006. *bell*, bell shape; *bpi*, bumpiness; *fs*, fruit shape; *fw*, fruit weight; *hrt*, heart shape; *lcn*, locule number; *ljfs*, extremely elongated fruit shape; *MF*, mass per fruit; *nfl*, flowers per inflorescence; *nsf*, seed number per fruit; *sblk*, stem-end blockiness

[b] Refers to collections with variable numbers and types of accessions

[c] Refers to the total of 61 genotypes

[d] Refers to 59 *S. lycopersicum* genotypes

[e] Not determined

estimated distance from QTLs affecting fruit size and/or shape <15 cM, according to the available maps (http://www.sgn.cornell.edu/cview/). For the QTLs reviewed by Grandillo et al. (1999), only those showing a percent of phenotypic variance explained by the locus greater than 20% in at least one study were taken into account. All of the QTLs reported by Van der Knaap and Tanksley (2003) were included as very frequently the QTLs that control traits not directly describing fruit size or shape were co-localized with QTLs responsible for fruit dimension.

DNA was obtained from young leaves by the mini-extraction method of Doyle and Doyle (1987). PCR amplification was performed in a 20 µl total volume, containing 20 ng of genomic DNA template, 50 pmol of each of the two primers, 200 µM dNTPs, 2 mM $MgCl_2$, 1× Taq polymerase buffer, and 1 U Taq DNA polymerase (Promega, Madison, WI, USA). One of the two SSR primers was end-labelled with Cy5. The amplifications were conducted with a Perkin-Elmer 9700 Thermal Cycler (Applied Biosystems, Foster City, CA, USA), with an initial 5 min at 94°C that was followed by 30 cycles of 30 s at 94°C, 30 s at X°C, and 30 s at 72°C, plus 15 min final extension at 72°C. X°C refers to the annealing temperature, which is specified for each primer in the relevant reference (see Table 1 and Supplementary Table S3). The amplification products were separated on 6% (w/v) denaturing polyacrilamide (1:19 bis:acrilamide) gels and visualised with the GenomyxSC system (Beckman, Palo Alto, CA, USA).

Each individual was genotyped at each locus by scoring the length of the amplified SSR band with reference to molecular ladders and genotypes of known allelic composition. Nine loci were monomorphic among the *S. lycopersicum*

genotypes (Supplementary Table S3) and were excluded from the subsequent analyses. For polymorphic loci, the mean number of observed (No) and expected (Ne) alleles per locus, the observed (Ho) and expected (He) unbiased heterozygosity, and the percentage of polymorphic loci were calculated, considering both the whole collection and the subsets of genotypes based on germplasm category or fruit type. The analysis was carried out with POWERMARKER Ver. 3.23 (Liu and Muse 2005; http://www.powermarker.net). Genetic distance (Nei et al. 1983) matrices and cluster analyses based on the unweighted pair group method (UPGM) of Sneath and Sokal (1973) were calculated with the TREECON Ver. 1.3b software (Van de Peer and De Wachter 1993) using the 20 informative loci together or the subsets of the Q-SSRs and NQ-SSRs. The morphology-based distance matrix was compared with those obtained from the molecular data using the 20 informative loci together or the subsets of Q-SSRs and NQ-SSRs by POWERMARKER, using Mantel's test.

Linkage disequilibrium (LD) analysis was performed for the 20 polymorphic SSRs, with the dedicated procedure of the TASSEL software, using 1,000 permutations (http://www.maizegenetics.net/).

To delineate clusters of individuals on the basis of their genotypes at multiple loci the model-based STRUCTURE Ver. 2 software (Pritchard et al. 2000) was used, adopting the "admixture model", a burn-in period of 30,000 iterations, and the data from 20 polymorphic SSR loci. The number of clusters (k) was set to five, as this number maximized the Δk parameter (Evanno et al. 2005). This model-based approach assumes unlinked loci and HW equilibrium. In our case, the first assumption is likely to be correct (even if some loci were of unknown map position and thus potentially linked to other SSR loci), but the HW would be violated. However, because we used this approach only for classification purposes, both aspects were not considered as providing bias towards our inferences.

Two analytical approaches were used to identify and validate putative marker-trait associations. First, the non-parametric Kruskal–Wallis K-test was chosen to take into consideration the nature of most of the data variables that significantly departed from the assumption of normality. This method can be regarded as the non-parametric equivalent of the one-way analysis of variance (Lehmann and D'Abrera 1975). The second approach employed was the logistic regression procedure, a parametric test that factors into the analysis the existing population structure (Flint-Garcia et al. 2005). For the non-parametric analysis, morphological data were categorized into five classes, as described previously (Mazzucato 1995), using the mean and σ values (1, lower values; 5, higher values). Associations were then detected according to the K-test (NPAR1WAY SAS procedure, option WILCOXON, SAS Institute Inc. 2002). The parametric test was carried out using the logistic regression procedure of the TASSEL software, which estimates associations between segregating sites and phenotypes, while accounting for population structure. The genotypic data were compared with the original mean values for each morphological trait, while factoring into the analysis the five-cluster structure obtained as described above. Admixed genotypes (i.e. genotypes with membership probabilities <0.8) were assigned to a mixed sixth group, as in Flint-Garcia et al. (2005). Logistic regression was carried out with 500 permutations. For both methods, an association was indicated when the mean values of the marker classes were significantly different at $P < 0.01$.

The frequency of significant marker-trait relationships was calculated separately for Q-SSRs and NQ-SSRs, considering the overall morphological traits and the traits divided between those directly describing fruit size and shape (LN, BES, SES, FW, PI, FS, plus IT, because this trait was highly correlated with them; F-traits) and those unrelated to fruit dimension (GH, PH, NF, SE, GS, FC, PUF, RT; NF-traits). The χ^2 homogeneity analysis was used to compare the proportions obtained for each combination, considering the two classes of markers. The expected values were determined on the basis of the totals of lines and columns in a 2×2 table.

Results

Morpho-physiological analysis

The analysis of variance revealed significant differences among genotypes for all of the traits, and significant genotype × environment interactions for all of the traits, with the exception of NF and PI (data not shown). The Spearman rank correlation coefficients calculated between pairs of variables (Supplementary Table S4) revealed how some were rather independent, whereas a group of traits clustered together because of a reciprocal tight correlation (Fig. 1, upper-right quadrant). Notably, all of these traits were F-traits, i.e. directly involved in the expression of fruit size and shape attributes. This reflected the tendency for large fruits (high FW) to have a curved stem-end shape (high SES), a flat blossom-end shape (low BES), a high LN, a flat or round fruit shape (low FS) and a thin pericarp (low PI; Fig. 2d (I–III)). These phenotypes were also significantly associated with compound inflorescence (high IT, Fig. 1). In contrast, the medium and small fruit-size genotypes were generally round or elongated, and since most of the improved material fell into this category, they had low plant heights, simple inflorescences and a thick pericarp (Fig. 2d (IV–VI)).

In the principal component analysis (PCA), the first six principal coordinates (PCs) explained a total of 82% of the

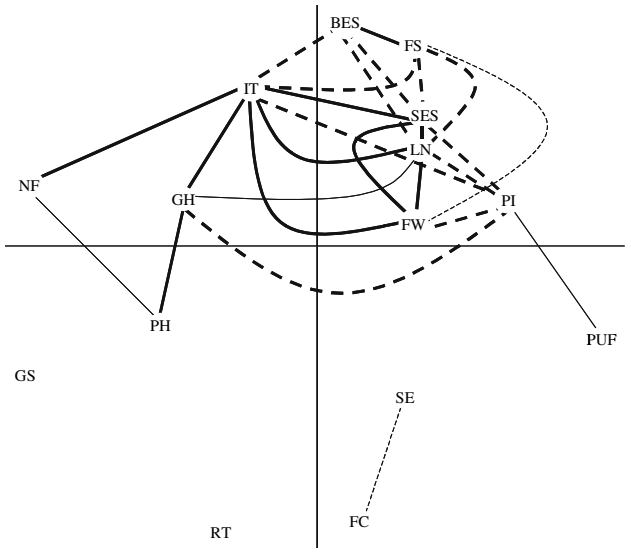

Fig. 1 Relationship map constructed by multidimensional scaling of the Spearman correlation coefficients calculated between pairs of 15 morphological traits. The distance between traits is inversely proportional to the sizes of the correlation coefficients. Correlations between traits significant for $P < 0.001$ and $P < 0.0001$ are represented with *thin* or *thick connecting lines*, respectively. *Solid* and *dashed lines* indicate positive and negative correlations, respectively. Trait acronyms are described in Supplementary Table S2 and in "Materials and methods"

morphological variation. The first and the second PC axes accounted for 34 and 15% of the variation and demonstrated all of the F-traits and part of the NF-traits (SE, FC, PUF, and RT), respectively. When the PCA was carried out with only the F-traits, the first six PCs explained 99% of the total variation (not shown).

The dissimilarity matrix obtained from the morphological descriptors was used for the clustering procedure. The resulting dendrogram displayed four major groups of genotypes (Fig. 2a). The first included all of the large-fruited tomatoes and could be further divided into three subclusters, representing the pear-shaped (Fig. 2a, pink stripe), the flattened-ribbed (yellow stripe), and the ox-heart (red stripe) types, respectively. Four genotypes remained outside of these three sub-groups. The second major cluster included mostly the small-round and elongated types, without a clear-cut distinction based on fruit morphology (green stripe). The two bell-pepper genotypes (IL-LBP03 and IL-MBP58) were positioned in a separate, third cluster (Fig. 2a). The last major cluster grouped the block-shaped genotypes (blue stripe), including the Rio Grande, M82 and Chico III cultivars. Fruits of accessions representative of each cluster are documented in Fig. 2d (I–VI).

Molecular analysis

Out of the 29 SSR loci assayed, four were monomorphic overall (two Q-SSRs and two NQ-SSRs) and five were monomorphic when the *S. lycopersicum* genotypes were considered (four Q-SSRs and one NQ-SSR), but showed polymorphism in the WS genotypes (Supplementary Table S3). These nine loci were excluded from subsequent analyses.

The percentage of the SSR loci pairs in linkage disequilibrium ($P < 0.001$) was about 4%, with the major part representing inter-chromosomal LD (data not shown).

Of the 20 informative loci, ten were Q-SSRs and ten NQ-SSRs (Table 1); the two groups yielded similar levels of polymorphism. Out of a total of 96 alleles detected in the *S. lycopersicum* genotypes overall, 49 were due to Q-SSRs (mean of 4.9 alleles per locus) and 47 to NQ-SSRs (4.7 alleles per locus). In *S. lycopersicum* genotypes, the No ranged from two to nine, with a mean of 4.8 (Table 1). When the No was calculated on subsets of genotypes based on the germplasm categories and normalized for the number of genotypes, the WSs showed the highest mean values (data not shown). Similarly, the WSs had the highest Ho values (Table 2). When the genotypes were grouped on the basis of the fruit shape, as indicated by the HCA analysis based on morphology, it was confirmed that the group of round-elongated genotypes (Fig. 2a, green stripe) was the most heterogeneous (Table 2). The dissimilarity matrices based on the full range of morphological and molecular descriptors showed significance, but were not highly correlated ($P < 0.05$, $r = 0.12$).

To take into account the structure of the molecular diversity that exists among the genotypes studied, a model-based analysis was performed using the STRUCTURE software and the data of 20 polymorphic SSR loci. After setting to five the number of clusters according to the statistic of Evanno et al. (2005), the software output the coefficients of estimated ancestry per each individual in each group. In the plot of ancestry estimates shown in Fig. 2c in parallel with the dendrogram based on the morphological data, each individual is represented by a single horizontal bar broken into five segments, with lengths proportional to the individual's estimated ancestry fraction from each of the five groups. Model-based groups were mostly consistent with the morphological classification. The groups that, based on the morphology, contained the accessions with pear-shaped, flattened-ribbed and heart-shaped fruits (Fig. 2a pink, yellow and red stripes) were paralleled, with few exceptions, by the clusters based on the SSRs (Fig. 2c pink, yellow and red bars). The fourth SSR-based cluster (green bars) loosely corresponded to the group of small-round and elongated tomatoes (green stripe); this cluster also contained the WSs that are not reported in the dendrogram (data not shown). Finally, the co-ancestry analysis formed a group (blue bars) that included the block-shaped cultivars (CV-CHICO, CV-RIOGR, partially CV-M82); therefore, this also paralleled the clustering based on the morphology. Some of the other different genotypes revealed co-ancestry to this group (Fig. 2c).

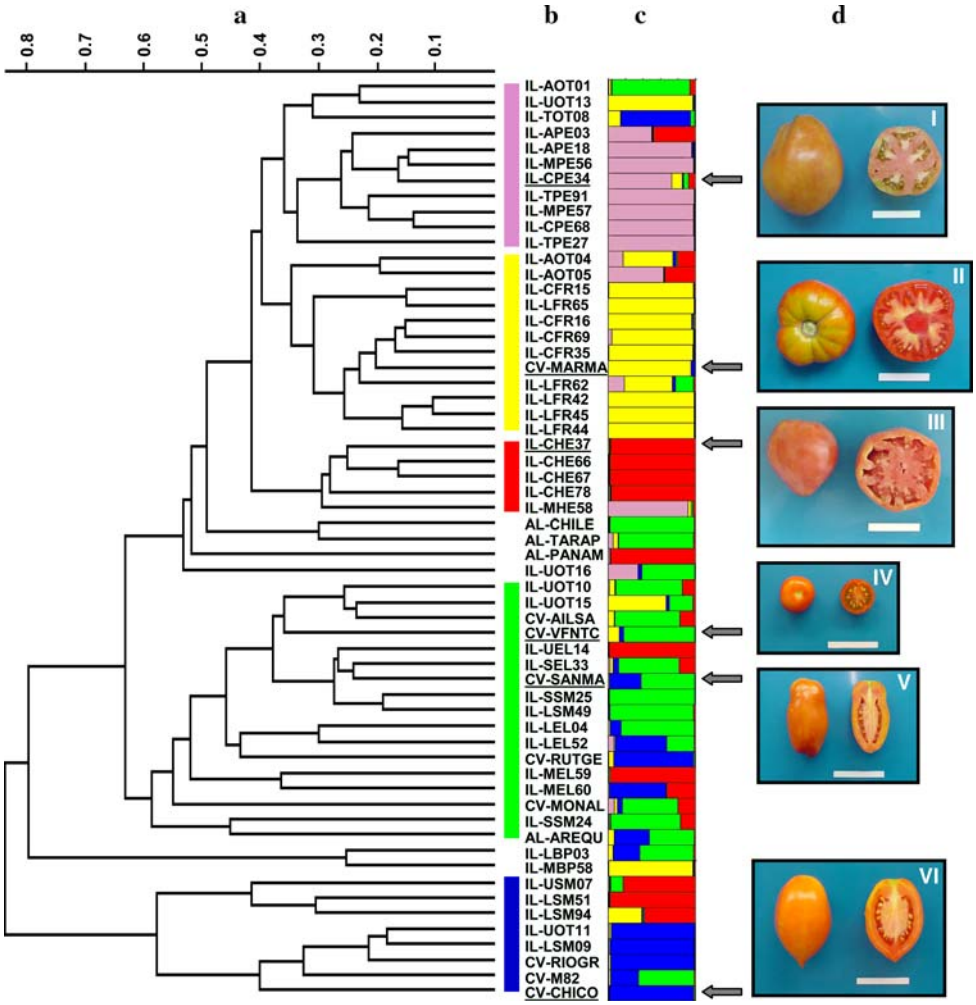

Fig. 2 Morphological and molecular analysis of 59 tomato accessions/genotypes mainly belonging to the Italian landrace germplasm. **a** Dendrogram of similarity based on 15 morpho-physiological descriptors (*colour stripes* highlight the major clusters). **b** Accession acronyms (for descriptions see "Materials and methods" and Supplementary Table S1). **c** Estimated population structure; each individual is represented by a *horizontal bar*, which is partitioned into five coloured segments that represent the individual estimated levels of the five clusters. **d** Images of fruits collected from accessions that are representative of the fruit type variations (*I*, *IL-CPE34*; *II*, *CV-MARMA*, *III*, *IL-CHE37*; *IV*, *CV-VFNTC*; *V*, *CV-SANMA*; *VI*, *CV-CHICO*). *Bar*, 5 cm

Table 2 Genetic diversity among the 61 genotypes as revealed by SSR analysis

Grouping criteria	Classes	No. of accessions	Mean observed alleles per locus (No)	Mean expected alleles per locus (Ne)	Observed heterozygosity (Ho)	Expected heterozygosity (He)	Polymorphic loci (%)
Total		61	5.2	2.28	0.05	0.44	100
Germplasm category	ILs	46	4.1	2.15	0.04	0.41	95
	CVs	9	2.6	1.88	0.04	0.36	85
	ALs	4	2.4	2.05	0.06	0.42	80
	WSs	2	1.8	1.70	0.15	0.33	65
Fruit shape typology[a]	Pear-shaped	11	2.5	1.72	0.03	0.31	70
	Flattened-ribbed	12	2.2	1.51	0.04	0.22	55
	Heart-shaped	5	1.9	1.46	0.03	0.26	70
	Round-elongated	17	3.3	2.13	0.05	0.42	95
	Block-shaped	8	2.5	2.01	0.03	0.34	60

CVs Cultivars, *ALs* Latin American landraces, *WSs* wild species, *ILs* Italian landraces

[a] WSs and six genotypes not falling into the major morphological clusters (Fig. 2a) were not included

Marker-trait associations

Upon running the *K*-test without taking into consideration the population structure, the 20 × 15 marker-trait comparison matrix revealed 29 significant associations (9.7%), about one order of magnitude larger than would be expected by chance (Table 3). Among these, 23 (15.3%) were due to Q-SSRs and six (4.0%) to NQ-SSRs (Fig. 3a). Because the

Table 3 Significant ($P < 0.01$) marker-trait associations detected by the Kruskal–Wallis non-parametric *K*-test and by logistic regression using marker loci linked to QTLs that contribute to fruit size and shape (Q-SSRs) or markers not mapped or having no a priori known linkage with genes affecting fruit traits (NQ-SSRs)

Marker type	Locus name	Significant associations detected by		
		Only *K*-test	Only logistic regression	Both analyses
Q-SSR	TMS63		IT	FS
	EST258529	FC		PI
	EST253712	PI	SES	FW, LN, IT
	TMS60	PH, BES, FW		IT
	TMS59	PI, FW, IT, SES		LN, FS, BES, GH
	Tom236-237		PUF, LN	GS
	TMS52	FW		PH, PI
NQ-SSR	Tom162-163		PI, PUF, RT	
	LEMDDNa	BES, GH		
	LELE25		NFI	LN, FW, SES
	LE20592		GH	
	LESSRPSPGb	PH		

For trait acronyms see "Materials and methods"

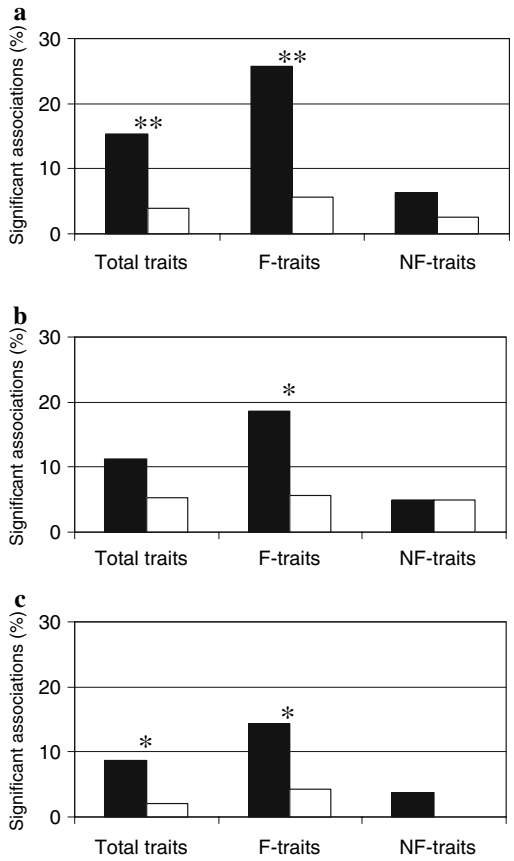

Fig. 3 Percentage of significant ($P < 0.01$) marker-trait associations as detected by Q-SSRs (*black bars*) and NQ-SSRs (*white bars*) considering all traits or separated between traits directly related to fruit size and shape (plus the highly correlated variable IT; *F-traits*) and other traits (*NF-traits*): **a** after Kruskal-Wallis non-parametric *K*-test; **b** after logistic regression, taking into account the population structure; and **c** those significant for both analyses. *, **, Proportions between Q-SSRs and NQ-SSRs significantly different for $P < 0.05$ and $P < 0.01$, respectively

Q-SSRs are qualified by their location in proximity to QTLs that control fruit size and/or shape, the 15 traits were also considered separately, dividing the F-traits from the NF-traits. After this distinction, the significant associations between Q-SSRs and F-traits were raised to 25.7%, whereas the other groups of combinations ranged between 2.5 and 6.3% (Fig. 3a). A similar trend was seen when the associations were calculated by logistic regression, taking into consideration the population structure (Fig. 3b). When only the associations detected by both analyses were considered, those between Q-SSRs and F-traits remained significantly higher than the other combinations (Fig. 3c; Table 3).

Some of the associations revealed by the non-parametric test were not recognized if the population structure among the genotypes was factored into the analysis. This was the case for clear false positives, as for the association between EST258529 and FC, which was essentially due to the ox-heart group sharing a specific, rare allele at the marker locus and the characteristic unpigmented fruit epidermis phenotype (*y*; Stevens and Rick 1986; FC = 4), and between LeMDDNa and GH, which was due to a rare allele almost specific to cultivars with a determinate growth habit (Table 3).

Among the relationships revealed by both analyses, the TMS63 marker showed significant association with FS (Table 3); almost all of the flattened-ribbed genotypes belonging to the Marmande group had a 130-bp allele, which was also shared by other genotypes with low or medium FS index (flattened ALs and some round genotypes; Fig. 4a). When the test was applied to the original diameter variables, it was clear that the significant association was essentially due to the polar diameter (not shown).

The EST253712 marker correlated with FW and with traits correlated to FW, such as LN and IT (Fig. 1). The 142-bp allele was associated with genotypes with two-loculed,

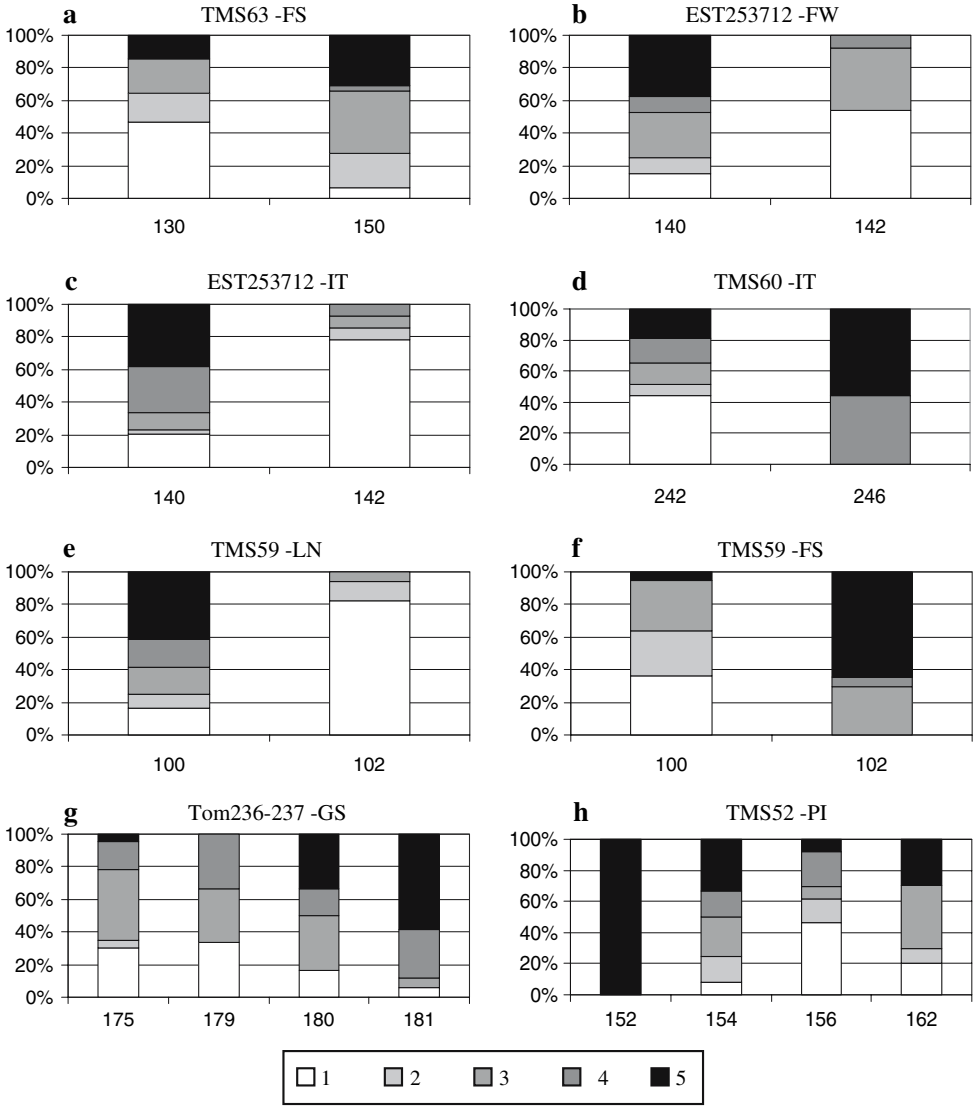

Fig. 4 Relative percentage histograms representing the significant marker-trait associations (defined in the title of each histogram). **a**–**h** *White* to *black bars* represent the distributions of genotypes according to their categorical phenotypic values (1–5), as referred to in the legend, for the classes of the most common alleles (*x axes*, in bp)

rather small fruits (low LN and FW; Fig. 4b) and simple inflorescence (low IT; Fig. 4c).

The TMS60 marker also showed significant association with IT (Table 3). After excluding the rare alleles (232, 240, and 244 bp), the significance was due to the 246-bp allele (frequency, 17%) that was shared by the genotypes with the highest IT values (Fig. 4d).

The allelic composition at the TMS59 locus was significantly associated with several morphological traits (Table 3). In the *S. lycopersicum* genotypes, the TMS59 marker showed two alleles, of 100 and 102 bp, with relative frequencies of 68 and 32%, respectively. The 102-bp allele was preferentially shared by genotypes with low locule number (low LN; Fig. 4e), round or elongated fruits (medium-high FS; Fig. 4f), and beaked bottom-end shape (high BES). These genotypes included all of the San Marzano and San Marzano-like accessions, the block-shaped cultivars and the IL-LBP03 bell-pepper accession. All of the other groups, plus some small-sized cherry types (CV-VFNTC, IL-LSM49, IL-SSM24), showed the 100-bp allele. The mean values of the phenotypic class for FS were 2.1 ± 0.2 for the 100-bp allele group and 4.4 ± 0.2 for the 102-bp allele group.

The alleles present at the Tom236-237 marker locus were significantly associated to GS. Genotypes with a strong intensity of the shoulder preferentially showed the 180- and 181-bp alleles (Fig. 4g); these included genotypes with very different fruit types, from flattened-ribbed (IL-LFR42 and IL-LFR45), to small-round (CV-AILSA, IL-LSM49, IL-SSM24, and IL-SSM25), and to elongated (IL-UEL14 and IL-LEL04).

While the association between TMS52 and PH did not reveal a pattern of particular interest, that with PI was due to the 152-bp allele being specific to genotypes with very high pericarp index (Fig. 4h).

The LELE25 marker was the only NQ-SSR to show marker-trait associations that were significant at both levels

of analysis (Table 3). This marker showed associations with LN, FW and SES due to two rather rare alleles (225 and 229 bp, with frequencies 3.4 and 8.5%, respectively) that were shared by genotypes with two-loculed, small-round fruits (CV-VFNTC and IL-LSM49 with the 225-bp allele, and CV-AILSA, CV-MONAL, IL-SSM25, IL-SSM24, and AL-AREQU with the 229-bp allele). This marker was significantly associated with both fruit diameters (not shown).

Discussion

Morphological and molecular diversity

The genetic diversity detectable in crops by molecular markers such as SSRs is generally dependent on the mating system, the domestication history, and the magnitude of the collection being analysed. In self-pollinating species, as the polymorphism is typically maintained between accessions, the estimated genetic diversity results in a function of the sample size. When wide collections are addressed (e.g. more than 100 entries), numbers of alleles per polymorphic locus above 10 have commonly been reported in studies of self-pollinating crops, such as wheat (Huang et al. 2002), rice (Xu et al. 2005), and soybean (Wang et al. 2006).

In tomato, since the first studies with molecular markers, it was clear that the level of genetic diversity in the cultivated gene pool was significantly lower than in other self-pollinating species (Williams and St. Clair 1993). Whereas still high numbers of alleles per polymorphic SSR locus (8.5) were reported when several wild accessions entered the collection (Alvarez et al. 2001), screenings limited to the cultivated germplasm yielded values near to 2.5 (He et al. 2003; Tam et al. 2005). Early studies also showed that regional, traditional cultivars from South America preserved more genetic diversity than modern tomato varieties (Williams and St. Clair 1993). Our data expand this result to the landraces differentiated in the Old World as we scored a number of alleles per polymorphic locus of 4.8 (only taking into account the 59 accessions of cultivated germplasm); this value was comparable to that reported in a collection of Spanish landraces of similar magnitude (García-Martínez et al. 2006). Because a screening of more than 500 European tomato cultivars with 20 SSR loci yielded a mean number of alleles of 4.7 (Bredemeijer et al. 2002), it can be concluded that a relatively small number of diverse landraces from relatively small geographical regions can encompass the same amount of genetic diversity that is shown by very large collections of varieties.

As it has been demonstrated that major phenotypic differences can often occur with only minor genotypic changes (Rick and Holle 1990; Williams and St. Clair 1993), a tight correlation between morphological and molecular profiling was not expected, and indeed the Mantel test between the two matrices was hardly below the threshold for statistical significance. As an example, two cultivars with rather different fruit shapes, such as San Marzano and Ailsa Craig, can nevertheless show a high level of co-ancestry with both AFLPs (Park et al. 2004) and SSRs (this study).

The model-based analysis of the genetic structure in the collection studied revealed several instances of co-ancestry that were not obvious using distance-based clustering methods, but were supported by historical and pedigree information. A high level of co-ancestry between *S. lycopersicum* var. *cerasiforme* and Latin American accessions is consistent with the hypothesis that much of the diversity from *cerasiforme* was incorporated into those regional cultivars (Rick 1976; Williams and St. Clair 1993). High co-ancestry to *S. lycopersicum* var. *cerasiforme* also characterised all of the cherry types (sometimes referred to as *cerasiforme* germplasm themselves), and the modern M82 cultivar and some of the old cultivars, like San Marzano and the derivative Corbarino landrace (IL-SEL33), which experienced introgressions from wild germplasm (Monti et al. 2004). Co-ancestry to this group was never found for the pear-shaped, heart-shaped or flattened-ribbed types.

Marker-trait associations

One of the reported constraints of association mapping studies is the easy detection of false positives resulting from the existence of genetic structure in the populations studied (Flint-Garcia et al. 2005). In addition to factoring the genetic structure into the analysis, the reliability of association studies has been addressed through case studies of previously known genes (Aranzana et al. 2005). Similarly, we carried out association studies in tomato by using a set of markers located in the same genomic region of known QTLs that contribute to fruit size and shape. The results demonstrate that this approach can be efficient in tomato because the proportion of significant associations was significantly higher when Q-SSRs were compared with F-traits than in all of the other combinations. Moreover, four out of five NQ-SSRs that gave significant marker-trait associations have actually not been precisely mapped yet, and thus they could be linked either to reported or unreported QTLs. It can be reasonably predicted that a suitable choice of markers, genotypes and traits can provide an analytical platform that is useful for the performance of genome-wide association mapping scans and to obtain insights into the function of specific QTLs over the phenotypic germplasm variation. To translate this information into tools for marker-assisted breeding would then be straightforward.

Our set of markers addressed six out of the nine key loci described as controlling size and shape variation in tomato (Tanksley 2004); only *fw4.1*, *fasciated* (*f*) and *ovate* (*o*) had no linked SSR analysed in this study. Of the linked markers, two showed significant associations with FW and/or FS (addressing the regions of *fw1.1* and *fs8.1*; see below) and three were monomorphic. One of these was Tom178–179, a marker located on chromosome 2, in a region which could span both locule number (ln), one of the two loci mainly controlling the number of carpels (Barrero and Tanksley 2004), and *fw2.2*, one of the major QTLs involved in tomato domestication that accounts for approximately 30% of the variance in fruit weight (Frary et al. 2000). Tom178–179 was monomorphic in our 59 cultivated genotypes, but in *S. lycopersicum* var. *cerasiforme* and in *S. pimpinellifolium* it showed two alleles that were different to that found in *S. lycopersicum* (Supplementary Table S3). This confirms the allelic fixation at this genomic region in the cultivated germplasm reported in previous studies (Nesbitt and Tanksley 2002; Tanksley 2004). Allelic fixation in cultivated germplasm for regions spanning key genes for tomato domestication is also suggested, as the majority (four out of five) of the markers that were monomorphic in *S. lycopersicum* but polymorphic in the wild species were Q-SSRs.

Among the associations found with previously known QTLs, the TMS63 marker, located in the genomic region of *fw1.1* (Grandillo et al. 1999) and of several QTLs for fruit shape (Van der Knaap and Tanksley 2003), was actually associated with FS. This was in agreement with the concept that FW and FS show a highly significant correlation (Van der Knaap and Tanksley 2003) and that the presence of QTLs for fruit length (*fl1.1*) and diameter (*fd1.1*) have also been reported in this region (Lippman and Tanksley 2001).

Similarly, the EST253712 marker, which is located near to a reported QTL for fruit weight on chromosome 6 (MF6a, Paterson et al. 1991), was associated with FW and to correlated traits in this study.

The TMS60 marker, which is linked to a QTL for fruit shape (*ljfs7*, Van der Knaap et al. 2002), was associated with IT because of a specific allele in the group of genotypes with pear-shaped fruits that typically shows double or compound inflorescence (IT > 1). It would be of interest to investigate whether this marker also reflects the presence of LD with QTLs for fruit weight and shape located in its proximity, but that do not contribute to its definition as a Q-SSR (*fw7.2*, Grandillo et al. 1999; *sun*, Van der Knaap and Tanksley 2001; *sblk7.1* and *hrt7.1*, Van der Knaap and Tanksley 2003).

As TMS59 is located on chromosome 8 (Areshchenkova and Ganal 2002) at a location that is linked to *fs8.1*, which is a major QTL that defines the largest portion of the phenotypic variation for fruit shape (Grandillo et al. 1999; Ku et al. 2000; Van der Knaap and Tanksley 2003), it is likely that the polymorphism found at this marker locus mirrors different allelic compositions of *fs8.1*. LN and BES are traits that correlate with FS (Fig. 1; Supplementary Table S4). This LD is consistent with the concept that *fs8.1* is located near to the centromere, where recombination is suppressed (Grandillo et al. 1996). *fs8.1* has been reported to control ovary growth before anthesis in only the longitudinal dimension (Ku et al. 2000), although here when this marker was tested against the original fruit diameters, it showed a higher association with ED than PD (not shown). This supports the more recent concept that *fs8.1* has a role in the expression of bumpiness (Van der Knaap and Tanksley 2003), a trait largely dependent on equatorial growth. Thus, if the allelic composition at the TMS59 locus truly reflects different alleles of the QTL, then the role of *fs8.1* can be extended to a more general control of tomato fruit-shape variation, from flat to extremely elongated.

The TMS52 marker, which is linked on chromosome 12 to a QTL for FS and one for LN (Van der Knaap and Tanksley 2003), was associated with PH and PI. The 152-bp allele was shared by the genotypes with fruits with the highest PI values. The same allele also caused an association with FW, although in a structure-dependent fashion, as it was specific to small-fruited genotypes. Because this allele was also present in *S. lycopersicum* var. *cerasiforme* and in a further eight small-sized, cherry-type tomatoes not included in the collection (not shown), this suggests the presence of LD with a QTL responsible for fruit size (indeed, a minor QTL for fruit weight, *fw12.2*, is linked to TMS52; Grandillo et al. 1999). If this is the case, the presence of the 152-bp allele, which probably derives from *S. lycopersicum* var. *cerasiforme*, represents an important contribution towards the expression of the small-round, cherry-type fruit phenotype.

In addition to the associations that have indicated previously reported QTLs, our investigation has also shown that loci defined as NQ-SSRs show association to the phenotype. One of these, the LELE25 marker, has been localised to chromosome 10 (Bredemeijer et al. 2002) and showed association with LN and the correlated traits of SES and FW. While the major genes controlling for LN are on different chromosomes (*f* on chromosome 11, and *ln* on chromosome 2; Barrero and Tanksley 2004), a QTL for locule number has also been localised to chromosome 10 (*lcn10.1*, Van der Knaap and Tanksley 2003), and thus the allelic composition of LELE25 may reflect a linkage with *lcn10.1*.

The present study has demonstrated that tomato landraces that are still grown in Italy maintain a high level of genetic diversity, and that molecular fingerprinting based on SSRs is a very attractive way for the implementation of association analysis in tomato. The landraces therefore offer a choice of genotypes that represent a "pseudo-segregant

population", which is useful for the simultaneous evaluation of numerous traits with multiallelic loci, whereas one or a few traits and only two alleles per locus are available in biparental crosses used in classical QTL analysis (Flint-Garcia et al. 2005). This approach provides a wider perspective on the action of known QTLs, and helps in the identification of new marker-trait relationships.

Acknowledgments The authors wish to thank Dr Paola Crinò, Dr Rosa Rao and the Tomato Genetic Resource Center, Davis (CA, U.S.A.) for providing some of the seed stocks, two anonymous reviewers for their constructive comments on the manuscript and Dr Christopher Berrie for editorial assistance with the English. This study was supported by the Italian Ministry of Education, Universities and Research (MIUR) in the framework of the SCRIGNO project (Sviluppo e Caratterizzazione delle Risorse Genetiche Native in Ortofrutticoltura), and by the Marche Region—ASSAM (Marche Agency for Agro-food Sector Services).

References

Alvarez AE, van de Wiel CCM, Smulders MJM, Vosman B (2001) Use of microsatellites to evaluate genetic diversity and species relationships in the genus *Lycopersicon*. Theor Appl Genet 103:1283–1292

Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, Lister C, Molitor J, Shindo C, Tang C, Toomajian C, Traw B, Zheng H, Bergelson J, Dean C, Marjoram P, Nordborg M (2005) Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. PLoS Genet 1:531–539. doi:10.1371/journal.pgen.0010060

Archak S, Karihaloo JL, Jain A (2002) RAPD markers reveal narrowing genetic base of Indian tomato cultivars. Curr Sci 82:1139–1143

Areshchenkova T (2000) Isolation, characterization and mapping of microsatellites from the tomato genome and their application in molecular analysis of centromeric regions. Ph.D. Dissertation, University of Sachsen-Anhalt, Germany

Areshchenkova T, Ganal MW (2002) Comparative analysis of polymorphism and chromosomal location of tomato microsatellite markers isolated from different sources. Theor Appl Genet 104:229–235

Bailey LH, Tracy WW, Kyle EJ, Watts RL (1960) Tomato. In: Bailey LH (eds) The standard cyclopedia of horticulture. The Macmillan Company, New York, pp 3353–3359

Barrero LS, Tanksley SD (2004) Evaluating the genetic basis of multiple-locule fruit in a broad cross section of tomato cultivars. Theor Appl Genet 109:669–679

Bredemeijer G, Cooke R, Ganal M, Peeters R, Isaac P, Noordijk Y, Rendell S, Jackson J, Röder MS, Wendehake K, Dijcks M, Amelaine M, Wickaert V, Bertrand L, Vosman B (2002) Construction and testing of a microsatellite database containing more than 500 tomato varieties. Theor Appl Genet 105:1019–1026

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177

Broun P, Tanksley SD (1996) Characterization and genetic mapping of simple repeat sequences in the tomato genome. Mol Gen Genet 250:39–49

Causse M, Duffe P, Gomez MC, Buret M, Damidaux R, Zamir D, Gur A, Chevalier C, Lemaire-Chamley M, Rothan C (2004) A genetic map of candidate genes and QTLs involved in tomato fruit size and composition. J Exp Bot 55:1671–1685

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Eshed Y, Zamir D (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics 141:1147–1162

Esquinas-Alcazar J, Nuez F (1995) Situacion taxonomica, domesticacion y diffusion del tomate. In: Nuez F (eds) El cultivo del tomate. Ed Mundi-Prensa, Madrid, Barcelona, Mexico, pp 13–42

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

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

Frary A, Nesbitt TC, Frary A, Grandillo S, van der Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tnaksley SD (2000) *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. Science 289:85–88

Frary A, Xu Y, Liu J, Mitchell S, Tedeschi E, Tanksley SD (2005) Development of a set of PCR-based anchor markers encompassing the tomato genome and evaluation of their usefulness for genetics and breeding experiments. Theor Appl Genet 111:291–312

Fulton TM, Bucheli P, Voirol E, López J, Pétiard V, Tanksley SD (2002) Quantitative trait loci (QTL) affecting sugars, organic acids and other biochemical properties possibly contributing to flavor, identified in four advanced backcross populations of tomato. Euphytica 127:163–177

García-Gusano M, García-Martínez S, Ruiz JJ (2004) Use of SNP markers to genotype commercial hybrids and Spanish local cultivars of tomato. Tomato Genet Coop Rep 54:12–15

García-Martínez S, Andreani L, García-Gusano M, Geuna F, Ruiz JJ (2006) Evaluation of amplified fragment length polymorphism and simple sequence repeats for tomato germplasm fingerprinting: utility for grouping closely related traditional cultivars. Genome 49:648–656

Grandillo S, Ku HM, Tanksley SD (1996) Characterization of *fs8.1*, a major QTL influencing fruit shape in tomato. Mol Breed 2:251–260

Grandillo S, Ku HM, Tanksley SD (1999) Identifying the loci responsible for natural variation in fruit size and shape in tomato. Theor Appl Genet 99:978–987

He C, Poysa V, Yu K (2003) Development and characterization of simple sequence repeat (SSR) markers and their use in determining relationships among *Lycopersicon esculentum* cultivars. Theor Appl Genet 106:363–373

Herrmann D, Boller B, Studer B, Widmer F, Kölliker R (2006) QTL analysis of seed yield components in red clover (*Trifolium pratense* L.). Theor Appl Genet 112:536–545

Huang XQ, Borner A, Röder MS, Ganal MW (2002) Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. Theor Appl Genet 105:699–707

Kraakman ATW, van Eeuwijk FA, Dourleijn CJ, Stam P (2001) Fingerprinting of barley to study yield stability. In: Gallais A, Dillman C, Goldgringer I (eds) Quantitative genetics and breeding methods: the way ahead. Proceed 11th meeting of Eucarpia, section Biometrics in Plant breeding. INRA edition, Paris, pp 117–124

Ku HM, Grandillo S, Tanksley SD (2000) *fs8.1*, a major QTL, sets the pattern of tomato carpel shape well before anthesis. Theor Appl Genet 101:873–878

Lehmann EL, D'Abrera HJM (1975) Nonparametrics: statistical methods based on ranks. Holden-Day series in probability and statistics. Holden-Day, San Francisco

Lippman Z, Tanksley SD (2001) Dissecting the genetic pathway to extreme fruit size in tomato using a cross between the small-fruited wild species *Lycopersicon pimpinellifolium* and *L. esculentum* var. Giant Heirloom. Genetics 158:413–422

Liu K, Muse S (2005) PowerMarker: new genetic data analysis software, Version 3.23. Available via DIALOG. http://www.powermarker.net. Accessed 9 Jan 2007

Lynch M, Walsh B (1997) Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, pp 413

Mazzucato A (1995) Italian germplasm of *Poa pratensis* L. I. Variability and mode of reproduction. J Genet Breed 49:111–118

McIntosh MS (1983) Analysis of combined experiments. Agron J 75:153–155

Miller JC, Tanksley SD (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor Appl Genet 80:437–448

Monti L, Santangelo E, Corrado G, Rao R, Soressi GP, Scarascia Mugnozza GT (2004) Il "San Marzano": problematiche e prospettive in relazione alla sua salvaguardia e alla necessità di interventi genetici. Agroindustria 3:161–169

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. J Mol Evol 91:153–170

Nesbitt TC, Tanksley SD (2002) Comparative sequencing in the genus *Lycopersicon*: implication for the evolution of fruit size in the domestication of cultivated tomatoes. Genetics 162:365–379

Noble D (1994) Tantalizing tomatoes from the past. Am Veget Grower 42:44–47

Noli E, Conti S, Maccaferri M, Sanguineti MC (1999) Molecular characterization of tomato cultivars. Seed Sci Technol 27:1–10

Orloci L (1972) An algorith for cluster seeking in ecological collections. Vegetatio 27:339–345

Park YH, West MAL, St Clair DA (2004) Evaluation of AFLPs for germplasm fingerprinting and assessment of genetic diversity in cultivars of tomato (*Lycopersicon esculentum* L.). Genome 47:510–518

Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinowitch HD, Lincoln SE, Lander ES, Tanksley SD (1991) Mendelian factors underlying quantitative traits in tomato: comparison across species generations and environments. Genetics 127:181–197

Podani J (1993) SYN-TAX-pc Version 5.0, User's Guide. Scientia, Budapest

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Rick CM (1976) Tomato, *Lycopersicon esculentum* (Solanaceae). In: Simmonds NW (eds) Evolution of crop plants. Longman Group, London, pp 268–273

Rick CM, Holle M (1990) Andean *Lycopersicon esculentum* var. *c*erasiforme: genetic variation and ist evolutionary significance. Econ Bot 44:69–78

Ruiz JJ, García-Martínez S, Picó B, Gao M, Quiros CF (2005) Genetic variability and relationship of closely related Spanish traditional cultivars of tomato as detected by SRAP and SSR markers. J Am Soc Hortic Sci 130:88–94

SAS Institute Inc. (2002) SAS guide for personal computers, 9th edn. SAS Institute, Cary

Smulders MJM, Bredemeijer G, Rus-Kortekaas W, Arens P, Vosman B (1997) Use of short microsatellites from database sequences to generate polymorphisms among *Lycopersicon esculentum* cultivars and accessions of other *Lycopersicon* species. Theor Appl Genet 94:264–272

Sneath PHA, Sokal RR (1973) Numerical taxonomy. WH Freeman, San Francisco

Soressi GP (1969) Il pomodoro. Ed Agricole, Bologna

Stevens MA, Rick CM (1986) Genetics and breeding. In: Atherton JG, Rudich J (eds) The tomato crop. Chapman and Hall, London, pp 35–109

Suliman-Pollatschek S, Kashkush K, Shats H, Hillel J, Lavi U (2002) Generation and mapping of AFLP, SSRs and SNPs in *Lycopersicon esculentum*. Cell Mol Biol Lett 7:583–597

Tam SM, Mhiri C, Vogelaar A, Kerkveld M, Pearce SR, Grandbastien M-A (2005) Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon-based SSAP, AFLP and SSR. Theor Appl Genet 110:819–831

Tanksley SD (2004) The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. Plant Cell 16:S181–S189

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

Van de Peer Y, De Wachter R (1993) TREECON: a software package for the construction and drawing of evolutionary trees. Comput Applic Biosci 9:177–182

Van der Knaap E, Tanksley SD (2001) Identification and characterization of a novel locus controlling early fruit development in tomato. Theor Appl Genet 103:353–358

Van der Knaap E, Lippman Z, Tanksley SD (2002) Extremely elongated tomato fruit controlled by four quantitative trait loci with epistatic interactions. Theor Appl Genet 104:241–247

Van der Knaap E, Tanksley SD (2003) The making of a bell pepper-shaped tomato fruit: identification of loci controlling fruit morphology in Yellow Stuffer tomato. Theor Appl Genet 107:139–147

Wang L, Guan R, Zhangxiong L, Chang R, Qiu L (2006) Genetic diversity of chinese cultivated soybean revealed by SSR markers. Crop Sci 46:1032–1038

Williams CE, St. Clair DA (1993) Phenetic relationships and levels of variability detected by restriction fragment length polymorphism and random amplified polymorphic DNA analysis of cultivated and wild accessions of *Lycopersicon esculentum*. Genome 36:619–630

Xu Y, Beachell H, McCouch SR (2005) A marker-based approach to broadening the genetic base of rice in the USA. Crop Sci 44:1947–1959

Zhang N, Xu Y, Akash M, McCouch S, Oard JH (2005) Identification of candidate markers associated with agronomic traits in rice using discriminant analysis. Theor Appl Genet 110:721–729