Characterization of a repertoire of tomato fruit genetic variants in the San Marzano genetic background

Gabriella Dono[a], Maurizio Enea Picarella[a], Clara Pons[b], Enrico Santangelo[c], Antonio Monforte[b], Antonio Granell[b] and Andrea Mazzucato[a,*]

[a] Dept. of Agriculture and Forest Sciences, University of Tuscia, Via S.C. de Lellis snc, 01100 Viterbo (Italy)

[b] Dept. of Plant Genomics and Biotechnology, IBMCP (CSIC-UPV), Ingeniero Fausto Elio s/n Valencia (Spain)

[c] Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria (CREA), via della Pascolare 16, 00016 Monterotondo (Roma, Italy)

***** Corresponding author at: Dept. of Agriculture and Forest Sciences, University of Tuscia, Via S.C. de Lellis snc, 01100 Viterbo, Italy. *E-mail address*: mazz@unitus.it

E-mail addresses: GD, gabriella.dono91@hotmail.it; MEP, picarella@unitus.it; CP, cpons@upvnet.upv.es; ES, enrico.santangelo@crea.gov.it; AMo, amonforte@ibmcp.upv.es; AG, agranell@ibmcp.upv.es

Declarations of interest: none

ABSTRACT

San Marzano (SM) is a worldwide famous tomato Italian traditional landrace characterized by elongated fruits with a dual-purpose use in the fresh and processing market. A repertoire of mutations affecting the fruit and of interest for commercial breeding were introduced into the SM genetic background following backcross schemes. The lines generated included 13 genotypes each carrying a single mutation in genes controlling a) the content of all pigments (*hp-1*, *hp-2*, *pd*), b) of carotenoids (*r*, *t*, *at*, *B*, *B_mo$_B$*), c) of chlorophyll (*gf*), d) of flavonoids (*y*) or e) the ripening process (*Nr*, *rin*, *Gr*). Five lines carrying a combination of two mutations were also included. Analysis of SNP polymorphisms showed that the genetic distance of the lines from the recurrent parent was very variable and not well predicted by the number of backcrosses because it was also a function of the dissimilarity of the donor parent. All the genotypes, together with an SM control, were grown in two consecutive years and characterized for vegetative, reproductive and fruit quality traits. Overall, the studied lines reproduced the SM typical phenotypes, but several differences also emerged as both possible negative or advantageous pleiotropic traits for fresh or processing uses and peeling. High pigment mutations confirmed the negative pleiotropic effects on plant fertility and fruit development described earlier and also negatively affected fruit post-harvest life. These latter defects were also reported in the carotenoid mutant *tangerine*. In contrast, absence of peel pigmentation in the *y* mutant was associated with positive postharvest properties as those fruit presented higher resistance to wrinkling and dehydration. Delayed ripening mutants showed positive post-harvest phenotypes, as expected. In conclusion, the study of the present repertoire of fruit variations in an elongated tomato genotype represents a contribution to expand the study of fruit physiology to unusual fruit types and to breed innovative tomato lines with valuable nutritional and technological properties.

Keywords: Fruit pigmentation mutations; Fruit ripening mutations; Introgression lines; San Marzano; *Solanum lycopersicum*.

Abbreviations: AC, Alisa Craig; BC, backcross; CHL, chlorophyll content; DAT, days after transplanting; FLAV, flavonoid content; FLOW, flowering date; FW, fruit weight; GBS, genotyping by sequencing; IL, introgression line; LD, linkage disequilibrium; NF, number of flowers; NIL, near isogenic line; PCA, principal component analysis; PDO, protected denomination of origin; PH plant heigh; PV, pollen viability; SI, shape index; SL, shelf life;

SM, San Marzano; SxF, seeds per fruit; TGRC, Tomato Genetics Resource Center; WT, wild type; WRINK, wrinkling.

1. Introduction

The development of experimental and breeding plant populations is a prerequisite to genetic and functional studies in plant biology. The tomato (*Solanum lycopersicum* L.) is one of the major vegetable crops and is recognized as a model for the study of fleshy fruit development. Populations of introgression lines (ILs), where specific genomic regions from a wild donor are introgressed with marker assisted selection into a common cultivated genetic background, have been a choice material for the study of quantitative traits related to fruit physiology and quality. Several donor species have been adopted to this purpose, including *S. pennellii* (Eshed and Zamir 1995), *S*. *habrochaites* (Finkers et al. 2007), *S. pimpinellifolium* (Barrantes et al. 2016) and *S. chmielewskii* (Ballester et al. 2016).

Of great experimental interest is also the development of near isogenic lines (NILs), where specific Mendelian mutations are introgressed into a recurrent genetic background by backcross (BC) schemes. Repertoires of NILs are very informative, because the near isogenicity between wild-type and mutant lines allows the comparison of gene effects and the physiological and molecular study of variants of interest. When NIL collections are developed into different recipients, the possibility is open to compare the effect of the same gene in different genetic backgrounds.

In tomato, several mutations have been described and cloned that affect important fruit traits such as pigmentation, ripening and shelf-life (Foolad et al. 2007). Such variants included those involving a general pigment intensification (“high-pigment” genes) and hampered fruit maturation (“delayed-ripening” genes), that have been widely used in breeding modern varieties and hybrids with increased pigments or prolonged shelf-life. Other mutants, such as those affecting the accumulation of single classes of pigments like carotenoids or chlorophyll are common in heirloom and garden varieties but had not been widely adopted into professional cultivars to date. Moreover, few works have been devoted

to combine two or several mutations into a single background line, a practice that can lead to novel genotypes of interest.

Efforts to introgress mutations into the same genetic background in tomato started in the second half of last century with the development of a collection of NILs in cv Ailsa Craig. Over 150 variants were introgressed in Ailsa Craig (Darby 1978; Smith et al. 1983) and more than 350 accessions with the same background are listed in the C.M. Rick Tomato Genetics Resource Center (TGRC) website (http://tgrc.ucdavis.edu). A selection of 11 different NILs carrying fruit colour and ripening mutations together with 18 double mutant combinations were analysed with emphasis on yield in comparison with the recurrent parent (Darby 1978).

Another effort to obtain NILs carrying mutations involved in the synthesis of plant and fruit pigments and in other reproductive aspects has been produced at INRA, France. Twelve NILs obtained by varieties in which the mutations appeared spontaneously or after irradiation and 25 NILs selected by recurrent backcrossing in diversified plant material were thoroughly described (Philouze 1991).

More recently, a collection of more than 80 NILs in the background of cv Micro-Tom have been produced and characterized (Carvalho et al. 2011; Sestari et al. 2014; http://www.esalq.usp.br/tomato/). Although this collection appears of great interest for the study of hormonal and photomorphogenic processes in the tomato plants, it also includes NILs directly involved in the fruit phenotype and is thus of interest for the study of fruit genetics and physiology.

In the early seventies, G.P. Soressi at the Experimental Station for Vegetable Research of the Italian Ministry of Agriculture in Montanaso Lombardo (Lodi, Italy) started an ambitious introgression program to develop repertoires of about 30 tomato fruit mutations derived from his own research and from the collection of L. Butler (University of Toronto, Canada). Five diversified genetic backgrounds, popular at that time, were chosen, including

the fresh market variety Marmande (with flattened fruit), the processing types New Yorker (round, selected for its earliness), Gimar (round, selected for the firmness) and Roma (medium-elongate) and San Marzano with elongate fruit. San Marzano (SM) is one of the most popular Italian tomato landraces, used with the dual-purpose of fresh consumption and processing.

Although its origin is controversial, it is certain that SM was widely cultivated at the beginning of the XX century in the Agro Sarnese Nocerino (province of Naples, Italy) as a preferred variety for peeling (Monti et al. 2004). The SM plant is characterized by indeterminate growth habit and produce fruits of about 60-80 g, with a strong green shoulder and a shape index ranging from 2.0 to 2.4 (Monti et al. 2004; Ercolano et al. 2008). Due to its outstanding agronomic, technological and organoleptic qualities, SM remained popular for more than one century and nowadays it is still inscribed to the Register of Varieties and awarded by EU Protected Denomination of Origin (PDO; Monti et al. 2004). For this reason, the fresh and processed products certified as SM reaches prices by far higher than those attained by standard varieties (García-Martínez et al. 2013). Due to its importance in Italy and all around the globe, the SM type has been the object of genetic studies aimed at discriminating the original types from modern varieties and hybrids that can be similar in phenotype, but very diverse for quality traits and at giving perspectives for traceability (Rao et al. 2006; Caramante et al. 2009; Savo Sardaro et al. 2013). For these reasons, SM was also characterized by biochemical and sensorial profiling (Ercolano et al. 2008) and by a partial resequencing of its genome (Ercolano et al. 2014).

In this paper, we describe the characterization of a repertoire of tomato fruit variants in the traditional SM background, including 13 lines with single introgressions and five lines carrying a combination of two mutations. The genetic distance of single lines from the recurrent parent was estimated by SNP genotyping, in order to define the degree of similarity with the recurrent SM background and provide support to the analysis of

phenotypic traits. Overall, the studied lines reproduced the typical SM phenotypes, but several differences also emerged, including both negative or advantageous pleiotropic traits for fresh or processing uses. The attribution of these traits to the introgressed mutations or to the remaining donor parent background is discussed. The characterization of this collection is valuable for developing lines with novel fruit phenotypes and for studying the biochemical effect of mutant alleles in this genetic background.

2. Materials and Methods

2.1. Plant material and growth conditions

Nineteen tomato lines with SM-type fruits have been studied (Table 1). The genotypes comprised a traditional accession of SM with normal red fruit (WT), 13 single mutant lines affected in different aspects of fruit physiology and five double mutants. The *B_mo*$_B$ line, obtained with the combination of the variant *High beta* (*B*) with its modifier *mo*$_B$, was considered as a single mutant. Introgressed lines were generated by crossing the original WT (an SM accession from Salerno, Italy, collected in 1973) with different donors of the mutations and following BC schemes, where the number of BCs varied from one to five. Positive phenotypic selection was applied during BC generations to recover the introgressed mutation and the recurrent parent phenotype (growth habit, leaf traits, green shoulder, fruit shape). Several cycles of selfing were carried out to maximize and stabilize SM phenotypic traits in the lines. Details on the mutations used are given in Supplementary Table S1.

To combine mutations, single mutants ILs were hand-crossed and the F_1 generation grown to obtain F_2 seed. Double mutant plants were selected based on the expected phenotype and selfed in order to fix the mutations. No further backcross was carried out on

105 double mutants and therefore the degree of backcrossing of these lines was estimated as
106 the mean of the number of BCs of the two parent lines (Table 1).
107 Two replicates of 20 seed for each line were germinated in Petri dishes with 3 ml of
108 sterile water. Germination was monitored after four and ten days. Eight plants per accession
109 at the 4-5th true leaf stage were transplanted in twin rows (100 cm between twins, 60 cm
110 between rows and 50 cm between plants within the row) in an unheated tunnel located at
111 the University of Tuscia's Experimental Farm at Viterbo, Italy (42°260′N, 12°040′E). Plants
112 were grown following standard cultural practices for indeterminate tomatoes, using tutors
113 and weekly removal of lateral shoots. Daily temperature was maintained below 30°C by a
114 ventilation system and the plants were irrigated through a drop irrigation system.
115 The trial was repeated with identical materials and methods for two consecutive years
116 (2017 and 2018).
117 As outgroups for the genotypic analysis, plants of an IL of *S. chmielewskii* in VF145-22-8
118 background (LA1563) and of a *S. pimpinellifolium* accession (LA1589) were grown to
119 extract DNA. The two accessions were obtained from TGRC.
120
121 *2.2. DNA extraction, GBS library preparation and genotyping*
122 Genomic DNA was isolated from young leaves samples with SpeedTools Plant DNA
123 Extraction kit (Biotools, Spain). The GBS was performed by LGC Genomics GmbH
124 (Germany) following the procedure reported by Elshire et al. (2011). Briefly, DNA was
125 digested with the restriction enzyme *Ape*KI and barcoded libraries were prepared by
126 accession and sequenced on an Illumina HiSeq 2000 platform. A total of 3 million 75-bp
127 reads per sample were generated. To obtain variant calls in form of VCF data, the FASTQ
128 reads were trimmed and mapped to Heinz reference genome v2.5. Freebayes SNP caller
129 (Garrison and Marth 2012) was used to call the SNPs on the mapped sequence reads
130 together with some public genome references from *S. pimpinellifolium*. Raw SNPs were

filtered with the maximum missing data of 30% and minimum allele frequency of 0.06. Heterozygous positions were corrected as missing data.

As the original SM genotype used in the crosses was not available, an SM "reference" genotype was composed by filtering only those SNP loci shared by eight SM accessions analyzed within the Traditom EU project, including landraces and registered lines (not shown). Such filtering yielded 1351 SNP positions that were used to conservatively estimate introgressions in the studied lines. The genetic relationships between the SM reference genotype, the 18 ILs studied and two outgroups were analyzed by principal component analysis (PCA) based on the dissimilarity matrix of the available 1351 filtered SNPs using TASSEL 5.0 (Bradbury et al. 2007). The distance was based on the identity by state (IBS) and calculated as (1 – IBS), with IBS defined as the probability that alleles drawn at random from two individuals at the same locus are the same. For loci sharing the same alleles, IBS=1, for loci with different alleles, IBS=0 and for intermediate situations IBS=0.5. The distance between two taxa is 1 - pIBS, with pIBS being the average IBS over all non-missing loci. PCA graphs were composed with CurlyWhirly 1.17.08.31 (https://ics.hutton.ac.uk/software).

To better estimate the genetic relationship among the ILs, the recurrent SM background and the background Ailsa Craig (AC), common to ten single mutant donors, subsets of SNPs have been created by filtering only sites polymorphic within SM and the ILs and only those polymorphic between SM and AC. Heatmaps to depict different alleles have been drawn with Heatmapper (Babicki et al. 2016). In addition, haplotypes were inspected visually and defined when the length of the haplotype was longer than 45 kb. The *Aft_atv* double mutant showed many unique haplotypes, suggesting a more complex breeding history than the rest of mutants, so it was discarded from further analysis. Haplotypes were transformed to genotypic data and Nei's genetic distance (Nei et al., 1983) among mutants

was calculated from the haplotype/genotype matrix. A Neighbor-Joining tree was obtained with PowerMarker 3.5 (Liu and Muse 2005).

2.3. Phenotypic analysis

Plant height (PH) was measured at 45 d after transplanting (DAT). At the same time, the leaf chlorophyll (CHL) and the leaf flavonoid (FLAV) contents were evaluated using the Dualex® scientific device (FORCE-A A, Orsay, France). For each genotype, four plants were selected and five fully developed leaves per plant were chosen. Dualex measurements were carried out at the centre of middle primary leaflets adaxial (upper side) lamina surface, avoiding midribs and reported as µg/cm^2 (Cerovic et al. 2012).

On a single plant basis, flowering time (FLOW) was calculated as the time to the first flower opening (expressed as DAT) and the number of flowers per inflorescence (NF) was counted on the first and the second inflorescence. To estimate pollen viability (PV), two flowers at anthesis were sampled from four plants of each line. PV was evaluated by light microscopy after staining the pollen with two drops of 1% (w/v) acetic orcein solution. A minimum of 100 pollen grains per slide were counted and classified as viable or non-viable based on their stainability and morphology.

For each line, fruits were harvested at full ripening. On eight fruits, the polar and equatorial diameter was measured and the shape index (SI) calculated as their ratio. On the juice obtained extracting the seeds, the total soluble solid content (Brix) was determined by a digital refractometer (MA871, Milwaukee, Milwaukee Instruments, Inc., NC, USA). The seed extracted from each fruit (SxF) was dried and counted.

Eight fruits collected at the full ripe stage were rinsed and analyzed with a Minolta chromameter (CR400, Konica Minolta). Colorimeter reading values of L*, a* and b* were measured using the D65 illuminant and each record was an average of four measurements on each fruit (in the equatorial zone). Later, fruits were divided into two replicates and stored

on plastic plates at room temperature in the dark. The initial fruit weight (FW) was measured and the weight loss was monitored four times at 5 d intervals. The percent of fruit weight remaining after 20 d of storage was referred to as shelf-life (SL). The fruits with severe cracks, considered commercially unacceptable, were discarded. The day of the first wrinkling (WRINK) was also recorded for each single fruit over a period of 40 days. When fruits remained smooth at the end of the experiment, the maximum value (40) was assigned.

All data were collected in both years following the same methodology.

2.4. Statistical analysis

All data were subjected to General Linear model (GLM) analysis. The differences between each line and the WT was assessed using Student's t test at the 5% significance level. For the Dualex data, values were retained with a confidence interval of 95% (± 2σ). Preliminary assumptions of constancy of variance and normal distribution of the data have been met. Data were analysed according to a two-factor design, considering Genotype (G) and Year (Y) as main factors. When the G*Y interaction resulted significant, one-way analyses in single years were carried out in order to discriminate which genotypes were at the origin of such interaction (Supplementary Table S3 and S4). To simplify reading of the data in the main text, allowance was made for the interaction and all genotypes were presented with a single mean value.

The statistical analyses were performed with the SAS software package (SAS® University Edition) and graphs were elaborated with Excel (Microsoft Office 2013).

3. Results

The collection comprised 19 lines (Table 1; Fig. 1). In addition to the original SM (WT), introgressions included variants affecting pigments in general (*hp-1*, *hp-2*, *pd*), specific classes of pigments such as carotenoids (*r*, *t*, *at*, *B*, *B_mo$_B$*), chlorophyll (*gf*) and flavonoids (*y*) and the process of ripening (*Nr*, *rin*, *Gr*). Among the double mutants, two combined *y* with variants of flesh pigments (*r_y* and *gf_y*), giving a light yellow and wine-coloured fruit phenotype, while two combined *gf* with *r* and *hp-2* giving a green and dark brown fruit phenotype respectively. The last double mutant combined two mutations involved in the synthesis of anthocyanins (*Aft* and *atv*) giving a purple fruit phenotype (Table 1). With the exception of *pd* and *mo$_B$*, the gene underlying all the mutations has been identified (Supplementary Table S1). The main effects of these mutations on the tomato fruit phenotype have been described elsewhere (Moore et al. 2002; Levin et al. 2006; Foolad 2007; Barry et al. 2009; Mazzucato et al. 2013).

3.1. Genotypic analysis

After filtering, 1351 SNP positions were used for calculating genetic distances. The distribution of SNPs on the tomato chromosomes was relatively even, except for Chr2, 5 and 6 that presented a higher number of sites (Supplementary Table S2). The distance between SM and most mutant lines ranged from 0.007 (*hp-2*) to 0.067 (*B*), with the exception of the *t*, *rin* and *Aft_atv* lines that showed genetic distances above to 0.100 (Table 1). PCA plotting of the first two principal components (95.7% of variance explained) showed that almost all the lines in SM background were tightly clustered together with the WT, in comparison to the outgroups that mapped outside, together with *t, rin* and *Aft_atv* (Fig. 2, left). When the cluster was relaxed (14.0% of variance explained), it was clear the strong similarity of WT and *hp-2* and that of carotenoid mutants (Fig. 2, right).

As the number of BCs was not significantly related to the genetic distance from the recurrent parent (*P*=0.53; Supplementary Figure S1), we investigated if the similarity of the

introgressed lines to SM was also a function of the donor parent used in crosses. Out of 15 single mutations, ten had a donor background of Ailsa Craig (AC), one of Garim, one of Fireball and three of unknown or hybrid origin (Table 1). Cultivar Garim, the donor of the *hp-2* allele, was an SM-like genotype (Soressi 1975); accordingly, the lines containing *hp-2* were the most similar to the recurrent SM. The donor of the *rin* mutation, cv Fireball, was likely the origin of the large introgression on Chr2 which is unique of the *rin* introgression line (Supplementary Fig. S2).

To better understand the relationship among the ILs, the recurrent SM background and the AC background common to ten donor parents, we focussed only on the 539 SNPs polymorphic amongst the SM lines and on the 129 sites polymorphic between SM and AC (Supplementary Table S2). The line *Aft_atv* was excluded from the analysis as it presented the highest divergence from SM due to both the distance of donor parents and the low number of BCs (Supplementary Figure S2); this high level of polymorphism tended to obscure differences in other genotypes. Considering all SNPs polymorphic amongst the SM introgression lines, it was evident that the two most distant lines (*t* and *rin*), despite having had four BCs (Table 1), maintained big introgressions from parents different from AC (Supplementary Fig. S3A). Notably, the *pd* line carried a conspicuous introgressions on the long arm of Chr3 that offered a candidate position for the underlying gene (Supplementary Fig. S3A). Considering only the SNPs polymorphic between SM and AC, large introgression were evident on Chr6 for the lines with mutations involving carotenoids (*r*, *B*, *B_mo$_B$*, *at*, *gf_r*), on Chr9 (*r*, *B*, *B_mo$_B$*, *y* and its double mutants) and on Chr10 (*hp-1*, *r*, *gf*, *gf_r*; Supplementary Fig. S3B).

Describing haplotypes instead of single SNPs, 22 introgressions could be highlighted, with estimated size ranging from 0.05 to 64.74 Mbp (Supplementary Table S5). Still the amount of introgressed genome was not directly related to the number of BCs, but a tree

constructed on the basis of haplotypes evidenced the similarity of SM and *hp-2*, of the lines containing *y*, of those with *B* and *r* and those with *gf* (Supplementary Fig. S4).

3.2. Vegetative traits

All the lines showed a good germination with no detectable differences in comparison with the WT; only *Aft_atv* showed germinability lower than 80% (data not shown). At 45 DAT, WT plants were on average about 100 cm tall. Eight lines were significantly shorter than the WT, whereas none was significantly taller (Fig. 3A).

The G*Y interaction was significant for both CHL and FLAV (Supplementary Table S3); when data were mediated over years, a leaf CHL content of 32.4 µg/cm^2 was estimated in the WT and no line showed significantly higher values (Fig. 3B). In contrast, ten lines, including three fruit carotenoid and two delayed ripening mutants, had CHL values lower than SM (Fig. 3B).

The average FLAV value measured in the WT was 0.69 µg/cm^2. Among the mutant lines, only *pd* had a significantly lower value, whereas seven lines had a higher value, including the high pigment single mutants, *Aft_atv*, *y* and the two carotenoid mutations *t* and *r* (Fig. 3C).

3.3. Reproductive traits

WT plants produced the first flower about 30 DAT; among the mutants, *pd* and *gf_hp-2* flowered significantly later accordingly to the combined analysis (Table 2). However, considering single years, *B* and *hp-2* showed late flowering in 2017 and *pd* in 2018 (Supplementary Table S4). The *hp-1*, *pd* and three carotenoid mutant lines showed an NF higher than the WT, together with *gf* and *gf_r*, all the ripening mutants and *Aft_atv* (Table 2). In particular, *rin* had a very high NF essentially due to the higher incidence of ramified inflorescences (not shown). PV above 95% was estimated in the WT and did not differed in

ten of the mutant lines. *hp-2*, *pd*, three carotenoid mutants, *y*, *rin* and *Aft_atv* had a significantly lower PV; in the latter line about one fourth of the pollen grains were not viable (Table 2). Interestingly, also the *hp-1* mutant had value of PV lower than 90%, although not different from the WT (Table 2).

3.4. Fruit traits

A significant G*Y interaction was detected also for FW and SI (Supplementary Table S3). Compared to the WT, whose fruits weighted on average about 60 g, no genotype had heavier fruits whereas two, *at* and *rin*, had lighter fruits (Table 2). San Marzano showed an SI of about 1.9; two genotypes had a higher value (*hp-2* and *Aft_atv*), but within the upper limit recognized for the SM PDO (http://www.consorziopomodorosanmarzanodop.it). Five genotypes had fruits less elongated than SM (Table 2).

No significant G*Y interaction was found for SxF and for the Brix value, but highly significant differences were found among genotypes and between years (Supplementary Table S3). Only *at* had a SxF significantly lower than SM and none had significantly higher values (Table 2).

Chromameter analysis resulted in the components of the CIELAB colour space, lightness (L) and chromaticity coordinates (‘‘a’’, a green-to-red scale, and ‘‘b’’, a blue-to-yellow scale). Although the GxY interaction was significant for all three variates (Supplementary Table S3), they behaved consistently in the two years (not shown) and a PCA analysis clearly separated the 19 lines under study (Fig. 4). The first principal component (PC1) was positively loaded by L and b and negatively by a, whereas PC2 was positively loaded by a. The red-fruited lines, having the highest values of a, mapped on the N-W quadrant of the PCA space. Lines with darker fruits, such as those containing lycopene plus chlorophyll or anthocyanins, mapped on the S-W quadrant. Orange and yellow-fruited lines, having high values of b, mapped on the N-E and S-E quadrants respectively (Fig. 4). The *pd* mutant had

311 an intermediate position on the upper part of the graph. Finally, the S-E quadrant also
312 hosted the *rin* mutant, whose green fruits had the highest b value (Fig. 4).
313 For Brix, the *B* line showed on average 0.5° Brix more than SM, although this difference
314 was not significant; four lines had a value lower than SM (Fig. 5A). The shelf-life behaviour
315 of the studied lines was investigated with a post-harvest experiment where FW losses and
316 wrinkling were monitored for 20 and 40 days of storage respectively. The three ripening
317 mutants plus *y_r* had higher SL than SM; in addition, also *y* and *y_gf* showed higher values
318 although not significant. Four genotypes (the mutants involving *hp-2* plus *t* and *Aft_atv*)
319 showed SL lower than SM (Fig. 5B). All the mutants with delayed ripening and those
320 involving the *y* mutation also showed a resistance to wrinkling higher than the WT, whereas
321 the high-pigment mutants, together with *B_moB*, *gf*, *pd* and *t* had lower resistance (Fig. 5C).
322
323
324 **4. Discussion**
325
326 *4.1. General features of the studied lines*
327 This study addressed the analysis of genotypic and phenotypic variation in a repertoire of
328 tomato lines harbouring different variants affecting fruit traits in the genetic background of
329 the SM Italian landrace. Each of these lines had undergone a variable number of BCs to the
330 recurrent parent and of selfing generations in order to fix all SM typical traits, such as
331 indeterminate habit, elongated fruit shape and strong green shoulder.
332 GBS analysis based on 1351 SNPs indicated that the number of BCs was not a good
333 predictor of the genetic distance from the recurrent parent, as conspicuous regions from the
334 donor parent remains introgressed also after several backcrosses in some genotypes. The *t*
335 line despite being a BC_4 generation, was genetically positioned rather far from the recurrent
336 WT, whereas genotypes with a lower number of BCs (*hp-1*, *hp-2*, *pd*) were much closer.

This was well explained by considering the genetic distance of the donor parent as *hp-2* was obtained in a SM-like background and *t* had an introgression in Chr6 that did not belong to AC and should have been present in the donor possibly as a heritage of cv Tangerine where the gene was first described (Tomes 1952). Accordingly, the distance of the *rin* line was explained by a big introgressions on Chr2 likely inherited from Fireball.

The background selection for indeterminate habit (Self pruning, Chr6), green shoulder (Uniform, Chr10) and elongate shape (ovate mutation, Chr2) was not in conflict with the genes in foreground selection whose genomic location is known except for *pd*. This may have favoured conspicuous linkage drag from the donor backgrounds. The only strict linkage was on Chr6 between *Sp* (45.97 Mbp) and *B* (45.90 Mbp). Although it is not known if there was a recombination between *B* and *Sp* in the Ailsa Craig line donor of *B*, it was certainly not problematic to select *Sp_B* lines as the two dominant alleles are expected to be in coupling in the ancestral wild species where *B* was derived from.

Although six BCs theoretically allow the recovery of >99% of the recurrent background, it is expected that much higher introgressions remain due to linkage or lack of negative selection (Stam and Zeven 1981). Thus, reliable NILs can only be obtained with strong background marker assisted selection or, as emerging, with new breeding techniques, by editing the genome of WT genotypes to recreate variations of fruit traits. In tomato, proof-of-concept that CRISPR/Cas9 technology allows to produce collection of variants at target loci has been advanced (Jacobs et al., 2017) and fruit variants such as *alcobaca* (Yu et al. 2017), *y* (Deng et al. 2018), *r* and *t* (Dahan-Meir et al. 2018) have been recapitulated by gene editing.

In the case reported here, as in other analysis of NIL collections (Darby 1978; Carvalho et al. 2011), the reported phenotypic differences could be due either to the mutation studied in the foreground, to the genetic background remaining from the donor parent or to both. The comparison of phenotypes produced by the same mutation in different genetic

backgrounds can be informative to elucidate which is the variant effect. At the same time, definition of traits that are characteristic of specific genotypes is important in the genetic improvement of landraces (Casañas et al. 2017).

4.2. Features of lines involving all pigments

Phenotypes of mutants involving all pigments were consistent with the knowledge that high-pigment mutations intensify chlorophyll and flavonoid contents (Yen et al. 1997; Mustilli et al. 1999; Bino et al. 2005; Levin et al. 2006). However, it was interesting to note that, in SM background, *hp-1* did not cause a significant increase in leaf CHL as did *hp-2*, albeit not significantly, whereas the high pigment mutants (*hp-1* and *hp-2*) behaved as expected, determining an increased content of leaf FLAV. Indeed, even if the suggested interaction between DET1 and DDB1 (Schroeder et al. 2002) reinforces the hypothesis that the roles of these proteins may have evolved from a common mechanism for facing light stress, both proteins had already been well characterized separately, and associated to their own different phenotypic properties (Mustilli et al. 1999; Lieberman et al. 2004).

The colorimetric analysis well described the pigment mutants, where the most important change in the values of L*, a* and b* regarded the a* value, related to chlorophyll degradation and lycopene synthesis (López Camelo and Gómez 2004). In fact, *hp-1* and *hp-2* mapped on the N-W side of the PCA because of their high positive a values, which correspond to their stronger red chromatic tones. *gf_hp-2* had lower a values correlated with shades of green, related to the “brownish” color of the berry.

Antithetic to high-pigment genotypes, the genetically anonymous variant *pd* diluted all pigments (Tigchelaar et al. 1970). A better description of the pigment composition of this genotype would be desirable, because the low flavonoid content reported here contrasts with the very high polyphenol content found in a previous study (Minoggio et al. 2003). Differently, the low PH found in *pd* agrees with the reported information that this line shows

389 semi-determinate habit (Minoggio et al. 2003) and the BC selection failed to fully recover the
390 indeterminate phenotype after backcrossing.

391 Lines containing *hp-2*, showed the known undesirable pleiotropic effects on fertility
392 (Mustilli et al. 1999), such as late FLOW and low PV. This indicated that the pleiotropic
393 effects that affect this mutation are entirely reproduced in the elongate SM background. On
394 the other side, although not significantly, *hp-1* increased FW in elongated fruits, a
395 phenotype also reported in Micro-Tom NILs (Carvalho et al. 2011).

396

397 *4.3. Features of lines involving carotenoids*

398 Three single and two double mutants involved in fruit carotenoid accumulation showed a
399 leaf CHL content lower than the WT, whereas two had higher leaf FLAV content, indicating
400 that the variation in one class of pigments may significantly impact the levels of other
401 classes. Apparently, a crosstalk among different metabolic pathways may be hypothesized
402 and some findings seems to support this hypothesis (Minoggio et al. 2003; Pal et al. 2019).
403 However, other evidences tend to limit the extent of the reciprocal influence between the
404 phenylpropanoid/flavonoid and carotenoid pathways (Long et al. 2006), thus giving an input
405 for deeper studies.

406 Because the value of L* indicates the brightness and a decreasing L* value indicates the
407 darkening of red color (López Camelo and Gómez 2004), the carotenoid mutant lines
408 mapped in the N-E part of the PCA graph, characterized by positive and high L* values.

409 The WT in our experiments scored a mean Brix value of 5.8, in line with values reported
410 previously in SM (Ercolano et al. 2008; Baldina et al. 2016). *B* was the only mutation to
411 show a Brix value higher than the WT (6.24) although this difference was not significant. *B*
412 was also characterized by an increased NF and by a delay in FLOW, a phenotype that was
413 also observed in *B* introgressions in different genetic backgrounds (A. Mazzucato,
414 unpublished data). Also, the *t* mutation, which deserves high interest because it

accumulates pro-lycopene which has been involved in nutritional advantages (Unlu et al. 2007), showed several undesirable traits such as lower Brix value and inferior post-harvest properties. Thus, breeding *B* or *t* orange tomatoes should take into consideration these drawbacks and try to counteract negative pleiotropic effects by genetic or agronomic means.

4.4. Features of lines involving chlorophyll and flavonoids

Compared with the recurrent line, the "*stay-green*" mutation *gf* was characterized by lower CHL (at an early growth stage). This was not surprising since the *gf* phenotype is based on a class C stay-green mutation ("cosmetic" stay-green) that is deficient in its ability to break down chlorophyll, not to increase chlorophyll synthesis (Hörtensteiner 2009). In fact, it is well reported how the effects of the *gf* mutation are confined to the senescence phase, which includes numerous degradative events, mostly associated with the disintegration of the photosynthetic apparatus (Akhtar et al., 1999). Chlorophyll loss in leaves and mature fruits is compromised, since thylakoid grana and light-harvesting chlorophyll-binding proteins persist during senescence (Barry and Pandey 2009; Hörtensteiner and Kräutler 2010). Indeed, the *gf* fruits retains visibly a substantial amount of chlorophyll during ripening. The high FN showed by the *gf* line is likely an effect of the genetic background since other double mutants involving the same gene had a NF comparable to SM.

The *colorless epidermis* line showed very little departures from the SM ideotype. Alone or in combination, the *y* mutant showed a higher resistance to storage, indicating that pigment variation in the peel implicates different mechanical properties and post-harvest behavior of the fruit. A higher resistance to wrinkling in *y* mutants was also reported previously and the peculiar mechanical properties of the *y* epicarp were clearly manifested by the fact that the peel of the mutant fruit was richer of lignin (Adato et al. 2009;

441 Dominguez et al. 2009). Indeed, low levels of polyphenols induced by silencing of chalcone
442 synthase (CHS) reduced the ability of the fruit to deform and decreased cuticle permeability
443 (España et al. 2014). Thus, the *y* mutant phenocopies variants with delayed ripening for the
444 resistance to wrinkling, although different underlying genetic mechanisms are responsible
445 for this phenotype in different lines.

446 Among the double mutants, the *Aft_atv* combination, giving purple fruits, was the one
447 with the highest departures from the original WT, differing for nearly all the traits that were
448 taken into consideration. Part of this variation, e.g. that for SI, is likely due to the low level of
449 backcrossing of this genetic combination and the genetic distance from SM still inherent in
450 this line. The extended SL of purple fruits is an interesting character that was reported in
451 round-fruited backgrounds (Bassolino et al. 2013; Mazzucato et al. 2013; Borghesi et al.
452 2016). In SM, however, a lower resistance to dehydration was reported, in disagreement
453 with previous data (Bassolino et al. 2013; Zhang et al. 2013). Therefore, further
454 investigation is needed in order to assess if the better post-harvest performances of purple
455 tomatoes can be generalized or if they are dependent on the fruit shape and, more
456 generally, on the genetic background.

457

458 *4.5. Features of lines with delayed ripening*

459 As expected, all the mutations for delayed ripening had higher SL and WRINK compared
460 with the WT. Lines with delayed ripening in this genetic background will help the breeding of
461 SM hybrids with the underlying genes in heterozygous state, the conditions in which they
462 are commonly used in modern cultivars.

463 Mutants for delayed ripening also showed the pleiotropic phenotype of an increased
464 NF, due to the occurrence of compound inflorescences. In the *rin* line, we observed large
465 sepals and indeterminate inflorescences as expected because the original *rin* mutation also
466 affects the *MACROCALYX* gene, a MADS-box transcription factor with a role in sepal size

and inflorescence determinacy regulation (Vrebalov et al. 2002; Samach et al. 2007). However, in the SM *rin* line the phenotype also included longer and bifurcated inflorescences that caused an increase in NF.

5. Conclusions

The collection described here represents an original repertoire of useful alleles into SM, a dual-purpose tomato cultivar with elongate fruit well appreciated in Italy and all over the world. Indeed, this material would be valuable for comparison of morphological, physiological and agronomic traits among variants within this tomato type. Evaluating the same variants in different genotypes will provide additional insights into the phenotype/background interactions. Biochemical characterization of this collection, which is under way, will give further insights on the effect of each mutation on fruit aesthetic, technological and flavor and nutritional properties. As a considerable interest exists for breeding novel tomato genotypes, the described collection represents a precious material to combine two or several mutations in SM and select tomato lines with new phenotypes.

Acknowledgments: The authors acknowledge Gian Piero Soressi, who developed most of the lines used in the experiments. We are also grateful to Domenico Grossi, for expert technical assistance in growing the plants, and Emiliano Chiaretti and Emanuele Radicetti for assistance in data measurement. The C.M. Rick Tomato Genetics Resource Center (TGRC, University of California, Davis, CA, USA) is acknowledged for seed supply and two anonymous reviewers for their constructive comments on the manuscript.

Funding: This work was supported by the Latium Region FILAS project “MIGLIORA” and by the Italian Ministry of Agriculture (MiPAAF) under the AGROENER project (D.D. n. 26329, 1 april 2016) - http://agroener.crea.gov.it/ and by the European Commission through-H2020 SFS-7a-2014 TRADITOM (634561).

Author Contributions: E.S. and A.Ma. conceived and designed the experiments; G.D. and M.E.P. performed the experiments; G.D., C.P., A.Mo., and A.Ma. analyzed the data; G.D., A.G. and A.Ma. drafted the paper. All authors critically read and approved the final version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

Adato, A., Mandel, T., Mintz-Oron, S., Venger, I., Levy, D., Yativ, M., Domínguez, E., Wang, Z., De Vos, R.C.H., Jetter, R., Schreiber, L., Heredia, A., Rogachev, I., Aharoni, A., 2009. Fruit-surface flavonoid accumulation in tomato is controlled by a SlMYB12-regulated transcriptional network. PLoS Genet. 5(12), e1000777, http://doi:10.1371/journal.pgen.1000777.

Akhtar, M.S., Goldschmidt, E.E., John, I., Rodoni, S., Matile, P., Grierso, D., 1999. Altered patterns of senescence and ripening in gf, a stay-green mutant of tomato (*Lycopersicon esculentum* Mill.). J. Exp. Bot. 50(336), 1115-1122, doi.org/10.1093/jxb/50.336.1115.

Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J.R., Maciejewski, A., Wishart, D.S., 2016. Heatmapper: web-enabled heat mapping for all. Nucl. Acids Res. 44(Web Server issue): W147–W153, http://doi:10.1093/nar/gkw419

Baldina, S., Picarella, M.E., Troise, A.D., Pucci, A., Ruggieri, V., Ferracane, R., Barone, A., Fogliano, V., Mazzucato, A., 2016. Metabolite profiling of Italian tomato landraces with different

519 fruit types. Front. Plant Sci. 7, 664, http://doi:10.3389/fpls.2016.00664.
520 Barrantes, W., López-Casado, G., García-Martínez, S., Alonso, A., Rubio, F., Ruiz, J.J., Fernández-
521 Muñoz, R., Granell, A., Monforte, A.J., 2016. Exploring new alleles involved in tomato fruit quality
522 in an introgression line library of *Solanum pimpinellifolium*. Front. Plant Sci. 7, 1–12,
523 http://doi:10.3389/fpls.2016.01172.
524 Barry, C.S., Pandey, P.A., 2009. survey of cultivated heirloom tomato varieties identifies four new
525 mutant alleles at the *green-flesh* locus. Mol. Breed. 24(3), 269-276, http://doi:10.1007/s11032-
526 009-9289-4.
527 Ballester, A.R., Tikunov, Y., Molthoff, J., Grandillo, S., Viquez-Zamora, M., de Vos, R., de Maagd
528 R.A., van Heusden S., Bovy, A.G., 2016. Identification of loci affecting accumulation of secondary
529 metabolites in tomato fruit of a *Solanum lycopersicum* × *Solanum chmielewskii* introgression line
530 population. Front. Plant Sci. 7, http://doi:10.3389/fpls.2016.01428.
531 Bassolino, L., Zhang, Y., Schoonbeek, H.J., Kiferle, C., Perata, P., Martin, C., 2013. Accumulation of
532 anthocyanins in tomato skin extends shelf life. New Phytol. 200(3), 650-655,
533 http://doi:10.1093/jhered/esg093, http://doi:10.1111/nph.12524.
534 Bino, R. J., De Vos, C. H., Lieberman, M., Hall, R. D., Bovy, A., Jonker, H.H., Tikunov, Y., Lommen,
535 A., Moco, S., Levin, I., 2005. The light-hyperresponsive *high pigment-2*^{dg} mutation of tomato:
536 alterations in the fruit metabolome. New Phytol. 166: 427-438. http://doi:10.1111/j.1469-
537 8137.2005.01362.
538 Borghesi, E., Ferrante, A., Gordillo, B., Rodríguez-Pulido, F.J., Cocetta, G., Trivellini, A., Mensuali-
539 Sodi, A., Malorgio, F., Heredia, F.J., 2016. Comparative physiology during ripening in tomato rich-
540 anthocyanins fruits. Plant Growth Regul. 80(2), 207-214., http://doi:10.1007/s10725-016-0158-y.
541 Bradbury, P.J., Zhang, Z,, Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler E.S., 2007.
542 TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics.
543 23:2633-2635.
544 Caramante, M., Rao, R., Monti, L.M., Corrado, G., 2009. Discrimination of ‘San Marzano’
545 accessions: a comparison of minisatellite, CAPS and SSR markers in relation to morphological
546 traits. Sci. Hort. 120(4), 560-564, doi.org/10.1016/j.scienta.2008.12.004.

547 Carvalho, R.F., Campos, M.L., Pino, L.E., Crestana, S.L., Zsögön, A., Lima, J.E., Vagner, A.B.,
548 Peres, L.E., 2011. Convergence of developmental mutants into a single tomato model system:
549 'Micro-Tom' as an effective toolkit for plant development research. Plant Methods. 7, 18,
550 doi.org/10.1186/1746-4811-7-18.
551 Casañas, F., Simó, J., Casals, J., Prohens, J., 2017. Toward an evolved concept of landrace. Front.
552 Plant Sci. 8:145. http://doi: 10.3389/fpls.2017.00145.
553 Cerovic, Z.G., Masdoumier, G. , Ghozlen, N.B., Latouche, G., 2012. A new optical leaf-clip meter for
554 simultaneous non-destructive assessment of leaf chlorophyll and epidermal flavonoids. Physiol.
555 Plant. 146: 251-260. http://doi:10.1111/j.1399-3054.2012.01639.x.
556 Dahan-Meir, T., Filler-Hayut, S., Melamed-Bessudo, C., Bocobza, S., Czosnek, H., Aharoni, A.,
557 Levy, A.A., 2018. Efficient in planta gene targeting in tomato using geminiviral replicons and the
558 CRISPR/Cas9 system. Plant J. 95:5-16. http://doi: 10.1111/tpj.13932.
559 Darby, L.A., 1978. Isogenic lines of tomato fruit colour mutants. Hort. Res. 18, 73-84.
560 Deng, L., Wang, H., Sun, C., L,i Q., Jiang, H., Du, M., Li, C.B., and Li, C., 2018. Efficient generation
561 of pink-fruited tomatoes using CRISPR/Cas9 system. J. Genet. Genom. 45, 51-54. http://doi:
562 10.1016/j.jgg.2017.10.002. Epub 2017 Nov 6.
563 Dominguez, E. , Lòpez-Casado, G., Cuartero, J., Ramìrez, L.E., 2009. Development of fruit cuticle
564 in cherry tomato (*Solanum lycopersicum*). Funct. Plant Biol. 36, 613–620.
565 Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E., 2011.
566 A robust simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS
567 ONE. 6(5): e19379. http://doi:10.1371/journal.pone.0019379
568 Ercolano, M.R., Carli, P., Soria, A., Cascone, A., Fogliano, V., Frusciante, L., Barone, A., 2008.
569 Biochemical, sensorial and genomic profiling of traditional Italian tomato varieties. Euphytica.
570 164(2), 571-582.
571 Ercolano, M.R., Sacco, A., Ferriello, F., D'Alessandro, R., Tononi, P., Traini, A., Barone A., Zago E.,
572 Chiusano M.L., Buson G., Delledonne M., Frusciante L., 2014. Patchwork sequencing of tomato
573 San Marzano and Vesuviano varieties highlights genome-wide variations. BMC Genom. 15(1), 1–
574 13, http://doi:10.1186/1471-2164-15-138.

575 Eshed, Y., Zamir, D., 1995. An introgression line population of *Lycopersicon pennellii* in the
576 cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics.
577 141(3), 1147-1162.
578 España, L., Heredia-Guerrero, J.A., Reina-Pinto, J.J., Fernández-Muñoz, R., Heredia, A.,
579 Domínguez, E., 2014. Transient silencing of CHALCONE SYNTHASE during fruit ripening
580 modifies tomato epidermal cells and cuticle properties. Plant Physiol. 166(3), 1371-1386,
581 http://doi:10.1104/pp.114.246405.
582 Finkers, R., van Heusden, A.W., Meijer-Dekens, F., van Kan, J.A., Maris, P., Lindhout, P., 2007. The
583 construction of a *Solanum habrochaites* LYC4 introgression line population and the identification
584 of QTLs for resistance to Botrytis cinerea. Theor. Appl. Genet. 114(6), 1071-1080, doi
585 10.1007/s00122-006-0500-2.
586 Foolad, M.R., 2007. Genome mapping and molecular breeding of tomato. Intl. J. Plant Genom.
587 Article ID 64358, http://doi:10.1155/2007/64358.
588 García-Martínez, S., Corrado, G., Ruiz, J.J., Rao, R., 2013. Diversity and structure of a sample of
589 traditional Italian and Spanish tomato accessions. Genet. Res. Crop Evol. 60(2), 789-798,
590 doi.org/10.1007/s10722-012-9876-9.
591 Garrison E., Marth G., 2012. Haplotype-based variant detection from short-read sequencing. arXiv
592 preprint arXiv:1207.3907.
593 Jacobs T.B., Zhang N., Patel D., Martin G.B., 2017. Generation of a collection of mutant tomato
594 lines using pooled CRISPR libraries. Plant Physiol. 174(4), 2023-2037. http://doi:
595 10.1104/pp.17.00489.
596 Hörtensteiner, S, 2009. Stay-green regulates chlorophyll and chlorophyll-binding protein degradation
597 during senescence. Trends Plant Sci. 14: 155-62.
598 Hörtensteiner, S., Kräutler, B., 2011. Chlorophyll breakdown in higher plants. Biochim. Biophys.
599 Acta. 1807(8), 977-88, http://doi: 10.1016/j.bbabio.2010.12.007.
600 Levin, I., De Vos, C.R., Tadmor, Y., Bovy, A., Lieberman, M., Oren-Shamir, M., Segev O., Kolotilin
601 I., Keller M., Ovadia R., Meir, A., Bino R.J., 2006. High pigment tomato mutants - more than just
602 lycopene (a review). Isr. J. Plant Sci. 54(3), 179-190.

603 Lieberman, M., Sege, O., Gilboa, N., Lalazar, A., Levin I., 2004. The tomato homolog of the gene
604 encoding uv-damaged dna binding protein 1 (ddb1) underlined as the gene that causes the *high*
605 *pigment-1* mutant. Theor. Appl. Genet. 108, 1574-1581. https://doi.org/10.1007/s00122-004-
606 1584-1.
607 Liu, K., Muse, S.V., 2005. PowerMarker: integrated analysis environment for genetic marker data.
608 Bioinformatics 21, 2128-2129. https://doi.org/10.1093/bioinformatics/bti282.
609 Long, M., Millar, D.J., Kimura, Y., Donovan, G., Rees, J., Fraser, P.D., Bramley, P.M., Bolwell, G.P.,
610 2006. Metabolite profiling of carotenoid and phenolic pathways in mutant and transgenic lines of
611 tomato: Identification of a high antioxidant fruit line. Phytochemistry. 67, 1750–1757,
612 http://doi:10.1016/j.phytochem.2006.02.022.
613 López Camelo, A.F., Gómez, P.A., 2004. Comparison of color indexes for tomato ripening. Hort.
614 Bras. Brasília. 22 (3), 534-537.
615 Mazzucato, A., Willems, D., Bernini, R., Picarella, M.E., Santangelo, E., Ruiu, F., Tilesi, F., Soressi,
616 G.P., 2013. Novel phenotypes related to the breeding of purple-fruited tomatoes and effect of
617 peel extracts on human cancer cell proliferation. Plant Physiol. Biochem. 72, 125-133, http://doi:
618 10.1016/j.plaphy.2013.05.012.
619 Minoggio M., Bramati L., Simonetti P., Gardana C., Iemoli L., Santangelo, E., Mauri P.L., Spigno P.,
620 Soressi G.P. Pietta, P.G., 2003. Polyphenol pattern and antioxidant activity of different tomato
621 lines and cultivars. Ann. Nutr. Metabol. 47(2), 64-69, http://doi:10.1038/385718a0.
622 Monti, L.M., Santangelo, E.: Corrado, G., Rao, R., Soressi, G.P., Scarascia Mugnozza, G.T., 2004. Il
623 “San Marzano”: problematiche e prospettive in relazione alla sua salvaguardia e alla necessità di
624 interventi genetici. Agroindustria. 3(2), 161-170.
625 Moore, S., Vrebalov, J., Payton, P., Giovannoni, J., 2002. Use of genomics tools to isolate key
626 ripening genes and analyse fruit maturation in tomato. J. Exp. Bot. 53(377), 2023-2030,
627 http://doi:10.1093/jxb/erf057.
628 Mustilli, A.C., Fenzi, F., Ciliento, R., Alfano, F., Bowler, C., 1999. Phenotype of the tomato *high*
629 *pigment-2* mutant is caused by a mutation in the tomato homolog of DEETIOLATED1. Plant Cell.
630 11(2), 145-157, doi.org/10.1105/tpc.11.2.145.

631 Nei, M., Tajima, F., Tateno, Y., 1983. Accuracy of estimated phylogenetic trees from molecular data.
632 II. Gene frequency data. J. Mol. Evol. 19, 153-70.
633 Pal, H., kundu, A., Sahu, R., Sethi, A., Hazra, P., Chatterjee, S., 2019. Unraveling the metabolic
634 behavior in tomato high pigment mutants (*hp-1*, *hp-2*^{dg}, *og*^{c}) and non ripening mutant (*rin*) during
635 fruit ripening. Sci. Hort. 246, 652-663.
636 Philouze, J., 1991. Description of isogenic lines, except for one, or two, monogenically controlled
637 morphological traits in tomato, *Lycopersicon esculentum* Mill. Euphytica. 56(2), 121-131.
638 Rao, R., Corrado, G., Bianchi, M., Di Mauro, A., 2006. $(GATA)_4$ DNA fingerprinting identifies
639 morphologically characterized 'San Marzano' tomato plants. Plant Breed. 125(2), 173-176,
640 http://doi:10.1111/j.1439-0523.2006.01183.x.
641 Samach, A., Lotan, H., 2007. The transition to flowering in tomato. Plant Biotechnol. 24(1), 71-82.
642 Savo Sardaro, M.L., Marmiroli, M. , Maestri, E. and Marmiroli, N., 2013. Genetic characterization of
643 Italian tomato varieties and their traceability in tomato food products. Food Sci. Nutr. 1: 54-62.
644 http://doi:10.1002/fsn3.8.
645 Schroeder, D.F., Gahrtz, M., Maxwell, B.B., Cook, R.K., Kan, J.M., Alonso, J.M., Ecker, J.R., Chory,
646 J., 2002. De-etiolated 1 and damaged DNA binding protein 1 interact to regulate Arabidopsis
647 photomorphogenesis. Curr. Biol. 12(17), 1462-72.
648 Sestari, I., Zsögön, A., Rehder G.G., de Lira Teixeira L., Mariko Aymoto Hassimoto N., Purgatto E.,
649 Vagner, A.B., Peres, L.E., 2014. Near-isogenic lines enhancing ascorbic acid, anthocyanin and
650 carotenoid content in tomato (*Solanum lycopersicum* L. cv Micro-Tom) as a tool to produce
651 nutrient-rich fruits. Sci. Hort. 175, 111-120, doi.org/10.1016/j.scienta.2014.06.010.
652 Smith, J.M., Ritchie, D.B., 1983. A collection of near-isogenic lines of tomato: research tool of the
653 future? Plant Mol. Biol. Rep. 1(1), 41-45.
654 Soressi, G.P., 1975. New spontaneous or chemically-induced fruit ripening tomato mutants. Rep.
655 Tomato Genet. Coop. 25 21–22.
656 Stam, P., Zeven, A.C., 1981. The theoretical proportion of the donor genome in near-isogenic lines
657 of self-fertilizers bred by backcrossing. Euphytica. 30: 227-238.
658 Tigchelaar, E.C., Tomes, M.L., Erickson, H.T., Graham, T.O., Barman, R.J., 1970. "Pigment diluter"

(*pd*), a new plant and fruit color mutant. Tom. Genet. Coop. 20,64.

Tomes, M.L., 1952). Flower color modification associated with the gene *t*. Tom. Genet. Coop. 2,12.

Unlu, N.Z., Bohn, T., Francis, D., Clinton, S.K., Schwartz, S.J., 2007. Carotenoid absorption in humans consuming tomato sauces obtained from tangerine or high-β-carotene varieties of tomatoes. J. Agric. Food Chem. 55(4), 1597-1603, http://doi:10.1021/jf062337b.

Vrebalov, J., Ruezinsky, D., Padmanabhan, V., White, R., Medrano, D., Drake, R., Schuch, W., Giovannoni, J.A., 2002. MADS-box gene necessary for fruit ripening at the tomato *ripening-inhibitor* (*rin*) locus. Science. 296(5566), 343-346, http://doi:10.1126/science.1068181V.

Yen, H.C., Shelton, B.A., Howard, L.R., Lee, S., Vrebalov, J., Giovannoni, J.J., 1997. The tomato *high-pigment* (*hp*) locus maps to chromosome 2 and influences plastome copy number and fruit quality. Theor. Appl. Genet. 95, 1069-1079, doi.org/10.1007/s001220050664.

Yu, Q.-h., Wang, B., Li N., Tang, Y., Yang, S., Yang, T., Xu, J., Guo, C., Yan P., Wang Q. et al., 2017. CRISPR/Cas9-induced Targeted Mutagenesis and Gene Replacement to Generate Long-shelf Life Tomato Lines. Sci. Rep. 7, 11874.

Zhang, Y., Butelli, E., De Stefano, R., Schoonbeek, H.-J., Magusin, A., Pagliarani, C., Wellner, N., Hill, L., Orzaez, D., Granell, A., Jones, J.D.G., Martin, C., 2013. Anthocyanins double the shelf life of tomatoes by delaying overripening and reducing susceptibility to gray mold. Current Biology, 23, 1094–1100. http://doi: 10.1016/j.cub.2013.04.072.

References to websites

C.M. Rick Tomato Genetics Resource Center

http://tgrc.ucdavis.edu

(accessed 13rd March 2019)

Consorzio di tutela del Pomodoro San Marzano dell'agro Sarnese Nocerino

http://www.consorziopomodorosanmarzanodop.it

(accessed on 13rd March 2019)

687 Information & Computational Sciences - James Hutton Institute

688 https://ics.hutton.ac.uk/software

689 (accessed on 25th January 2019)

690

691 Micro-Tom Mutants HCPD Lab

692 http://www.esalq.usp.br/tomato/

693 (accessed 27th May 2019)

Table 1

List of the 19 lines with a San Marzano genetic background and of the outgroups used in the study, divided according to the class of variation, extended names of the mutations, genetic symbols used, number of backcrosses (BCs) carried out with the recurrent parent, number of selfing generations (Self), genetic distance (D) from SM estimated after GBS analysis and genetic background of donors.

Class of material	Class of variation	Name	Genetic symbol	No. of BCs	No. of Selfs	D	Donor parent background
Wild-type	- [a]	San Marzano	WT	-	-	-	-
San Marzano fruit variants	All pigments	*high pigment-1*	*hp-1*	2	5	0.024	Ailsa Craig (AC)
		high pigment-2	*hp-2*	2	6	0.007	Garim
		pigment diluter	*pd*	1	4	0.044	Unkown or hybrid
	Carotenoids	*yellow flesh*	*r*	5	6	0.056	AC
		tangerine	*t*	4	6	0.157	AC
		apricot	*at*	1	4	0.049	AC
		High Beta	*B*	1	2	0.067	AC
		High Beta + Beta modifier	*B_mo$_B$*	2	4	0.062	AC
	Chlorophyll	*green flesh*	*gf*	4	5	0.025	AC
	Flavonoids	*colourless fruit epidermis*	*y*	3	2	0.018	AC
	Ripening	*Never ripe*	*Nr*	4	5	0.015	AC
		ripening inhibitor	*rin*	4	5	0.132	Fireball
		Green ripe	*Gr*	2	5	0.035	Unkown or hybrid
	Double mutants	*yellow flesh + colourless fruit epidermis*	*r_y*	4 [b]	4	0.022	AC / AC
		green flesh + colourless fruit epidermis	*gf_y*	3.5	4	0.048	AC / AC
		green flesh + yellow flesh	*gf_r*	4.5	4	0.052	AC / AC
		green flesh + high pigment-2	*gf_hp-2*	3	4	0.017	AC / Garim
		Anthocyanin fruit + atroviolaceum	*Aft_atv*	1	3	0.268	Unknown / AC
Outgroups	-	*S. chmielewski* IL	Sc IL	-	-	0.619	-
		S. pimpinellifolium	Sp	-	-	0.915	-

[a] Not applicable

[b] In double mutants, the number of BCs has been assigned as the mean of BCs carried out in the two parent line.

700 **Table 2**

701 Flowering date (FLOW, d after transplant), number of flowers per inflorescence (NF), pollen
702 viability (PV, %), fruit weight (FW, g), shape index (SI) and seeds per fruit (SxF) measured
703 on plants of the San Marzano cv (WT) and of 18 fruit variant lines. Mean values significantly
704 higher and lower than the WT for $P \leq 0.05$ after Student's *t* test are in bold and underlined
705 respectively.

Class of variation	Genetic symbol	FLOW	NF	PV	FW	SI	SxF
Wild type	WT	30.6	7.3	96.3	60.7	1.89	44.3
All pigments	*hp-1*	30.9	**11.0**	89.0	74.2	1.68	51.4
	hp-2	33.6	7.2	81.3	44.0	**2.16**	21.8
	pd	**36.0**	**11.1**	94.6	63.2	1.63	37.9
Carotenoids	*r*	27.7	8.9	95.7	57.7	1.70	44.5
	t	29.5	6.8	93.7	46.3	2.04	42.1
	at	33.0	**10.5**	96.5	29.3	1.44	21.2
	B	33.1	**9.9**	87.7	48.7	1.55	45.0
	B_mo$_B$	33.4	**9.1**	91.5	54.0	1.86	43.7
Chlorophyll	*gf*	30.3	**11.5**	91.0	68.0	1.85	53.2
Flavonoids	*y*	32.6	8.2	93.6	69.7	2.02	38.4
Ripening	*Nr*	29.2	**9.4**	95.6	45.3	1.65	63.5
	rin	29.1	**14.1**	92.5	34.1	1.91	30.2
	Gr	32.0	**9.7**	93.0	71.8	1.75	31.0
Double mutants	*r_y*	30.6	8.6	94.7	44.3	1.64	24.3
	gf_y	28.7	8.0	96.0	52.7	1.82	41.8
	gf_r	30.2	**8.7**	96.6	57.8	1.70	44.3
	gf_hp-2	**33.1**	6.2	94.8	44.0	1.88	35.8
	Aft_atv	30.5	**9.2**	75.6	66.8	**2.22**	40.4

706

707

708

Figure legends

Fig. 1. Representative fruits of the San Marzano cultivar (WT) and of 18 lines carrying mutations for fruit phenotype in the San Marzano background (line symbols are reported in Table 1).

Fig. 2. Distribution according to the first two principal components of the 21 lines studied (left) and of 16 clustered San Marzano lines (right) after GBS analysis at 1351 SNP markers (line symbols are reported in Table 1).

Fig. 3. Absolute variation in (**A**) plant height (ΔPH, cm), (**B**) leaf chlorophyll (ΔCHL, μg/cm^2) and (**C**) flavonoid content (ΔFLAV, μg/cm^2) of 18 fruit mutant lines in San Marzano background compared with the recurrent parent. Line symbols are reported in Table 1. Bars coloured in grey and black indicate means significantly lower and higher than San Marzano for $P \leq 0.05$ after Student's *t* test respectively.

Fig. 4. Distribution according to the first two principal components (PC) of the 19 San Marzano lines studied according to the chromameter parameters a, b and L (line symbols are reported in Table 1).

Fig. 5. Variation in (A) soluble solids content (ΔBrix), (B) shelf-life after 20 d storage (ΔSL, % of initial FW) and (C) days to wrinkling (ΔWRINK) of 18 fruit mutant lines in San Marzano background compared with the recurrent parent. Line symbols are reported in Table 1. Bars coloured in grey and black indicate means significantly lower and higher than San Marzano for P≤0.05 after Student's t test respectively.

Supplementary Tables and Figures

Supplementary Table S1. List of the mutations used in the lines with a San Marzano genetic background adopted in the study divided according to the class of variation, extended names, genetic symbols, first descriptor of the variant and details on their molecular characterization.

Supplementary Table S2. Distribution of SNPs in the tomato chromosomes according to the filtering strategy in the comparisons of the San Marzano (SM) recurrent parent, the introgression lines with SM background and the Ailsa Craig (AC) background occurring in several donor parents.

Supplementary Table S3. F values and degree of significance in the factorial analysis for plant height (PH), chlorophyll (CHL) and flavonoid (FLAV) content, flowering date (FLOW), number of flowers per inflorescence (NF), pollen viability (PV), fruit weight (FW), shape index (SI), total soluble solids (Brix), number of seeds per fruit (SxF), days to fruit wrinkling (WRINK), weight decrement in 20 days of shelf-life (SL) and for the colorimetric parameters a, b and L. *, ** and *** indicate significant F values for $P \leq 0.05$, 0.01 and 0.001 respectively.

Supplementary Table S4. Mean values within levels of the main factor “Year” (1 and 2) for traits showing significant Genotype*Year interaction. Values are reported for plant height (PH, cm), chlorophyll (CHL, µg/cm2) and flavonoid (FLAV, µg/cm2) content, flowering date (FLOW, days from transplant), number of flowers per inflorescence (NF), fruit weight (FW, g), shape index (SI), fruit weight remaining after 20 days of shelf-life (SL, %), days from harvesting to first fruit wrinkling (WRINK, d) and for the colorimetric parameters a, b and L. Mean values significantly higher and lower than in the WT for $P \leq 0.05$ after Student's *t* test are written in bold and underlined respectively.

Supplementary Table S5. Chromosome, physical position and size of the 22 haplotypes detected and introgression lines harbouring each haplotype.

Supplementary Figure S1. Linear regression between the number of backcrosses (BCs) carried out for each line and the genetic distance from the San Marzano reference recurrent parent estimated by GBS analysis.

Supplementary Figure S2. Introgressions from the donor parent estimated in the 18 studied lines after GBS analysis at 1351 SNP markers. Short vertical blue lines indicate SNPs polymorphic compared with the San Marzano reference. Black arrowheads indicate the position of the introgressed mutations (line symbols are listed and explained in Table 1 and Supplementary Table S1).

Supplementary Figure S3. Polymorphisms between the 17 studied introgression lines (*Aft_atv* has been removed), the San Marzano (SM) recurrent background and Ailsa Craig (AC), the most recurrent donor parent background. Heatmaps are constructed using 539 SNPs polymorphic amongst these 19 genotypes (A) or only the 129 SNPs polymorphic between SM and AC. The blue color indicate presence of the SM allele, yellow of the AC allele, red of an allele from other genotypes and black a missing value. Line symbols are described in Table 1.

Supplementary Figure S4. Neighbour joining tree constructed on the basis of haplotypes different from San Marzano (SM) detected between the 17 studied introgression lines (*Aft_atv* has been removed). Line symbols are described in Table 1.

Figure1

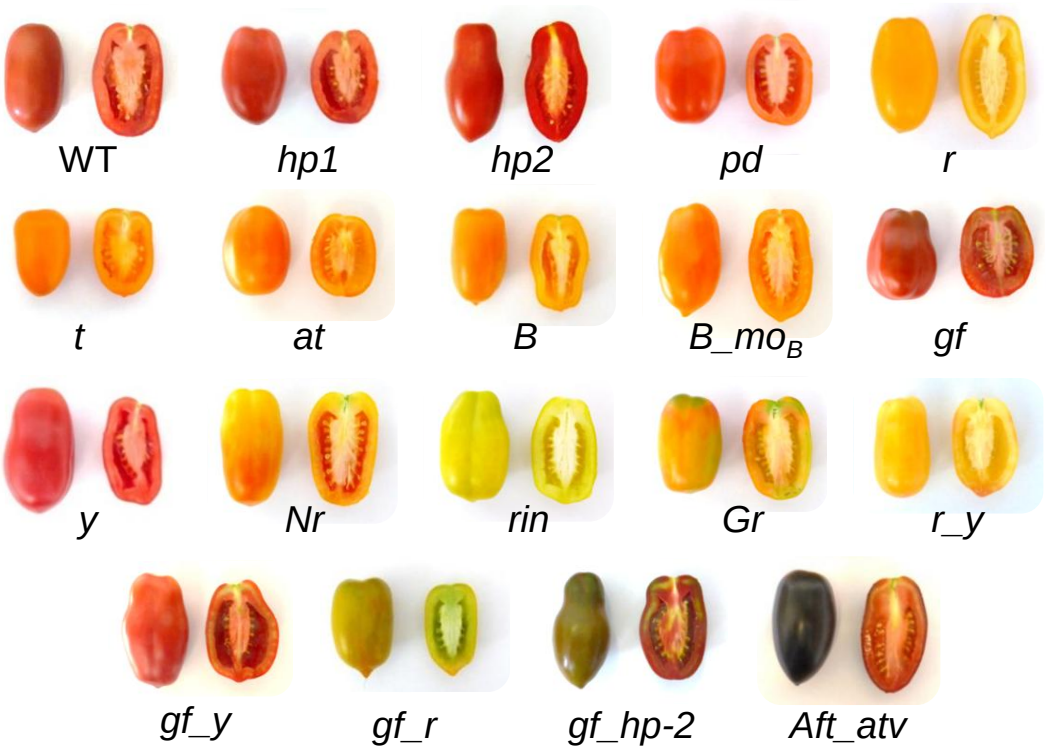


Figure2

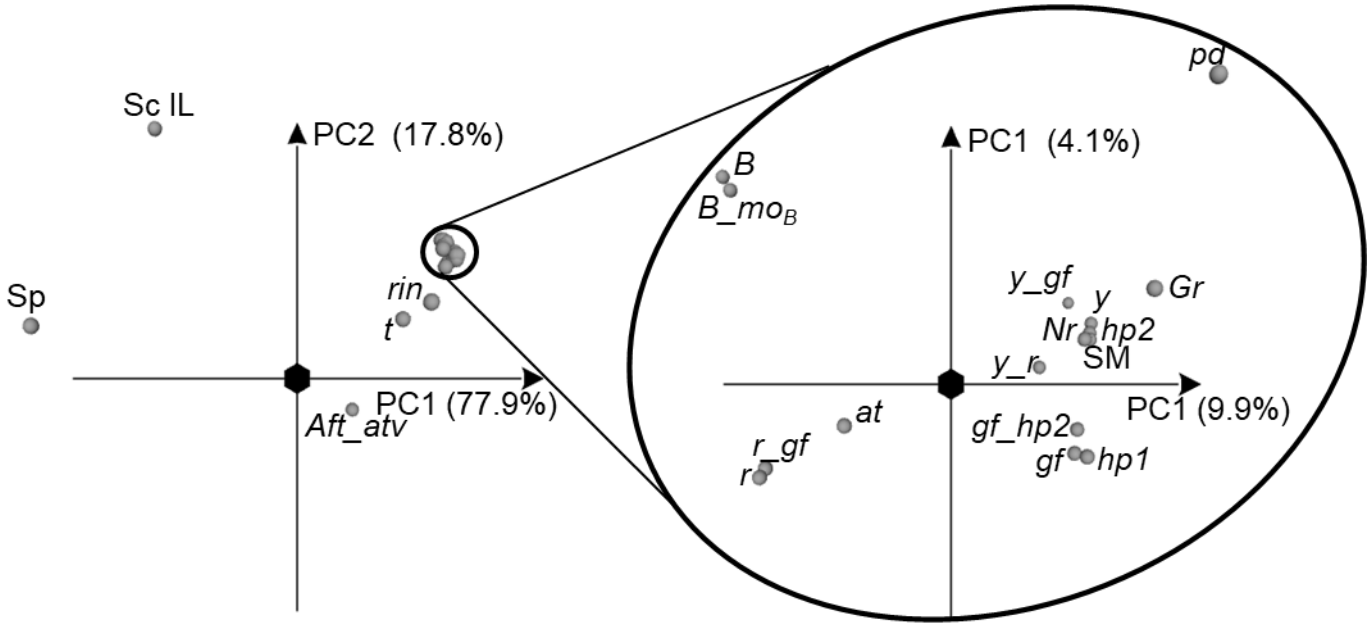

Sc IL
PC2 (17.8%)
Sp
rin
t
Aft_atv
PC1 (77.9%)
pd
B
B_mo_B
PC1 (4.1%)
y_gf
y
Gr
Nr
hp2
SM
y_r
at
PC1 (9.9%)
gf_hp2
r_gf
r
gf
hp1

Figure3

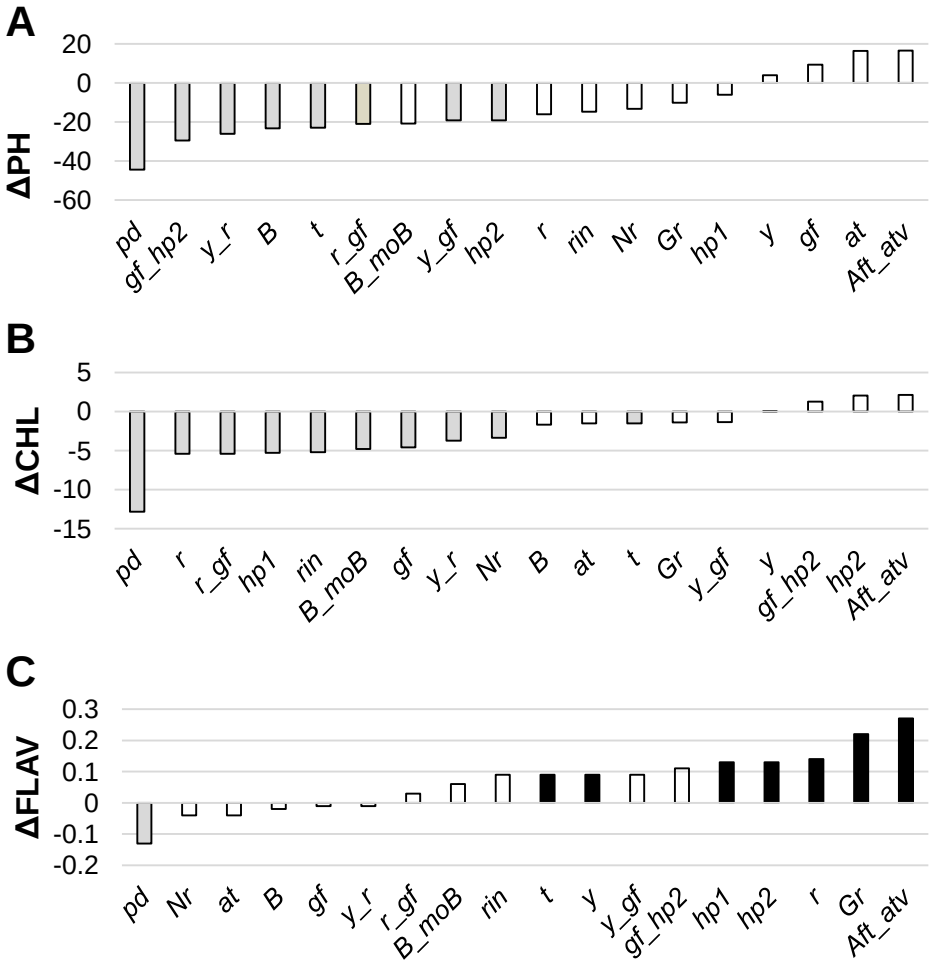

A
ΔPH
20
0
-20
-40
-60
pd
gf_hp2
y_r
B
t
r_gf
B_moB
y_gf
hp2
r
rin
Nr
Gr
hp1
y
gf
at
Aft_atv
B
ΔCHL
5
0
-5
-10
-15
pd
r
r_gf
hp1
rin
B_moB
gf
y_r
Nr
B
at
t
Gr
y_gf
y
gf_hp2
hp2
Aft_atv
C
ΔFLAV
0.3
0.2
0.1
0
-0.1
-0.2
pd
Nr
at
B
gf
y_r
r_gf
B_moB
rin
t
y
y_gf
gf_hp2
hp1
hp2
r
Gr
Aft_atv

Figure4

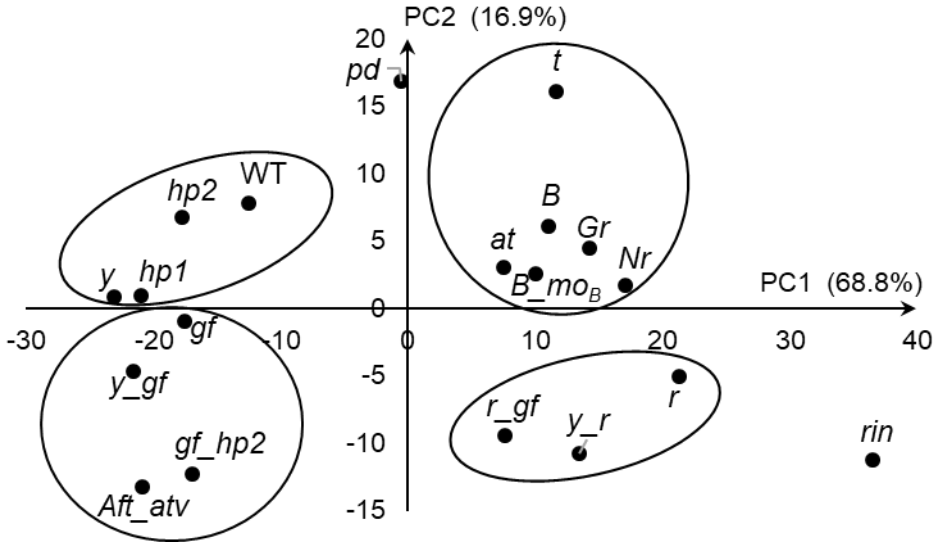
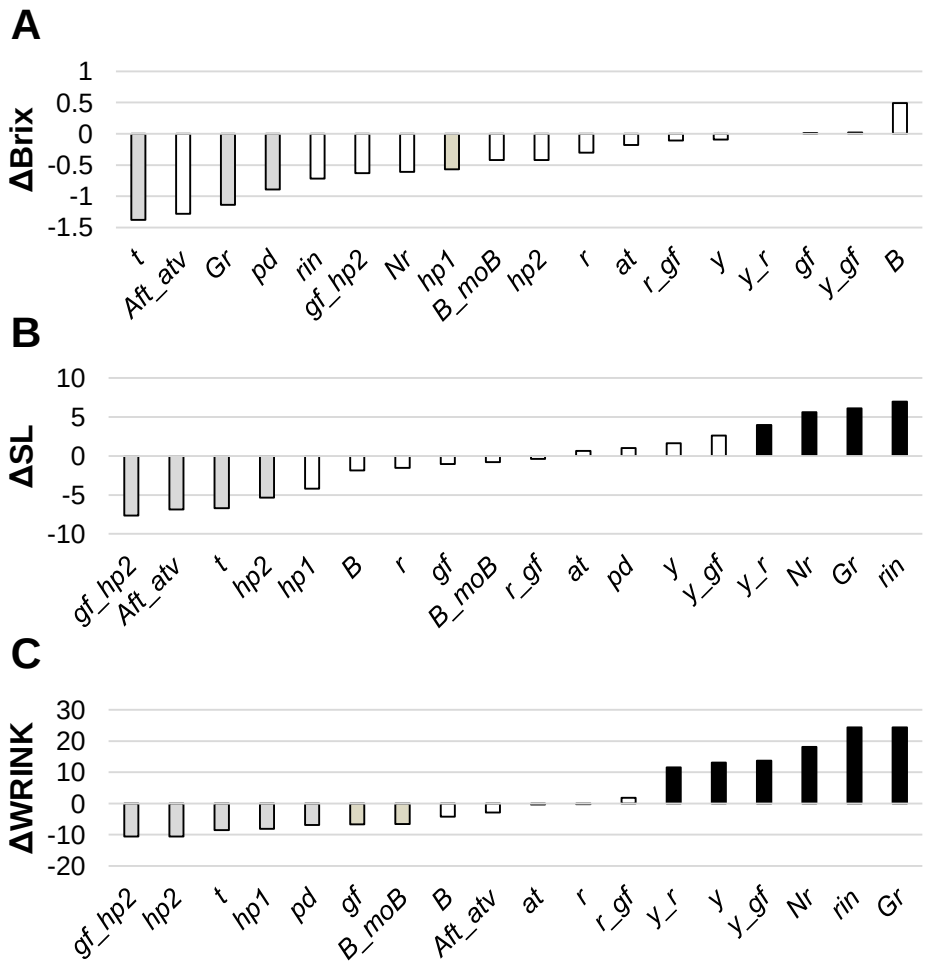

PC2 (16.9%)
PC1 (68.8%)
20
15
10
5
0
-5
-10
-15
-30
-20
-10
0
10
20
30
40
pd
t
B
Gr
at
Nr
B_mo_B
hp2
WT
y
hp1
gf
y_gf
gf_hp2
Aft_atv
r_gf
y_r
r
rin

Figure5


A
ΔBrix
1
0.5
0
-0.5
-1
-1.5
t
Aft_atv
Gr
pd
rin
gf_hp2
Nr
hp1
B_moB
hp2
r
at
r_gf
y
y_r
gf
y_gf
B
B
ΔSL
10
5
0
-5
-10
gf_hp2
Aft_atv
t
hp2
hp1
B
r
gf
B_moB
r_gf
at
pd
y
y_gf
y_r
Nr
Gr
rin
C
ΔWRINK
30
20
10
0
-10
-20
gf_hp2
hp2
t
hp1
pd
gf
B_moB
B
Aft_atv
at
r
r_gf
y_r
y
y_gf
Nr
rin
Gr

Supplementary Tables

Click here to download Supplementary Material: Supplementary Tables.docx

Supplementary Figures

Click here to download Supplementary Material: Supplementary Figures.docx