UNIVERSITA' DEGLI STUDI DELLA TUSCIA

Dipartimento Di Scienze Agrarie e Forestali (DAFNE)

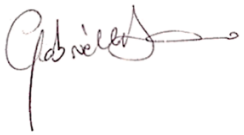
CORSO DI DOTTORATO DI RICERCA IN

SCIENZE DELLE PRODUZIONI VEGETALI E ANIMALI - CICLO: XXXII

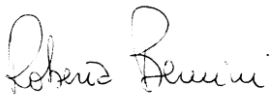
CHARACTERIZATION OF A COLLECTION OF TOMATO MUTANT LINES IN THE SAN MARZANO BACKGROUND

(AGR/07 – Genetica agraria)

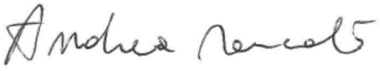
Tesi di dottorato di:
Dott. Gabriella Dono



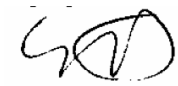
Coordinatore del corso:
Prof.ssa Roberta Bernini



Tutore:
Prof. Andrea Mazzucato



Co-tutori:
Dott. Gianfranco Diretto e Dott. Josè Luis Rambla



ANNO (2019/2020)

"...Apri la mente a quel ch'io ti paleso
e fermalvi entro; ché non fa scienza,
sanza lo ritenere, avere inteso..."

(Dante Alighieri)
Divina Commedia 1265 - 1321
Beatrice: V, 40-42

Chapter 1.

General Introduction

1.1. Traditional italian tomato varieties

First introduced into Europe from Central and Southern America by Spanish explorers at the beginning of the sixteenth century, the tomato (*Solanum lycopersicum L.*) was initially considered a botanical curiosity, used for ornamental purposes, and its potential as a foodstuff was hindered by the suspicion of the presence of poisonous compounds (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 (Rick 1976).

In Europe, the tomato found success mainly in the Mediterranean countries for their suitable agro-climatic conditions, including Italy and Spain, where it found a secondary centre for diversification, after its probable domestication in Mexico, and acquired commercial importance (Bailey et al., 1960).

Although tomato production is almost exclusively based on modern hybrid cultivars, there are still several traditional local cultivars renowned for their high quality. Indeed, after its introduction in the Old World, a wide range of local cultivars was developed; 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).

Among the traditional Italian varieties, San Marzano, Sorrento and Corbarino, which originated from the Campania region, are among the most popular traditional varieties (García-Martínez et al., 2013).

These landraces, although equipped with excellent organoleptic features, have over time lost part of their economic importance as they were replaced with modern hybrids, more productive and equipped with resistances to the most common pathogens. However, the improvement of local varieties of different plants, including tomato, was

recovered following the current trends in the development of the territory and all its components and the genetic re-evaluation of typical products of certain areas of our countries (Mazzucato et al., 2008).

1.2. The “San Marzano” traditional variety

San Marzano (SM) is one of the most popular Italian tomato landraces, used with the dualpurpose of fresh consumption and processing (Figure 1.1). 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). Currently, fruits of a few traditional cultivars are premium products in the domestic market. For instance, in 2011, for the production of San Marzano DOP canned tomatoes, fruits were purchased by processing factories at a price that was nearly five times higher than that of other varieties (http://www.consorziopomodorosanmarzanodop.it/).

Figure 1.1. Red fruits of the San Marzano variety

1.3. Breeding in tomato

It is likely that tomato breeding in Europe started just after its introduction. In tomato, the productivity of the crop has always been a major goal, together with the establishment of cultivars genetically resistant to pathogens and parasites, to the size of the berry and to post-harvest quality. The development of experimental and breeding plant populations is a prerequisite to genetic and functional studies in plant biology.

1.3.1. Development of populations of near isogenic lines (NILs)

The tomato is one of the major vegetable crops worldwide 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, 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 and Ritchie, 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).

1.3.2. Development of a repertoire of tomato fruit genetic variants in the San Marzano genetic background

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 (mediumelongate) and San Marzano with elongate fruit. San Marzano (SM) is one of the most popular Italian tomato landraces, used with the dualpurpose of fresh consumption and processing.

1.4. Description of the mutations introgressed into the San Marzano collection

Tomato fruit variants, backcrossed into the traditional San Marzano background, included 13 single mutant lines affected in different aspects of fruit physiology and five double mutants; respectively three mutants affecting the content of all pigments, *high pigment-1* (*hp-1*), *high pigment-2* (*hp-2*), *pigment diluter* (*pd*), five for carotenoids, *yellow flesh* (*r*), *tangerine* (*t*), *apricot* (*at*), *Beta* and its modifier (*B*, *B_Mo$_b$*), a variant for chlorophyll, *green flesh* (*gf*), one for flavonoids, *colorless epidermis* (*y*), three cariants for delayed ripening, *Green ripe* (*Gr*), *Never ripe* (*Nr*), *ripening inhibitor* (*rin*) and five double mutant lines (*r_y*, *r_gf*, *y_r*, *y_gf*, *Aft_atv*) (Figure 1.2).

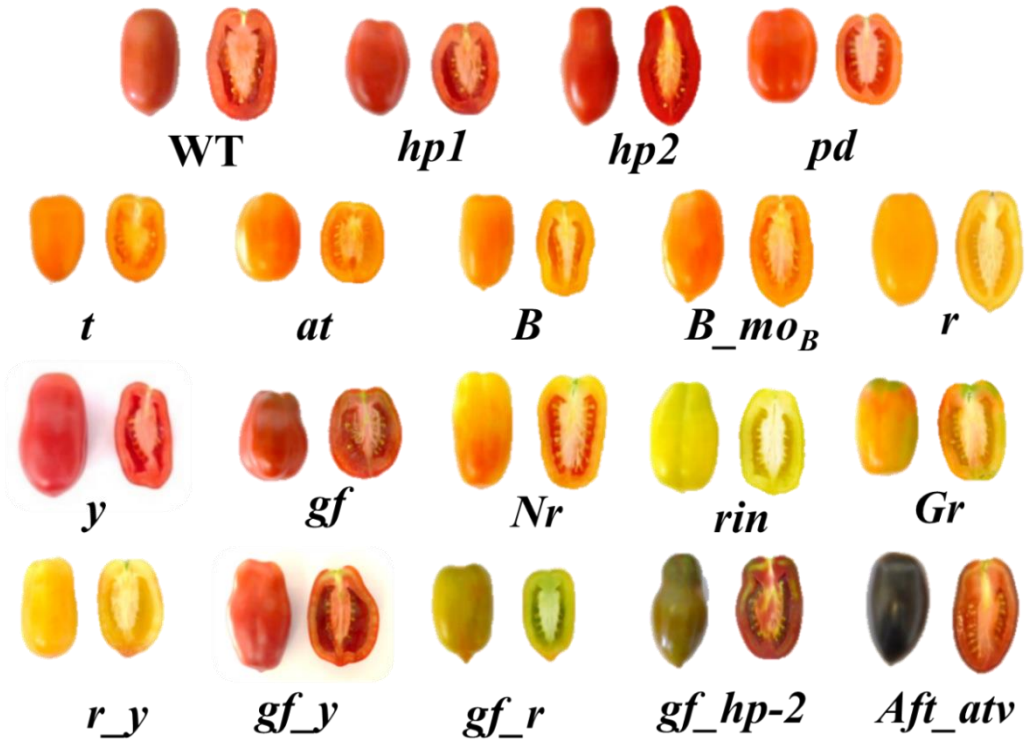


Figure 1.2. Representative fruits of the San Marzano cultivar (WT) and of 18 lines carrying mutations for fruit phenotype in the San Marzano background.

1.4.1. Mutations increasing the content of all pigments

Photomorphogenesis light is a critical environmental signal controlling many aspects of plant development and it is perceived by a series of photoreceptors covering a wide spectral range. Several light-hypersensitive mutants have been described in tomato (*Solanum lycopersicum*) over the last decades; in general, these mutants may be classified either as defective in photoreceptors, or altered in some elements of the light signal transduction chain (Chory 1993).

Among the latter, the monogenic recessive, nonallelic *high pigment* (*hp-1* and *hp-2*), mutants affect fairly specifically the responses mediated by phytochromes, they do not display any obvious phenotypes in darkness but are characterized by their exaggerated light responsiveness. Indeed, they display higher anthocyanin levels in their seedlings, shorter hypocotyls, high chlorophyll content in leaves and unripe fruits and more deeply pigmented mature fruits when compared with wild-type plants (Mustilli et al., 1999).

Because of their effect on fruit color, due to elevated levels of carotenoids (lycopene and carotenes) and flavonoids, a considerable interest has developed in introgressing these mutations into several commercial processing and fresh-market tomato varieties, that are currently marketed as Lycopene Rich Tomatoes (LRT) (Wann 1997; http://www.lycored.com/).

1.4.1.1. *high pigment-1*

The UV-damaged DNA binding protein (*DDB*) is an heterodimer with two subunits, *DDB1* and *DDB2* and it is an important factor involved in DNA repair and cell cycle regulation processes, as it has high affinity for a variety of DNA lesions including UV photoproducts (Fu et al. 2003; Ishibashi et al. 2003). The *DDB1* gene encodes two alternative high pigment mutant alleles, *hp-1* and *hp-1^{w}* mutant phenotypes. The *hp-1* mutant allele was discovered as a spontaneous variant in 1917 at the Campbell Soup Company farms (Riverton, NJ) (Reynard, 1956) and it was located in the centromeric region of the tomato chr2 (Lieberman et al., 2004). The *hp-1^{w}* mutant appeared among the progeny of a plant raised from ethyl methanesulfonate (EMS)-treated seeds of the genotype GT (Peters et al. 1989). A single A931T base transversion in the coding sequence of the *DDB1* gene in the *hp-1* mutant plants resulted in the substitution of the conserved asparagine at position 311 to a tyrosine residue. In the *hp-1^{w}* mutant, on the other hand, a single

G2392A transition was observed, with the substitution of the conserved glutamic acid at position 798 to a lysine residue (Lieberman et al., 2004).

1.4.1.2. *high pigment-2*

Previous studies have demonstrated that *hp-2*, *hp-2^{j}* and *hp-2^{dg}* represent different mutations in the gene encoding the tomato homologue of the *Arabidopsis thaliana* nuclear protein *DEETIOLATED1* (*DET1*), a negative regulator of photomorphogenesis (Mustilli *et al*., 1999; Levin *et al*., 2003).

The *hp-2* mutant was reported in the italian San Marzano variety (Soressi 1975) and it was mapped on the chr1 (Mustilli et al., 1999). The *hp-2^{j}* mutant was found among progeny of a T-DNA-transformed plant (cv. Moneymaker) (van Tuinen et al. 1997), and the *hp-2^{dg}* mutant appeared in plantings of the Manapal variety (Konsler 1973).

hp-2 mutants showed an A-to-T transversion directing an alternative splicing of intron 10, that leads to a nine-base deletion in exon 11, and results in a deletion of the first three amino acids encoded by this exon. In the more phenotypically extreme *hp-2^{j}* mutation, a C -to-T transition was found in exon 11, which gave rise to a substitution of a conserved proline for a serine residue in the C-terminal region of the DET1 protein (Mustilli et al. 1999). Sequence analysis of the *DET1* gene in *hp-2^{dg}* mutants revealed a single A -to- T base transversion in the second exon, resulting in a substitution of the conserved asparagine at position 34 by isoleucine (Levin et al. 2006).

1.4.1.3. DDB1 and DET interaction model

DDB1 and *DET1* are supposed to be involved in the same pathway, indeed previous studies already described the interaction of *DEB1* with the nonacetylated H2B histone tail (Benvenuto et al. 2002; Schroeder et al., 2002; Liu et al. 2004). Furthermore, both *DDB1* and *DDB2* were also previously found to interact with either histone acetyltransferase (HAT) proteins or HAT complexes (Brand et al. 2001; Datta et al. 2001; Martinez et al. 2001).

A functional model proposed that they regulate the expression of hundreds of genes via an interaction with chromatin; the DET1/DDB1 complex binds the histon tail repressing the transcription in the dark; light stimulates the *DDB1* interaction with the HAT protein,

resulting in the acetylation of the H2B histone tail, the release of the *DET*-complex and the upregulation of the transcription (Figure 1.3).

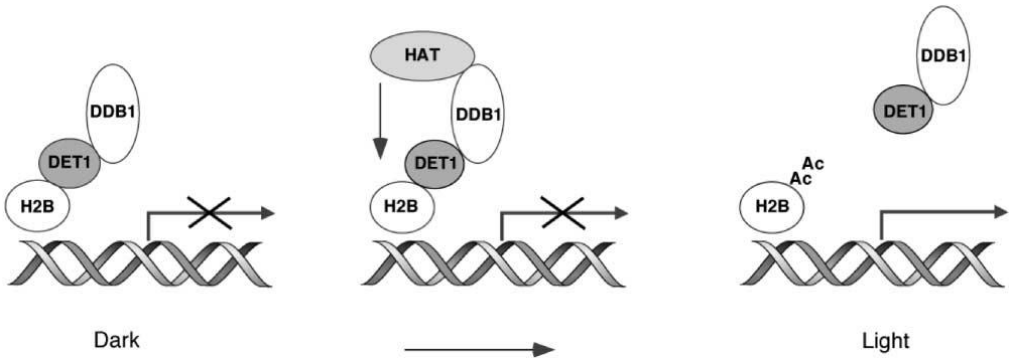


Figure 1.3. Interaction of DET1 and DDB1 with H2B and HAT, in a model proposed by Schroeder et al. (2002).

1.4.1.4. *Pigment diluter*, the opposite case

The gene underlying this mutation has not been cloned yet, remaining still unknown and paving the way for deeper studies; indeed this phenotype was described to influence the plant habitus, the chlorophyll content in leaves and fruit which appeared reduced and the fruits, which show reduced level of lycopene but high content of polyphenols (Minoggio et al., 2003). This interesting feature of berries drives to further insights.

1.4.2. Mutations affecting the content of carotenoids

The San Marzano collection includes five mutants for the content of carotenoids (*yellow flesh*, *apricot*, *tangerine*, *Beta* and its modifier) (Figure 1.4), which exhibit different shades of orange; this change in color is due to the accumulation of the intermediate and end-products of the carotene biosynthetic pathway (neurosporene, *ζ*-, *α*-,*γ*-, *δ*-carotene, prolycopene, and also the colorless polyenes, phytoene and phytofluene) at different levels, compared to the main pigments accumulated in the original tomato.

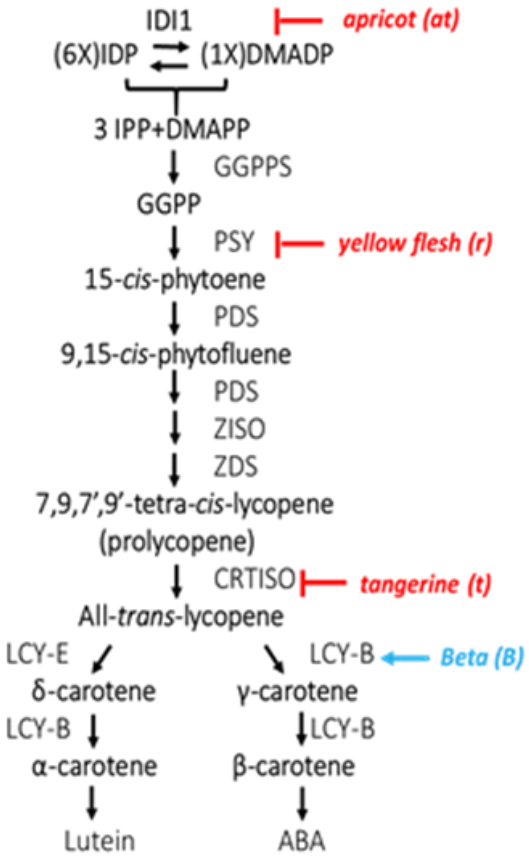


Figure 1.4. Carotenoid biosynthetic pathway in tomato. Mutants affecting the carotenoid biosynthesis are represented in italics. Blue color indicates positive regulation and red color indicates negative regulation of carotenogenesis. *IDI1*, isopentenyl diphosphate isomerase; *CRTISO*, carotene isomerase; *DMAPP*, dimethylallyl diphosphate; *GGPP*, geranylgeranyl diphosphate; *GGPPS* geranylgeranyl diphosphate synthase; *IPP*, isopentenyl diphosphate; *LCY-B*, lycopene cyclase; *LCY-E*, lycopene "-cyclase; *PDS*, phytoene desaturase; *PSY*, phytoene synthase; *ZDS*, ζ-carotene desaturase; *ZISO*, ζ-carotene isomerase (Yoo et al., 2017).

1.4.2.1. *yellow flesh*

In plants, carotenoids serve as precursors for the biosynthesis of hormones (ABA), signalling molecules and volatiles. Characterization of carotenoid mutants has greatly contributed to the knowledge of carotenoid biosynthesis pathway in plants.

The geranylgeranyl diphosphate (GGPP) is the ubiquitous isoprenoid precursor in the formation of carotenoids and the phytoene synthase (PSY-1) is the enzyme responsible for the conversion of GGPP to phytoene. PSY catalyzes a rate-limiting step in the carotenoid pathway, so the lack of expression of this gene eliminates carotenoid biosynthesis (Kachanovsky et al., 2012). Yellow-fruited tomatoes have been grown for centuries, indeed two phenotypes, *r* and *ry*, with a pale yellow fruit flesh and yellow fruit skin have been observed since the first half of the nineties (Jenkins and Mackinney, 1955; Hunt and Baker, 1980). The *r* phenotype also shows a paler lemon-yellow flower

corolla, while *ry* has a normal colored corolla (Darby et al., 1978). Both *r* and *ry* are mapped to the same location on chromosome 3, and thus considered as mutations of the same gene (Fray and Grierson, 1993). Furthermore PSY-2 is predominantly responsible for carotenoid formation in chloroplast-containing tissues, but it can still be detected in ripe tomato fruit (Fraser et al., 1999) and PSY-3 that was predicted to function in roots under stress conditions similarly to cereals.

1.4.2.2. *tangerine*

The carotenoid Isomerase gene (*CRTISO*) encodes a redox-type enzyme that converts *cis*-lycopene to *trans*-lycopene by isomerization reactions at 7, 9 and 70, 90 cis-bonds. The recessive mutation *tangerine* (*t*) (Tomes, 1952), was mapped to the long arm of chr10. A deletion of 348 bp was discovered in the promoter region of the t^{3183} allele, that abolishes its expression in fruits. Fruit of *t* are orange and accumulate prolycopene instead of all-*trans*-lycopene, which normally is synthesized in wild-type fruit (Isaacson et al., 2002). This deletion was discovered in six lines (LA0030, LA3183, LA3533, LA3682, Cispene, and Gold Minichal) by the INDEL marker of CRTISO (Yoo et al., 2017). Yoo et al. 2017 also found three new alleles that would be valuable for improving the prolycopene content in tomato breeding, in addition to the previously identified t^{3183} allele. The three new mutations were identified in the coding region, to be two nonsense mutations caused by one nucleotide insertion in the 2nd (LA3002) and 8th exons (LA3128), respectively, and a missense mutation (C to T) found in the 7th exon in LA0351 (Figure 1.5).

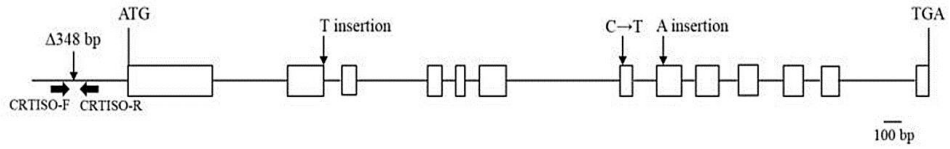


Figure 1.5. Gene structure of *CRTISO. CRTISO* mutation at the promoter region was previously identified (Isaacson et al., 2002) and three new mutations were identified in the coding region (Yoo et al., 2017).

1.4.2.3. *apricot*

Two isopentenyl diphosphate isomerase (*IDI*) enzymes exist in plants (Blanc et al., 1996; Ramos-Valdivia et al., 1997; Campbell et al., 1998; Sun et al., 1998; Cunningham and Gantt, 2000; Nakamura et al., 2001): one is active in the cytoplasm, including

mitochondria and peroxisomes, and the other is active in plastids. In this latter organelles *IDI1* sustains a balanced ratio of Isopentenyl diphosphate (IDP) to dimethylallyl diphosphate (DMADP), which are precursors in the biosynthetic pathways of plastidial isoprenoids, as monoterpenes, carotenoids and other compounds derived from geranylgeranyl diphosphate, such as chlorophylls, plastoquinone, tocopherols and gibberellins (Figure 1.6; reviewed by Gutensohn and Dudareva, 2013; Vranova et al., 2013). The gene encoding IDI1 was mapped on the chr4 (Eshed and Zamir, 1995) and identified as a candidate gene for the locus *Fcd1* (*Fruit Carotenoid Deficient*), which still includes four single recessive mutations (e1535, e0321, e0955 and e9292). They show the typical *fcd* phenotype, characterized by the pale orange color of their fruit, due to the reduction of the total fruit carotenoids. An effort of map-based cloning revealed the allele $fcd1^{at}$ (Jenkins and Mackinney, 1955), called *apricot* (*at*), which has a stop codon at position 234 (K234). Its phenotype is defined by a significant decrease in total fruit carotenoids, mainly because of an 80–95% decrease in the concentration of lycopene (Pankratov et al., 2016). Based on this evidence, *IDI* plays a significant role in plastidial isoprenoid biosynthesis.

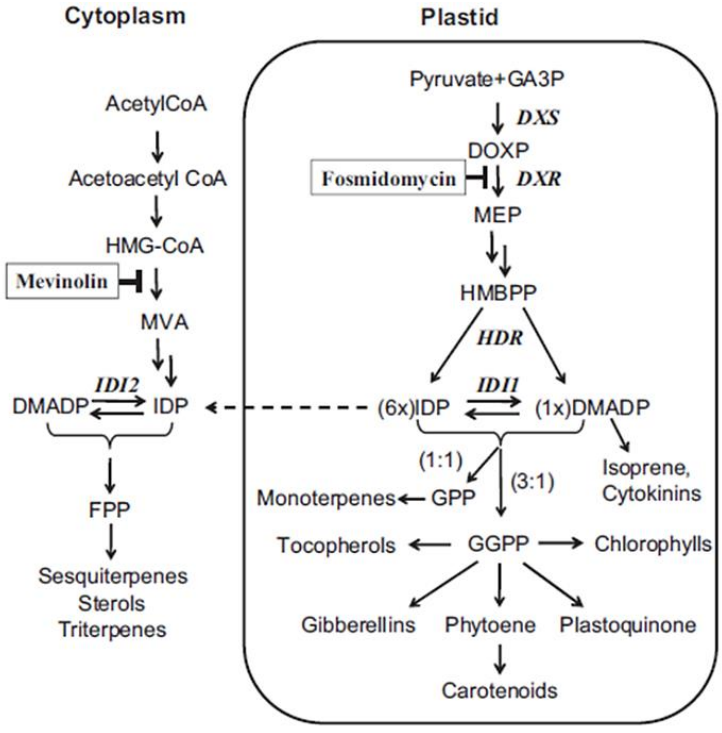


Figure 1.6. model for the role of IDI1 in the plastidial biosynthesis of isoprenoids. IDI1 sustains a balanced ratio of Isopentenyl diphosphate (IDP) to dimethylallyl diphosphate (DMADP), precursors in other biosynthetic pathways (Pankratov et al., 2016).

1.4.2.4. *Beta* and its modifier

Lycopene *β*-cyclase is the gene underlying *Beta* (*B*), a partially dominant, single-locus mutation that causes an orange color in the fully ripe fruit because of the accumulation of *β*-carotene at the expense of lycopene. In the wild type, β-carotene constitutes 5–10% of total fruit carotenoids, whereas in *Beta* it is synthesized *de novo* during tomato fruit development, ranging 45–50% and can exceed 90% when B is in combination with the *Beta-modifier* (Mo_B) gene (Ronen et al., 2000). The *B* allele originated in wild *Lycopersicon* species and was introduced into the cultivated tomato by crossing; B represents the wild-type form of the gene, which has gone through a loss of function mutation in the red tomato (Tomes et al., 1956). The locus *B* previously was found to be tightly linked on chr6 of the tomato linkage map, in proximity to the *self-pruning* (*sp*) locus (Lincoln et Porter, 1950).

1.4.3. Mutations affecting chlorophyll catabolism

Fleshy fruits predominantly accumulate carotenoids, anthocyanins, and flavonoids, and the *de novo* synthesis of these compounds at the onset of ripening is preceded by, or occurs concomitantly with, the degradation of chlorophyll (Figure 1.7; Seymour et al.,1993); However, little is known about the role of chlorophyll degradation and the regulation of this pathway during the ripening process, nor the overall contribution that it makes to fruit quality.

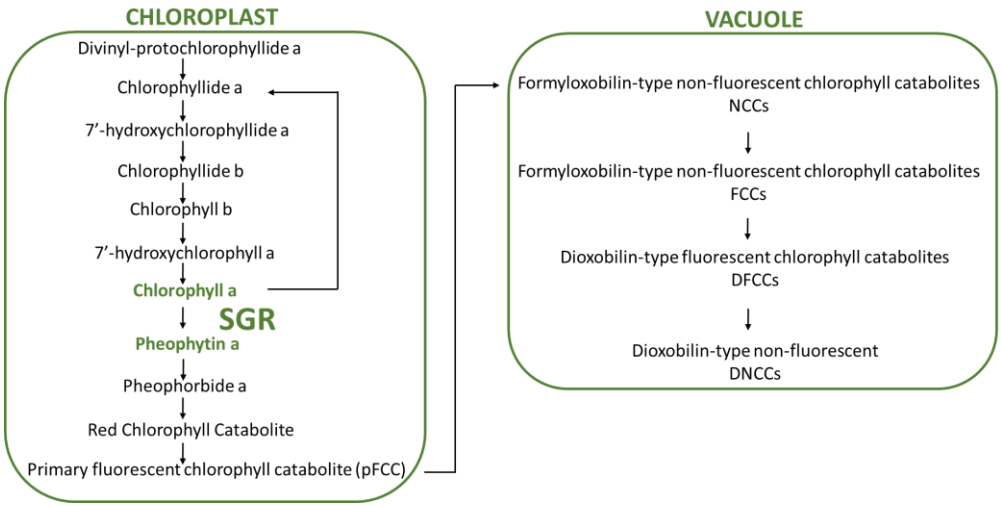


Figure 1.7. chlorophyll catabolism pathway in tomato. STAY-GREEN gene (SGR) underlying the *green flesh* mutation is indicated.

1.4.3.1. *green flesh*

Stay-green mutants have been identified in several plant species and classified on the basis of their chlorophyll retention and general senescence phenotypes; In particular, staygreen phenotypes have been divided into five classes (A, B, C, D, E) based on their physiological characteristics. A and B are functional, C, D and E are defined as cosmetic *stay-greens* (Thomas and Howarth, 2000). In this contest, mutations of the expression of *sgr* resulted in type C cosmetic staygreen phenotypes. Fruit of the *green-flesh* (*gf*) mutant of tomato ripen to a muddy brown color due to the accumulation of lycopene coupled with a lack of chlorophyll degradation (Kerr, 1956). The lack of chlorophyll degradation in *gf* is not restricted to fruits because dark- and nutrient-induced chlorophyll loss in leaves is also compromised in the mutant (Akhtar et al.,1999). The *gf* locus has previously been mapped on the long arm of chr8 (Tanksley et al.,1992), targeted to the chloroplast and the mutation results in the single amino acid substitutions R143S. Barry and Pandey (2009) analyzed a number of tomato heirloom varieties which displayed the staygreen phenotype and discovered more mutant alleles at the *gf1 locus*. These include additional single base substitutions and single base insertion and deletions of different sizes Also, these mutations affect different parts of the protein but they all result in the same phenotype (Figure 1.8).

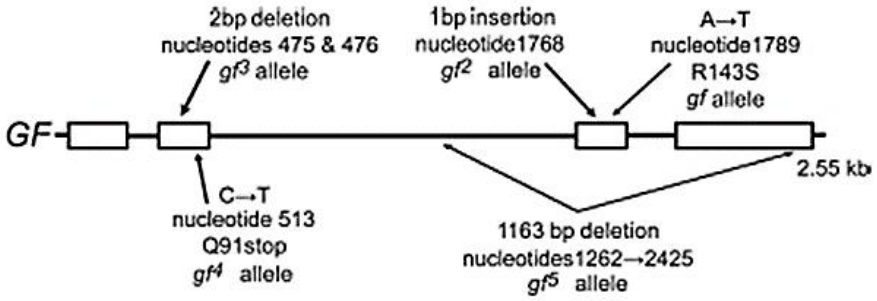


Figure 1.8. Genomic structure of GF. The nucleotide positions and changes that give rise to the five gf mutant alleles are shown (Barry and Pandey 2009).

1.4.4. Mutations affecting the ripening process

Tomato ripening is a highly coordinated developmental process that coincides with seed maturation, at the same time other processes occur, such ad fruit softening, accumulation of pigments, sugars, acids, and volatile compounds that increase attraction to animals (Klee and Giovannoni 2011). To achieve full ripening, climacteric

fruits, such as tomato require synthesis, perception and signal transduction of the plant hormone ethylene. Altered ethylene responsiveness in plant tissues influences development and can compromise the plants ability to respond to environmental stimuli (Barry and Giovannoni 2006). Although mutations affecting the ripening process can negatively alter the organoleptic quality of fruits, some of them, such as *rin*, *nor* and *alc* have been already used for long-shelf life breeding programs with the heterosis for yield (Figure 1.8).

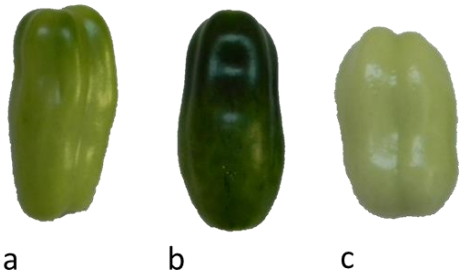


Figure 1.8. immature fruits of (a) San Marzano, (b) *high pigment-2* and (c) *pigment diluter.*

1.4.4.1. *Never ripe*

The ripening-impaired tomato mutant *Never-ripe* (*Nr*) is insensitive to the plant hormone ethylene. The gene that cosegregates with the *Nr* locus encodes a protein with homology to the Arabidopsis ethylene receptor ETR1, *LeETR1*. A single amino acid change in the sensor domain confers ethylene insensitivity.

1.4.4.2. *ripening inhibitor*

The strongest evidence for non-ethylene-mediated ripening control comes from the analysis of gene expression in fruit of the *ripening-inhibitor* (*rin*) that fail to produce autocatalytic ethylene, to ripen, and to ripen in response to exogenous ethylene, yet display signs of ethylene sensitivity and signalling, including the induction of some ethylene-regulated genes (Tigchelaar et al., 1978; Yen et al., 1995). The *rin* locus was mapped on the short arm of chr5 and the first clue was that two sequences were fused into a chimeric gene in the *rin* mutant as a result of a genome deletion. Both genes are members of the MADS-box family of transcriptional regulators, one (LeMADS-RIN)

regulates ripening while the other (LeMADS-MC) is responsible for the large sepal (macrocalyx) phenotype, thus associated with the *rin* mutation (Vrebalov et al., 2002).

1.4.4.3. *Green ripe*

Inhibition of fruit ripening in the *Green-ripe* (*Gr*) mutants of tomato is the result of ethylene insensitivity showing responses associated with floral senescence, abscission, and root elongation too. However, ethylene-mediated inhibition of hypocotyl elongation and petiole epinasty remain normal, suggesting that this locus affects a subset of ethylene responses in tomato, with the strongest phenotypes observed in fruit (Barry et al., 2005). The positional cloning revealed the presence of a 334-bp deletion that resides in the 5-flanking region of a gene encoding an evolutionary conserved protein of unknown function located on the long arm of chr1, that is predicted to be membrane localized.

The deletion causes ectopic expression of *GR*, a phenomenon consistent with a dominant gain of function mutation, that disrupts ethylene signalling, causing the inhibition of fruit ripening (Barry et al., 2006).

1.4.5. Mutations affecting cuticle, synthesis and accumulation of pigments

Most aerial plant surfaces are covered with a cuticle, a heterogeneous layer composed mainly of cutin and wax lipids. Despite the cuticle's key role in organ development and in protecting against stresses, very little is known about the regulation of the metabolic pathways. Flavonoids, often embedded in the cuticle, have been suggested to impact the characteristics of this structure and to provide protection against radiation and pathogens. Furthermore, they are an integral part of the human diet and are likely responsible for the observed beneficial effects of a fruit-rich diet (Adato et al 2009). Two mutations affecting the cuticle composition in tomato fruits are deeply analyzed in this work, and here below described.

1.4.5.1. *colourless fruit epidermis*

The *colorless fruit epidermis* (*y*) fruit mutant was originally described as a monogenic recessive variant leading to the formation of a colorless fruit peel (Lindstrom 1925). The mutation, mapped on the short arm of Chr1 (Rick and Butler 1956), involves the *SlMYB12* transcription factor, causing the lack of naringenin chalcone, one of the major flavonoids in tomato fruit peel, which gives the yellow color and has been proposed to influence the characteristics and function of the cuticular layer (Adato et al., 2009; Ballester et a., 2010), by playing an important role in controlling water transport across the polymer matrix (Luque et al., 1995).

In contrast to the wild-type, *y* mutant fruit does not accumulate the yellow pigment that typically suffuses throughout the cell walls of wt fruit epidermis. The *y* fruit exhibits a pink and less glossy appearance at the late orange and red stages of fruit development.

1.4.5.2. Combinations affecting anthocyanin sinthesis, *Aft_atv*

Anthocyanin fruit (*Aft*) from *L. chilense*, *Aubergine* (*Abg*) from *S. lycopersicoides*, and *atroviolacium* (*atv*) from *L. cheesmanii* cause anthocyanin expression in tomato fruit (Giorgiev 1972).

The dominant gene *Aft* triggers anthocyanin accumulation in immature green fruit upon stimulation by high light, causing their subsequent production continuously throughout development (Mes et al., 2008). The *Aft* gene identity has been recently revealed, as analternative Splicing in the *Anthocyanin Fruit* Gene Encoding an R2R3 MYB Transcription Factor Affects Anthocyanin Biosynthesis in Tomato Fruits (Colanero et al., 2019).

A recessive gene, *atv* (atroviolaceum), has been shown to influence anthocyanin pigmentation in the entire tomato plant, particularly in the vegetative tissues (Mes et al., 2008). the atroviolacea Gene has been located to chromosome 7 (Rick et al., 1968) and encodes an R3-MYB Protein Repressing Anthocyanin Synthesis in Tomato Plants (Colanero et al 2018).

Double mutant combinations with both *Aft* and *atv* alleles have been obtained and they can be distinguished by the presence of intensely pigmented fruits for the higher production of anthocyanins in the peel (Mes et al., 2008).

1.5. The role of the tomato secondary metabolites and their impact on the human health

Tomato is known as a plant that contains many secondary metabolites. Profiling of such metabolites in tomato plants, particularly in the fruits, has been performed by analysis of carotenoids (Moco et al., 2007), phenolic compounds (Slimestad and Verheul, 2009), volatile organic compounds (Buttery et al., 1987), and alkaloids (Friedman and Levin, 1998).

Tomato ripening involves a number of physiological processes that include the visible breakdown of chlorophyll and build-up of carotenoids, with massive accumulation of antioxidant components such as lycopene, α- and β-carotene, which possess provitamin A activity (Bendich & Olson 1989), within the plastids (Laval-Martin et al., 1975). *Trans*-lycopene represents the most abundant lycopene isomer in tomato, but also the *cis*-lycopene component has demostrated to be better absorbed in the intestine than *trans* isomers (Takeoka et al., 2001).

Besides carotenoids, the contents of other important antioxidant compounds, such as ascorbic acid, *α*-tocopherol and phenolics, as well as glycoalkaloids, *α*-tomatine and dehydrotomatine, vary during ripening, thus varying the nutritional value and the antioxidant activity of the fruit (Friedman et al., 1997; Martínez-Valverde et al., 2002).

These compounds have important functional roles, indeed they can act as mediators in the interaction of the plant with its environment, such as plant–insect, plant–microorganism and plant–plant interactions, as part of the plant's defence system, as attractors of pollinators for reproductive aims (Dixon et al., 2001), but they have also a potential in preventing diseases and promoting health in animals and humans.

Secondary metabolism is thus an interesting target for plant breeding. Molecular breeding is in this respect a promising approach, to decrease or more often increase the quantity of a certain compound or group of compounds. Also, there is interest in the production of novel compounds not yet produced in nature by plants, or specifically not expressed in tomato, such as the expression of genes belonging to the flavonoid and anthocyanin pathway in tomato (Verpoorte et al., 2000).

1.5.1. Carotenoids

Plant carotenoids are a wide group of 40-carbon lipophylic isoprenoid pigments with chemical properties essential for all the photosynthetic organisms. They are divided in two groups: carotenes, such as *β*-carotene, that are linear hydrocarbons and xanthophylls (derivatives of carotenes possessing in their molecule one or more oxigen containing groups – hydroxylic, epoxidic, carbonylic), such as lutein, violaxhantin, neoxhantin and zeaxhantin (Chalukova et al., 1991).

They are synthetized in the chloroplasts and chromoplasts of the plant cell in the normal isoprenoid pathway by a serial of sequential conversions (Figure 1.9); in the process of fruit ripening, chlorophyll is usually degraded and an active synthesis of carotenoids is accomplished (Chalukova 1991).

Carotenoids play vital roles in the light-harvesting photosinthesis process of the chloroplasts and they protect the photosynthetic apparatus against harmful reactive oxygen species produced by overexcitation of chlorophyll. They are also indispensable in chromoplasts, providing flowers and fruits with distinctive colors to attract animals for pollination and seeds dispersion (Hirschberg et al., 2001).

Carotenoids in plants are also precursors for the synthesis of the hormon abscisic acid (ABA) which originates from 9-cis-violaxanthin and –neoxanthin (Schwartz et al., 1997) and other signaling molecules and volatiles.

Dietary carotenoids are essential for the health of humans as they are unable to synthesize carotenoids *de novo* (Nisar et al., 2015). For example, *β*-carotene, the precursor of Vitamin A, prevents damages to the eyes (Fraser et al., 2004); similarly, lycopene protects against chronic, cardiovascular diseases and decreases the risk of cancers (Ford et al., 2013).

The properties of carotenoids as natural pigments have been industrially exploited for a long time; for instance some carotenoids are added as colorants for human food, they also have an important industrial value as aromas (Giuliano et al., 2003) and are widely used in the cosmetic and pharmaceutical industries (Fraser and Bramley 2006).

Therefore, improving individual carotenoid levels is an important trait in the breeding of many crops.

Tomato fruit are a big source of lycopene, approximately 90% of total carotenoids, the remainder being composed of *β*-carotene and traces of lutein. Through metabolic

engineering and other strategies, the large pool of lycopene could be converted into high value-added down- or up-stream compounds (Dharmapuri et al., 2002).

Besides the carotenoid mutants, specifically *r*, *at*, *t* and *B* of this collection, the total carotene content is determined by the biosynthetic level of their precursors, which is also affected by other mutated alleles of the collection. This is the case of *Never ripe, ripening inhibitor* and *Green ripe*, which mainly influence steps in fruit ripening, but can also disturb the biosynthesis of carotenoids (Palmieri et al., 1976; Tigchelaar et al., 1978). The quantity of carotenes in tomato fruits and the resulting color intensity are enhanced in the presence of *high pigment* genes (Thompson et al., 1967).

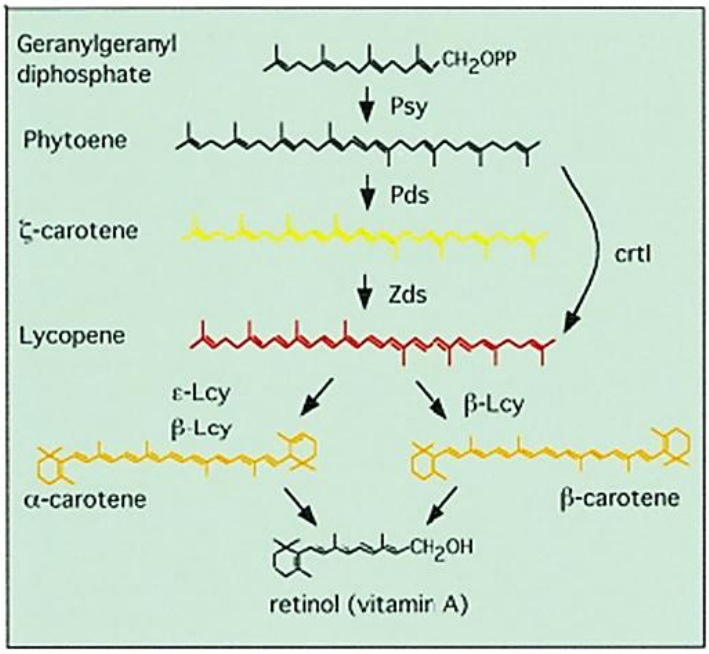


Figure 1.9. Carotenoid biosynthetic pathway in plants. Reactions leading to the vitamin A synthesis (Rosati et al., 2000).

1.5.2. Chlorophylls

Plants depend upon radiant energy to carry out photosynthesis and other physiological processes. The green plant has been called the converter of solar energy. In the presence of sunlight, plants synthesize complex organic compounds such as sugars, fats, proteins, etc., from simple inorganic compounds such as water, carbon dioxide, minerals, salts, etc. The leaf of a plant is the primary photosynthesizing organ with photosynthesis occurring in the chloroplasts where the chlorophyll pigment is located. The chloroplasts are generally more abundant towards the upper side of the leaf in the palisade cells and hence account for the darker appearance of the upper leaf surface compared with the

lower lighter surface. The pigments generally found in chloroplasts are chlorophyll (65%), carotenes (6%), and xanthophylls (29%) although the percentage distribution is highly variable. Chlorophyll a and chlorophyll *b* are most frequent in higher plants (David et al., 1965). Chlorophyll, the most abundant pigment on earth, is a key component of photosynthesis required for the absorption of sunlight (Hörtensteiner et al., 2012). The San Marzano collection includes the *green flesh* mutant, already described before, in which chlorophyll contents are mantained in the fruit during ripening, as well as the three delayed ripening mutants, *Never ripe*, *Green ripe* and *ripening inhibitor*, enriched in chlorophyll due to the blockage in the ripening process.

1.5.3. Tocochromanols

Tocochromanols are a group of amphipathic, lipid-soluble organic molecules composed of a polar moiety derived from tyrosine and a hydrophobic polyprenyl side chain originating from the isoprenoid pathway. Tocochromanols with a phytyl-derived side chain are termed tocopherols, whereas those with a geranylgeranyl-derived side chain are termed tocotrienols. The four different forms of tocopherols and tocotrienols (α, β, γ and δ) differ by the degree of methylation of the polar moiety. All tocochromanols are potent lipid-soluble antioxidants and are essential dietary nutrients for mammals as vitamin E (Schneider et al., 2005).

Only photosynthetic organisms such as plants, algae and some cyanobacteria synthesize tocochromanols. Tocopherols, considered more widespread than tochotrienols, have been detected in leaves, seeds, roots, tuber, fruits, stems, hypocotyls and cotyledons of higher plants; their composition is very heterogeneous, and α-tocopherol is in general the predominant form (Horvath et al., 2006).

Vitamin E content and composition in tomato fruits, associated with its beneficial roles and bio-fortification, has been deeply discussed (Raiola et al., 2015).

Interestingly, it has been observed that the advantageous properties associated with the consumption of tomatoes are related to the synergistic properties of the tomato molecules, in particular lycopene and α-tocopherol, which have been shown to inhibit HL-60 (human promyelocytic leukemia cells) cell differentiation, low-density lipoprotein (LDL) oxidation, and prostate carcinoma cell proliferation (Zanfini et al., 2010).

Combined efforts have pointed at increasing the nutrient composition of plants to enhance human nutrition and health, In order to develop foods with increased levels of

vitamin E, and in the last decade new findings of the biosynthetic pathway steps have been established (Grusak et al., 1999; Mène-Saffrané, 2010).

1.5.4. Quinones

Photosynthesis inevitably produces toxic molecules derived from oxygen. Indeed, molecular oxygen can interact with the photosynthetic electron transport chain, leading to the formation of reduced forms of oxygen; chloroplasts contain a variety of antioxidant mechanisms including soluble and lipophilic low molecular weight antioxidants, as carotenoids, considered to be the first line of defense, ascorbate and gluthatione, tocochromanols, but also the less known detoxification enzymes (Edge et al., 1997; Noctor et al., 1998; Tripathi et al., 2009; Falk et al., 2010).

However plastoquinone (PQ) and ubiquinone (UQ) are two important prenylquinones functioning as electron transporters in plants and acting an antioxidant role in plant leaves, being able to dissipate energy in the chlorophyll antennae, so they provide an additional protection against photooxidative stress (Swiezewska et al., 2004).

They are located in chloroplast thylakoids and mitochondrial inner membrane respectively; they are both made up of an active benzoquinone ring attached to a polyisoprenoid side chain and the length of polyisoprenoid side chain determines the type of PQ and UQ.

Plastoquinone-9 is well known as a photosynthetic electron carrier involved in electron transfer between PSII and PSI, to which has also been attributed a role in the regulation of gene expression and enzyme activities via its redox state. The plastoquinone biosynthesis pathway was manipulated to generate plant with increased levels of quinones (Ksas et al., 2015).

UQ, particularly UQ_{10}, has also been widely used in people's life. It is effective in treating cardiovascular diseases, chronic gingivitis and periodontitis, and shows favorable impact on cancer treatment and human reproductive health (Liu et al., 2016).

1.5.5. Sterols

Plant sterols, also called phytosterols, have been reported to include over 250 different sterols and related compounds in various plant and marine materials (Akihisa et al., 1991). They are essential components of the membranes of all eukaryotic organisms,

with functions in the control of membrane fluidity, permeability and adaptation of membranes to temperature. Sterols also play an important role in cellular and developmental processes in plants as precursors to the brassinosteroids. They also act as substrates for a wide variety of secondary metabolites such as the glycoalkaloids, cardenolides and saponins (Piironen et al., 2000).

They represent mostly a storage form of sterols, analogously to cholesteryl esters in mammalian organism. The commonly consumed plant sterols are sitosterol, stigmasterol and campesterol which are predominantly supplied by vegetable oils, vegetables and cereals are a less important source (Dutta et al., 1996). The nutritional interest of manipulating levels of sterols in food plants, using them as potential natural dietary products, entirely derives by their capacity to lower plasma cholesterol and LDL cholesterol, because of their similar structure to cholesterol. Indeed, the mortality from cardiovascular disease have been dramatically reduced using cholesterol-lowering drugs (statins) (Piironen et al., 2000).

Few data are available on quantitative changes in plant sterol contents; some data are available for tomato, indeed large increase in the sterol content, specifically of free sterols, steryl glycosides and particularly that of steryl esters increased in Rutgers outer pericarp tomato tissues, ranging from mature green to red ripe (Whitaker et al., 1988).

1.5.6. Fatty acids

Plant oils that are used for human consumption are comprised essentially of only five main fatty acids, the saturated palmitic (C16:0) and stearic (C18:0) acids, monounsaturated oleic acid (C18:1), and the polyunsaturated linoleic (C18:2) and α-linolenic (C18:3) acids.

The production of two very long chain polyunsaturated fatty acids, arachidonic acid (AA) and eicosapentaenoic acid (EPA), in substantial quantities in a higher plant was a breakthrough in the search for alternative sustainable sources of fish oils (Qi et al., 2004). Furthermore, in recent years, considerable focus has been placed on genetically engineering oilseed plants to incorporate additional fatty acids to provide the low-cost production of industrial fatty acids that are currently sourced from petrochemicals or from low-yielding plants (Singh et al., 2005). In general, tomato ripe fruits contained greater enzymatic activities but smaller amounts of linoleic and linolenic acid than green

fruits, since they are used as precursors of many volatile fatty acid derivatives, such as hexanal (Jadhav et al., 1972).

1.5.7. Phospholipids

Phospholipids are a major and vital component of all biological membranes and play a key role in processes such as signal transduction, cytoskeletal rearrangement, and in membrane trafficking; they also play a signalling role in plant growth and development and in the response of plants to biotic and abiotic stresses, inextricably linked to the mechanism of action of plant hormones. Because of their multiple attitudes they are used in agriculture as management tools to improve crop performance and enhance product quality (Cowan et al., 2006).

1.5.8. Amino-acids

Amino acids are precursors of proteins; plants subjected to stress show accumulation of proline and other amino acids. As primary compounds, their role in plants varies from acting as osmolyte, regulators of ion transport, modulators of stomatal opening and detoxyficators of heavy metals (Rai et al., 2002).

Free amino acid content of tomato fruit pericarp increases markedly during ripening transition (Boggio et al. 2000), that involves a substantial turnover of existing and newly synthesized proteins.

The predominant amino acids found in tomato fruit were alanine, arginine, asparagine, aspartic acid, β-alanine, γ-aminobutyric acid, glutamic acid, leucine (and/or isoleucine), serine, threonine and valine. In addition, traces of glycine, lysine and/or histidine and tyrosine are found (Freeman et al., 1960). Glutamic and aspartic acids increased with ripening, reaching a high point in the red ripe fruits. In particular, free glutamate content of ripe tomato fruit is much higher in all the cultivated varieties (Carrari et al. 2006) than in tomato wild species (Schauer et al. 2005). Anyway, the concentration of this amino acid is higher in tomato compared with many vegetables such as carrots, onions or pepper.

1.5.9. Amines and amides

Among low molecular mass nitrogenous compounds, a wide variety of amines, derived from amino acids and precursors of alkaloids, occur in higher plants (Smith et al., 1975). The most widespread free amines in plants are conveniently divided into three groups: aliphatic monoamines, aliphatic di- and polyamines, and aromatic amines, but amines may also be found associated with fatty acids and macromolecules (Cohen et al., 1998). It has been hypothesized that the accumulation of both common and uncommon higher polyamines may serve specific protection roles in plants adapted to extreme environments, but for many of these molecules their actual function in plant tissues has yet to be elucidated. Furthermore, the diet can supply enough amounts of amines and polyamines to support cell renewal and growth. An interesting study reported that fruits of a landrace of tomato have prolonged keeping qualities, determined by higher levels of polyamine putrescine than the standard Rutgers variety at the ripe stage (Rastogi et al., 1990).

1.5.10. Sugars and polyols

In plants, carbohydrates such as sucrose, glucose, fructose, galactose, mannose and ribose, are used as a universal energy currency, as building blocks for cell walls, for signaling, to maintain osmotic homeostasis under certain abiotic stress conditions and for various other purposes. Furthermore, a distint clade of sugar alcohols, called polyols, such as sorbitol, mannitol or inositol and organic acids such as malate or citrate, which are chemical reduced forms of aldose or chetose sugars, can perform different roles in plants.

Although sucrose is probably the most widely studied plant carbohydrate, it has been estimated that up to 30% of all photosynthetically fixed carbon in plants is transported as polyols (Bielesky 1982).

Sugars are also one of the most important source of calories and taste in cultivated crops and are thus critical for both the nutritional and the commercial value of crops (Reuscher et al., 2014).

Regarding specifically the developing tomato fruits, they rappresent a strong sinks for sugars, and an adequate supply from photosynthetic tissues is critical for accumulation of sugars during fruit development, determining the final taste; ripe fruits of commercial

tomato varieties contain mostly glucose and fructose at an equimolar concentration but hardly any sucrose (Ho et al., 1982).

1.5.11. Alkaloids

The steroidal alkaloids , also known as solanum alkaloids, are common constituents of numerous plants belonging to the *Solanaceae* family, in particular members of the genus *Solanum* (Rahman et al., 1998), which comprises 1350 species. Their synthesis, which is presumed to start from cholesterol, likely occurs in the cytosol. In plants, they serve as phytoanticipins, providing a preexisting chemical barrier against a broad range of pathogens (Hoagland et al., 2009).

The major alkaloid in tomato, *α*-tomatine, was reported to be present in the green tissues of the plant together with dehydrotomatine (Kozukue et al., 2004).

Whereas *α*-tomatine levels decrease as the fruit matures and ripens, recent studies suggest a ripening-dependent conversion of *α*-tomatine into esculeoside A, the most abundant alkaloid in the red-ripe tomato fruit (Moco et al., 2007), which also appeared to be ethylene dependent (Iijima et al., 2009).

In tomato, very little is known about the enzymes and genes contributing to their biosynthesis and few studies have provided deeper insights in the biosynthetic pathway in tomato (Itkin et al., 2011).

1.5.12. Phenylpropanoids

General phenylpropanoid metabolism begins with phenylalanine, and involves the activity of three enzymes, phenylalanine ammonia lyase, cinnamate 4-hydroxylase and 4-coumaroyl CoA ligase, to generate *p*-coumaroyl CoA, the activated intermediate for the various branches of phenylpropanoid metabolism (Figure 1.10; Zhang et al., 2015).

Engineering the expression of AtMYB12, a TF regulating flavonol biosynthesis in *Arabidopsis thaliana*, seemed the best option to activated the expression of genes encoding enzymes of flavonol and hydroxycinnamic ester biosynthesis in tomato, resulting in accumulation of exceptionally high levels of both flavonols and caffeoyl quinic acids, amounting to as much as 100 mg g−1 dry weight (Zhang et al., 2015).

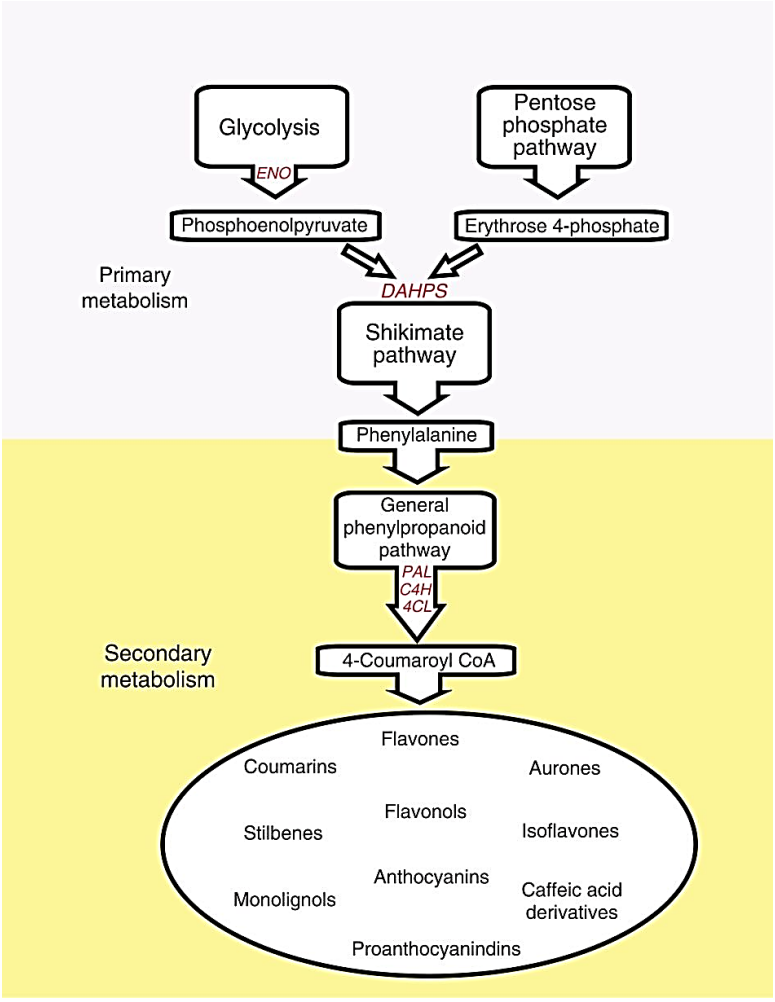


Figure 1.10. Schematic representation of the phenylpropanoid pathway in plants and its relationships to primary metabolic pathways (Zhang et al., 2015).

Phenylpropanoid derivatives accumulate in the stress-affected tissues and are thought to protect plants against various biotic and certain abiotic stressors, serving a specific role in pathogen defence, antiherbivory, ultraviolet screening, energy dissipation and radical scavenging, as well as structural components of cell walls (Grace et Logan 2000). The benefits of specific flavonoids and other phenylpropanoid-derived compounds to human health and their potential for long-term health benefits, due to their antioxidant activities, were also exploited, leading the scientific community to join in the effort to obtain higher amounts of the desired molecules (Ververidis et al., 2007). Indeed tomato is not rich in phenylpropanoids, with the only exception of small amounts in the peel, so the intent of introducing some mutated alleles, such as *Aft* and *atv*, into the San Marzano collection, leads to broaden the range of the already present metabolites in the tomato fruits.

1.5.13. Vitamins

Among vitamins, the nutritional and health value of ascorbic acid (Vitamin C) is of great importance in the human diet, in order to fight diseases such as scurvy, to maintain collagen, to reduce stress damage, and as an antioxidant. Ascorbic acid is important in the biosynthesis of amino acids, formation of adrenaline, and detoxification in the liver (Smirnoff 1996). Almost all of the vitamin C in the human diet is supplied by fruits and vegetables (Goddard and Matthews, 1979), which levels are very variable. Oranges and tomatoes are the principal source of vitamin C for humans, primates, and a few other mammals and passerines that are unable to synthesize this vitamin (Klein and Perry 1982).

Furthermore its content is also valuable from an agronomic point of view, because it contributes to both biotic and abiotic stress tolerance in crops (Davey et al., 2000).

1.6. The role of the tomato volatile compounds in plant and their impact on the consumer preferences

As with many fruits that are part of the human diet, tomato has been domesticated since a few centuries to satisfy human preferences, but originally evolved to attract seed dispersers.

A ripe tomato tipically produces few hundreds of volatile compounds (Tikunov et al., 2005), they are synthetized during the process of fruit ripening, that still include a variety of physical and chemical changes, such as softening of the fruit, increase in the organic acid levels and the synthesis of secondary metabolites to attract seed dispersers. When the fruit is physically damaged, a high amout of selected volatile compounds previous accumulated in a conjugated non-volatile form such as glycosides, is released; This mechanism of volatile storage through glycosylation and wound-induced deconjugation suggests a strategy of discouraging predators from feeding on the fruit before the maturation of the seed, and encouraging to feed on them once seed maturation has been achieved (Rambla et al., 2014).

Volatile compounds are responsable for our perception of flavour and aroma in fruits, due to the difference of many orders of magnitude between their abundance levels (Buttery et al., 1988); indeed a traditional approach to understand which compounds define the aroma consists in establishing the lowest concentration of a compound that

can be perceived by the human nose, a method that allowed to develop a list of almost twenty compounds most influent in determining the human preferences (Buttery et al., 1989). Subsequently the methods was revealed too simplicistic, firstly because the minimum concentrations of each compounds were singularly calculated by the orthonasal perception in a water solution, avoiding the interaction of each volatile compounds in a tomato matrix; furthermore (Bender et al., 2009) the orthonasal and retronasal perception for the same compound are different (Negoias et al., 2008). The last, but not least important factor contributing to the flavour perception of aroma is connected with the wide range of variation in the level of volatile compounds between cultivars (Causse et al., 2002).

Defining what consumers consider a good tomato remains the big challenge for the identification of new targets for genetic improvements.

This work aimed at characterizing the entire San Marzano collection for flavour, focusing on some classes of volatile compounds, benzenoids, branched-chain amino-acids related, apocarotenoids, esters, fatty acid derivatives, phenylalanine derivatives, sulfur compounds and monoterpenoids (Figure 1.11).

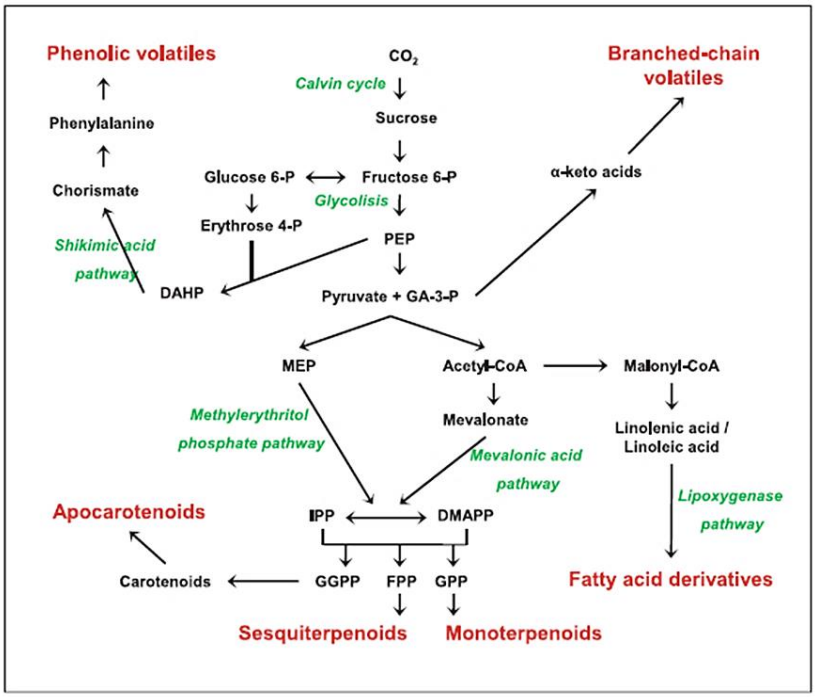


Figure 1.11. Biosynthetic pathways of the most relevant classes of volatiles in the tomato fruit. Volatile classes are highlighted in bold; metabolic pathways are represented in italics. Abbreviations: DAHP, 3-deoxy-d-arabino-heptulosonate 7-phosphate; DMAPP, dimethylallyl diphosphate; FPP, farnesyl diphosphate; GA-3-P, glyceraldehyde-3-phosphate; GGPP, geranyl diphosphate; IPP, isopentenyl diphosphate; MEP, 2-C-methyl-d-erythritol 4-phosphate; PEP, phosphoenolpyruvate (Rambla et al., 2014).

1.6.1. Amino acid derivatives

This class of volatile compounds considered relevant for tomato aroma and derived from amino acids, can be divided in two categories: phenolic and branched chain compounds. Phenolic volatiles come from phenylalanine and can in turn be divided in C6-C2 phenolic volatiles (phenylalanine derivatives), that are considered the most important for the tomato aroma, as 2-phenylethanol (Tieman et al., 2007). The other group (benzenoids) originates from the phenylpropanoid branch of phenylalanine catabolism, such as eugenol and guaiacol, other important compounds for tomato aroma (Koeduka et al., 2006).

Branched-chain amino-acids are a set of compounds with particularly low molecular weight and high volatility, some of which are considered to participate in tomato aroma, such as 3- and 2-methylbutanal, 3-methylbutanol, and 2-isobutylthiazole (Buttery *et al.*, 1989).

1.6.2. Apocarotenoids

Apocarotenoids are synthesized from the oxidative cleavage of double bonds in carotenoids (C40 terpenoids), compounds which are accumulated at high levels in the ripe fruit. These volatile compounds are produced at low levels in the ripe fruit, but are important in our perception of tomato flavour due to their very low odour thresholds, particularly for some cyclic apocarotenoids, such as the C13 ketones β-ionone or β-damascenone (Buttery et al., 1989); carotenoid-derived volatiles have proved to have an important role in tomato flavour,

as their levels positively correlate with tomato flavour acceptability (Vogel *et al.*, 2010). Specific enzymes located in the cytosol can cleave both ciclic and linear carotenoids (Simkin et al., 2004).

Finally two volatiles, 6-methyl-5-hepten-2-one and geranial, are derived from open-chain carotenoids. phytoene or phytofluene and lycopene respectively and the volatile products correlate strongly to the levels of the carotenoid precursors, contributing to the tomato aroma (Lewinsohn *et al.*, 2005).

1.6.3. Esters

Although very abundant and extremely important for the aroma of fruit in many species such as strawberry (Zorrilla-Fontanesi *et al.*, 2012), peach (Sánchez *et al*., 2012), or even some citrus species (González-Mas *et al.*, 2011), few esters are found in the volatile fraction of tomato and they are not relevant for tomato flavour. Indeed they were found to be accumulated in the gree-gruited wild tomatoes, decreasing in the red-fruits, with new evidences supporting that this lack has a positive effect on tomato liking.

1.6.4. Fatty acid derivatives

Volatiles derived from fatty acids constitute a class of compounds which includes the most abundant volatiles produced in the tomato fruit: the C6 volatiles 1-hexanol, (*Z*)-3-hexenal,

(*E*)-2-hexenal, or hexanal, and the C5 volatile 1-penten-3-one and they are classified as green leaf volatiles due to their characteristic 'green', fresh aroma of cut grass. In tomato fruit, the production of these compounds is increased at ripening, probably due to the loss of integrity of cellular membranes (Klee, 2010). The amount of free fatty acids available in the fruit is very limited, as plants accumulate them as acylglycerides rather than in the toxic free form, so the initial step of these volatiles formation was supposed to be the catabolism of the acylglycerides by a lipase, followed by the free fatty acids catabolism through lipoxygenase pathway, being linoleic and linoleic acids the most important substrates (Liavonchanka et al., 2006).

1.6.5. Terpenoids

This class of volatiles are among the most abundant in tomato vegetative tissues and particularly in trichomes, but only a few of them, such as limonene, linalool, or α-terpineol, are present in the ripe fruit, and their impact on tomato aroma is negligible. Geranyl diphosphate IS the precursor of all monoterpenoids and farnesyl diphosphate of all sesquiterpenoids (Granell and Rambla 2013).

1.7. CRISPR-Cas9, its origin as biological system, engineering and application in synthetic biology

Bacteria and archaea defend themselves against invasive DNA and RNA using adaptive immune systems comprising CRISPR (clustered regularly interspaced short palindromic repeats) loci and CRISPR-associated (Cas) genes. In association with Cas proteins, two RNA molecules are required, called a crRNA (CRISPR RNA) and a tracrRNA (trans-activating CRISPR RNA), which perform the same targeting functions, guiding the detection and cleavage of complementary DNA sequences (Jinek et al., 2014). Furthermore, three main types of CRISPR–Cas systems (type I–type III) are mainly distinguishable by the presence of: Cas3 for type I, Cas9 for type II and Cas10 for type III (Makarova et al., 2015).

Engineering of this biological system led to the development of a Synthetic CRISPR/Cas system, which has become the most interesting feature of the CRISPR gene editing. This system makes use of a single molecule, originated by the crRNA and tracrRNA fusion, called single guide RNA (sgRNA; Doudna and Charpentier, 2014).

The single guide RNA can be split into two components: one is specific for each target, is 20 bp in size and must be chosen by the user specifically for every gene editing experiment; the other is an invariant 80 bp sequence, called scaffold, which has a structural role and mediates the interaction of the sgRNA with the Cas9 endonuclease. The other important requirement for a target sequence is the PAM sequence, which is 5'-NGG-3' for Cas9. The Cas9 enzyme generates breaks in double-stranded DNA by using its two catalytic centers to cleave each strand of a DNA target site next to a PAM sequence and matching the 20-nucleotide sequence of the single guide RNA. The sgRNA includes a dual-RNA sequence derived from CRISPR RNA and a separate transcript that binds and stabilizes the Cas9 protein (Figure 1.12). Cas9-sgRNA–mediated DNA cleavage produces a blunt double-stranded break that triggers repair enzymes to disrupt or replace DNA sequences at or near the cleavage site. Catalytically inactive forms of Cas9 can also be used for programmable regulation of transcription and visualization of genomic loci (Doudna & Charpentier, 2014).

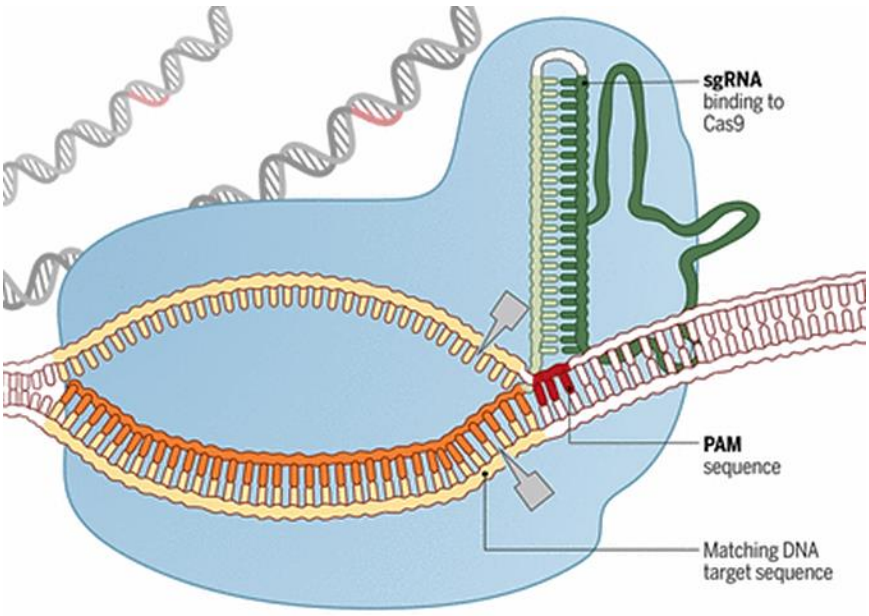


Figure 1.12. The Cas9 enzyme (blue); catalytic centers (blades); DNA target site (gold); PAM sequence (red); 20-nucleotide sequence (orange); single guide RNA (sgRNA); dual-RNA sequence derived from CRISPR RNA (light green); tracrRNA (dark green).

1.7.1. CRISPR-Cas9 applications in tomato

CRISPR/Cas9 mediated gene editing has been widely and successfully used in various organisms for precise targeted gene editing, great advantages in plant functional genomics research and crop improvement, in contrast to the traditional muthagenesis. It has been demonstrated that the CRISPR/Cas9 system can induce mutations in tomato plants by *Agrobacterium tumefaciens*-mediated transformation or transient assays in hairy roots with *A. rhizogens* (Brooks et al., 2014; Ron et al., 2014). Pan et al. (2016) efficiently targeted and obtained editing on two tomato genes, phytoene desaturase (*SlPDS*, Solyc03g123760.2.1) and phytochrome interacting factor PIF4 (*SlPIF4*, Solyc07g043580.2.1), using computationally designed gRNAs with the stable transformed CRISPR/Cas9 system.

One of the mutations used in breeding programs for extend the shelf-life in fruits, *alc*, was also reproduced by editing, obtaining promising results for tomato genetics and breeding (Yu et al., 2017). The *rin* mutation has been also target of the CRISPR-Cas9-mediated mutagenesism to regulate tomato fruit ripening (Ito et al., 2015), as well as *SlMYB12* gene, which has been recently edited by Deng et al. (2018) to the generation of pink-fruited tomato plants.

1.8. Objectives

San Marzano (SM) is one of the most popular Italian tomato landraces, used with the dual purpose of fresh consumption and processing. This PhD research is based on the characterization of a collection of tomato mutant lines that were backcrossed to SM, with different numbers of backcrosses (BC). Three main targets were addressed in this PhD work.

- Chapter II: Firstly, we wanted to analyze how the different levels of BC could impact on the genetic distance of each line from SM, due to the distinct parental donors of each mutation, as well as to the linkage drag phenomenon. This aspect helps us to reflect on how much time a backcross scheme needs to be pursued, with not always adequate results. The collection was phenotyped for aspects of vegetative, reproductive and fruit quality.
- Chapter III: To contribute with a further characterization of these mutation, many of which had already been studied in other different backgrounds, but never compared in a single one, the second target consisted in some metabolic analysis, including polar, non polar and volatile compounds.
- Chapter IV: The last aim was to apply the CRISPR-Cas9 technique to recapitulate two selected mutations, with the final purpose to easily obtain the desired phenotype with the variations typical of the mutation itself.

Bibliography

Adato, A., Mandel, T., Mintz-Oron, S., Venger, I., Levy, D., Yativ, M., et al. Fruit-surface flavonoid accumulation in tomato is controlled by a SlMYB12-regulated transcriptional network. *PLoS Genet.* **2009**, *5* (12), doi: 10.1371/journal.pgen.1000777.

Akhtar, M. S., Goldschmidt, E.E., John, I., Rodoni, S., Matile, P., Grierson, D. Altered patterns of senescence and ripening in gf, a stay-green mutant of tomato (Lycopersicon esculentum Mill. *Journal of Experimental Botany* **1999**, 50(336), 1115-1122.

Akihisa, T., Toshihiro, Y. Naturally occurring sterols and related compounds from plants. Physiology and biochemistry of sterols, **1991**, 172-228.

Bailey, Liberty Hyde. "The standard encyclopedia of horticulture [C]. Vol. II." (**1960**).

Ballester, A. R., Molthoff, J., de Vos, R., te Lintel Hekkert, B., Orzaez, D., Fernández-Moreno, J. P., et al. Biochemical and molecular analysis of pink tomatoes: deregulated expression of the gene encoding transcription factor SlMYB12 leads to pink tomato fruit color. *Plant physiol.* **2010**, *152*(1), 71-84.

Ballester, A.R., Tikunov, Y., Molthoff, J., Grandillo, S., Viquez-Zamora, M., de Vos, R., de Maagd, R.A., van Heusden, S., Bovy, A.G. Identification of loci affecting accumulation of secondary metabolites in tomato fruit of a *Solanum lycopersicum × Solanum chmielewskii* introgression line population. *Front. Plant Sci.* **2016**, 7 (1428), doi.org/10.3389/fpls.2016.01428.

Barrantes, W., Lopez-Casado, G., Garcia-Martinez, S., Alonso, A., Rubio, F., Ruiz, J.J., Fernandez-Munoz, R., Granell, A., Monforte, A.J. Exploring new alleles involved in tomato fruit quality in an introgression line library of *Solanum pimpinellifolium*. *Front. Plant Sci*. **2016**, 7, 1–12. doi.org/10.3389/fpls.2016.01172.

Barry, C. S., Pandey, P. A. survey of cultivated heirloom tomato varieties identifies four new mutant alleles at the green-flesh locus, *Molecular Breeding*, **2009**, 24(3), pp. 269–276. doi:10.1007/s11032-009-9289-4.

Barry, C. S., Giovannoni, J. J. Ripening in the tomato Green-ripe mutant is inhibited by ectopic expression of a protein that disrupts ethylene signaling. *Proceedings of the National Academy of Sciences*, **2006**, *103*(20), 7923-7928.

Barry, C. S., McQuinn, R. P., Thompson, A. J., Seymour, G. B., Grierson, D., & Giovannoni, J. J. Ethylene insensitivity conferred by the Green-ripe and Never-ripe 2 ripening mutants of tomato. *Plant Physiology*, **2005**, *138*(1), 267-275.

Barry, Cornelius S., James J. Giovannoni. Ripening in the tomato Green-ripe mutant is inhibited by ectopic expression of a protein that disrupts ethylene signaling. *Proceedings of the National Academy of Sciences*, **2006**, 103.20: 7923-7928.

Bender, G., Hummel, T., Negoias, S., Small, D.M. Separate signals for orthonasal vs. retronasal perception of food but not nonfood odors. *Behavioral Neuroscience*, **2009**, 123, 481–489.

Bendich, A., Olson, J. A. Biological actions of carotenoids. *The FASEB journal*, **1989**, *3*(8), 1927-1932.

Benvenuto, G., Formiggini, F., Laflamme, P., Malakhov, M., & Bowler, C. The photomorphogenesis regulator DET1 binds the amino-terminal tail of histone H2B in a nucleosome context. *Current Biology*, **2002**, *12*(17), 1529-1534.

Blanc, V.M., Mullin, K., Pichersky, E. Nucleotide sequence of Ipi genes from Arabidopsis and Clarkia (PGR 96-036). *Plant Physiol*. **1996**, 111, 651–652.

Boches, P.S., Peterschmidt, B.C., Myers, J.R. Breeding tomato for increased fruit phenolics. *J. Am. Soc. Hortic. Sci.* **2009**; 44:1055–6.

Brand, M., Moggs, J.G., Oulad-Abdelghani, M., Lejeune, F., Dilworth, F.J., Stevenin, J., Almouzni, G., Tora, L. UV-damaged DNAbinding protein in the TFTC complex links DNA damage recognition to nucleosome acetylation. *EMBO J*, **2001**, 20:3187 3196.

Brooks, C. *et al.* Efficient Gene Editing in Tomato in the First Generation Using the Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-Associated9 System, *Plant Physiology*, **2014**, 166(3), pp. 1292–1297. doi:10.1104/pp.114.247577.

Buttery, RG, Teranishi R, Flath RA, Ling LC. 1989. Fresh tomato volatiles: composition and sensory studies. In: Buttery RG, Shahidi F, Teranishi R, eds. *Flavor chemistry: new trends and developments*. ACS Symposium series 388. Washington, DC: American Chemical Society, 213–222.

Buttery, R.G., Teranishi, R., Ling, L.C., Flath, R.A., Stern, D.J. Quantitative studies on origins of fresh tomato aroma volatiles. *Journal of Agricultural and Food Chemistry,* **1988**, 36, 1247–1250.

Buttery, R. G., Teranishi, R., Flath, R. A., Ling, L. C. Fresh tomato volatiles: Composition and sensory studies. **1989**. 213-222.

Campbell, M., Hahn, F.M., Poulter, C.D. and Leustek, T. Analysis of the isopentenyl diphosphate isomerase gene family from Arabidopsis thaliana. *Plant Mol. Biol*. **1998**, 36, 323–328.

Caramante, M., Rao, R., Monti, L.M., Corrado, G. Discrimination of 'San Marzano' accessions: a comparison of minisatellite, CAPS and SSR markers in relation to morphological traits. *Sci. Hort*. **2009**, 120 (4), 560–564. https://doi.org/10.1016/j.scienta. 2008.12.004.

Carrari, F., Baxter, C., Usadel, B., Urbanczyk-Wochniak, E., Zanor, M. I., Nunes-Nesi, A., Sweetlove, L. J. Integrated analysis of metabolite and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. *Plant Physiology*, **2006**, 142(4), 1380-1396.

Causse, M., Saliba-Colombani, V., Lecomte, L., Duffe, P., Rousselle, P., Buret, M. QTL analysis of fruit quality in fresh market tomato: a few chromosome regions control the variation of sensory and instrumental traits. *Journal of Experimental* Botany, **2002**, 53, 2089–2098.

Chalukova, M., Manuelyan, H. Breeding for Carotenoid Pigments in Tomato. In: Kalloo G. (eds) Genetic Improvement of Tomato. Monographs on Theoretical and Applied Genetics, **1991**, 14. Springer, Berlin, Heidelberg.

Chory, J. Out of darkness: mutants reveal pathways controlling light-regulated development in plants. *Trends Genet*, **1993**, 9:167–172

Cohen, S. S. *Guide to the Polyamines*. **1998**, Oxford University Press.

Colanero, S., Perata, P., Gonzali, S. The atroviolacea gene encodes an R3-MYB protein repressing anthocyanin synthesis in tomato plants. *Frontiers in plant science*, **2018**, *9*, 830.

Cowan, A. K. Phospholipids as plant growth regulators. *Plant Growth Regulation*, **2006**, *48*(2), 97-109.

Cunningham, F.X. Jr and Gantt, E. Identification of multi-gene families encoding isopentenyl diphosphate isomerase in plants by heterologous complementation in Escherichia coli. *Plant Cell Physiol*. **2000**, 41, 119– 123.

Darby, L.A. Isogenic lines of tomato fruit colour mutants. Hortic. Res. 1978, 18, 73–84.

Darby, L. A., D. B. Ritchie, and I. B. Taylor. "Isogenic lines of the tomato 'Ailsa Craig'." *Annual Report Glasshouse Crops Research Institute*, **1977**, 168-184.

Datta, A., Bagchi, S., Nag, A., Shiyanov, P., Adami, G.R., Yoon, T., Raychaudhuri, P. The p48 subunit of the damaged-DNA binding protein DDB associates with the CBP/p300 family of histone acetyltransferase. *Mutat Res*, **2001**, 486:89–97.

Davey, M.W., Van Montagu, M., Inzé, D., Sanmartin, M., Kanellis, A., Smirnoff, N., Benzie, I.J.J., Strain, J.J., Favell, D., Fletcher, J. Plant L-ascorbic acid: chemistry, function, metabolism, bioavailability and effects of processing. *Journal of the Science of Food and Agriculture*, **2000**, 80, 825–860.

Gates, D. M., Keegan, H. J., Schleter, J. C., & Weidner, V. R. Spectral properties of plants. *Applied optics*, **1965**, 4(1), 11-20.

Deng, L., Wang, H., Sun, C., Li, Q., Jiang, H., Du, M., Li, C. Efficient generation of pink-fruited tomatoes using CRISPR/Cas9 system. *Journal of genetics and genomics= Yi chuan xue bao*, **2018**, *45*(1), 51.

Dharmapuri, S., Rosati, C., Pallara, P., Aquilani, R., Bouvier, F., Camara, B., Giuliano, G. Metabolic engineering of xanthophyll content in tomato fruits. *Febs Letters*, **2002**, 519(1-3), 30-34.

Dixon, R. A. Natural products and plant disease resistance. *Nature*, **2001**, 411(6839), 843-847.

Doudna, J. A., Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science*, **2014**, 346(6213), 1258096–1258096. doi: 10.1126/science.1258096.

Dutta, P. C., Appelqvist, L.A. Saturated sterols (stanols) in unhydrogenated and hydrogenated edible vegetable oils and in cereal lipids. *J Sci Food Agric*. **71**, 383–391 (1996).

Edge, R., McGarvey, D. J., Truscott, T. G. The carotenoids as anti-oxidants – a review. *J. Photochem. Photobiol. B: Biol*. **1997**, 41, 189–200.

Ercolano, M.R., Carli, P., Soria, A., Cascone, A., Fogliano, V., Frusciante, L., Barone, A. Biochemical, sensorial and genomic profiling of traditional Italian tomato varieties. *Euphytica*. **2008**, 164 (2), 571–582.

Ercolano, M.R., Sacco, A., Ferriello, F., D'Alessandro, R., Tononi, P., Traini, A., Barone, A., Zago, E., Chiusano, M.L., Buson, G., Delledonne, M., Frusciante, L. Patchwork sequencing of tomato San Marzano and Vesuviano varieties highlights genome-wide variations. *BMC Genom.* **2014**, 15 (1), 1–13. https://doi.org/10.1186/1471-2164-15-138.

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

Eshed, Y., Zamir, D. An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics*. **1995**, 141 (3), 1147–1162.

Falk, J., Munné-Bosch, S. Tocochromanol functions in plants: antioxidation and beyond. J. *Exp. Bot.* **2010**, 61, 1549–1566.

Finkers, R., van Heusden, A.W., Meijer-Dekens, F., van Kan, J.A., Maris, P., Lindhout, P.The construction of a *Solanum habrochaites* LYC4 introgression line population and the identification of QTLs for resistance to Botrytis cinerea. *Theor. Appl. Genet*. **2007**, 114 (6), 1071–1080. https://doi.org/10.1007/s00122-006-0500-2.

Foolad, M.R. Genome mapping and molecular breeding of tomato. *Intl. J. Plant. Genom.* **2007**, 64358. https://doi.org/10.1155/2007/64358. Article ID.

Ford, N., Erdman, J. W. Lycopene and cancer. In *Carotenoids and Human Health*, **2013**, 193-214, Humana Press, Totowa, NJ.

Fraser, P.D., Bramley, P.M. The biosynthesis and nutritional uses of carotenoids. *Prog. Lipid.Res*. **2004**, 43: 228-265.

Fraser, P. D., Bramley, P. M. Metabolic profiling and quantification of carotenoids and related isoprenoids in crop plants. *Plant Metabolomics*, **2006**, 229-242, Springer, Berlin, Heidelberg.

Fraser, P.D., Kiano, J.W., Truesdale, M.R., Schuch, W., and Bramley, P.M. Phytoene synthase-2 enzyme activity in tomato does not contribute to carotenoid synthesis in ripening fruit. *Plant. Mol*. **1999**, Biol. 40: 687–698.

Fray, R.G., Grierson, D. Identification and genetic analysis of normal and mutant phytoene synthase genes of tomato by sequencing, complementation and co-suppression. *Plant Mol. Biol.* **1993**, 22, 589–602.

Freeman, J. A., Woodbridge, C. G. Effect of maturation, ripening and truss position on the free amino acid content in tomato fruits. In *Proceedings. American Society for Horticultural Science*, **1960**, 76, 515-23.

Friedman, M., Levin, C. E. Dehydrotomatine content in tomatoes. *Journal of agricultural and food chemistry*, **1998**, *46*(11), 4571-4576.

Friedman, M., Kozukue, N., Harden, L. A. Structure of the tomato glycoalkaloid tomatidenol-3-β-lycotetraose (dehydrotomatine). *Journal of Agricultural and Food Chemistry*, **1997**, *45*(5), 1541-1547.

Friedman, M., Kozukue, N., Harden, L. A. Structure of the tomato glycoalkaloid tomatidenol-3-β-lycotetraose (dehydrotomatine). *Journal of Agricultural and Food Chemistry*, **1997**, *45*(5), 1541-1547.

Fu, D., Wakasugi, M., Ishigaki, Y., Nikaido, O., Matsunaga, T. cDNA cloning of the chicken DDB1 gene encoding the p127 subunit of damaged DNA-binding protein. *Genes Genet. Syst*. **2003**, 78:169–77.

Garcia-Martinez, S., Corrado, G., Ruiz, J.J., Rao, R. Diversity and structure of a sample of traditional Italian and Spanish tomato accessions. *Genet. Res. Crop Evol.* **2013**, 60 (2), 789–798 doi.org/10.1007/s10722-012-9876-9.

Giorgiev, C. Anthocyanin fruit tomato. *Rep. Tomato Genet*. **1972**, 22:10.

Giuliano, G., Al-Babili, S., von Lintig, J. Carotenoid oxygenases: cleave it or leave it. *Trend Plant. Sci.* **2003**, 8:145-149.

Goddard, M.C. and matthews, R. H. Contribution of fruitsand vegetables to human nutrition. *Hort. Science*, **1979**, 14, 245-247.

González-Mas, M.C., Rambla, J.L., Alamar, M.C., Gutiérrez, A., Granell, A. Comparative analysis of the volatile fraction of fruit juice from different Citrus species. *PLoS One*, **2011**, 6, e22016.

Grace, S.C., Logan, B.A. Energy dissipation and radical scavenging by the plant phenylpropanoid pathway. *Phil Trans Royal Soc London Series, B – Biol Sci*, **2000**, 355: 1499–1510.

Granell, A., Rambla, J.L. Biosynthesis of volatile compounds. In: Seymour G, Tucker GA, Poole M, Giovannoni JJ, eds. *The molecular biology and biochemistry of fruit ripening. Oxford: Wiley-Blackwell,* **2013**, *135–161.*

Granell, A., Rambla, J.L. Biosynthesis of volatile compounds. In: Seymour G, Tucker GA, Poole M, Giovannoni JJ, eds. *The molecular biology and biochemistry of fruit ripening. Oxford: Wiley-Blackwell,* **2013**, *135–161.*

Grusak, M. A., DellaPenna, D. Improving the nutrient composition of plants to enhance human nutrition and health. *Annual review of plant biology*, **1999**, *50*(1), 133-161.

Gutensohn, Michael, Dinesh A. Nagegowda, and Natalia Dudareva. Involvement of compartmentalization in monoterpene and sesquiterpene biosynthesis in plants. *Isoprenoid synthesis in plants and microorganisms.* Springer, New York, NY, **2012**. 155-169.

Yoo, H.J., Park, W.J., Lee, G.M., Oh, C.S., Yeam, I., Won, D.C., Lee, J.M. Inferring the genetic determinants of fruit colors in tomato by carotenoid profiling. *Molecules*, **2017**, 22 (5), 764.

Ho, L.C., Sjut, V., Hoad, G.V. The effect of assimilate supply on fruit growth and hormone levels in tomato plants. *Plant Growth Regulation*, **1982**, 1(3), 155-171.

Hoagland, R.E. Toxicity of tomatine and tomatidine on weeds, crops and phytopathogenic fungi. *Allelopathy J.* **2009**, 23: 425–436.

Hörtensteiner, S., Kräutler, B. Chlorophyll breakdown in higher plants. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, **2011**, 1807(8), 977-988.

Horvath, G., Wessjohann, L., Bigirimana, J., Jansen, M., Guisez, Y., Caubergs, R., Horemans, N. Differential distribution of tocopherols and tocotrienols in photosynthetic and non-photosynthetic tissues. *Phytochemistry*, **2006**, *67*(12), 1185-1195.

Hunt, G.M., Baker, E.A. Phenolic constituents of tomato fruit cuticles. *Phytochemistry*, **1980**, 19, 1415–1419.

Iijima, Y., Fujiwara, Y., Tokita, T., Ikeda, T., Nohara, T., Aoki, K., Shibata, D. Involvement of ethylene in the accumulation of esculeoside A during fruit ripening of tomato (*Solanum lycopersicum*). *J. Agric. Food Chem.* **2009**, 57: 3247–3252

Isaacson, T., Ronen, G., Zamir, D., & Hirschberg, J. Cloning of tangerine from tomato reveals a carotenoid isomerase essential for the production of β-carotene and xanthophylls in plants. *The Plant Cell*, **2002**, *14*(2), 333-342.

Ishibashi, T., Kimura, S., Yamamoto, T., Furukawa, T., Takata, K., Uchiyama, Y., Hashimoto, J., Sakaguchi, K. Rice UVdamaged DNA binding protein homologues are most abundant in proliferating tissues. *Gene*, **2003**, 308:79–87

Itkin, M., Rogachev, I., Alkan, N., Rosenberg, T., Malitsky, S., Masini, L., Prusky, D. GLYCOALKALOID METABOLISM1 is required for steroidal alkaloid glycosylation and prevention of phytotoxicity in tomato. *The Plant Cell*, **2011**, *23*(12), 4507-4525.

Ito, Y., Nishizawa-Yokoi, A., Endo, M., Mikami, M., & Toki, S. CRISPR/Cas9-mediated mutagenesis of the RIN locus that regulates tomato fruit ripening. *Biochemical and biophysical research communications,* **2015**, 467(1), 76-82.

Jadhav, S., Singh, B., & Salunkhe, D. K. Metabolism of unsaturated fatty acids in tomato fruit: linoleic and linolenic acid as precursors of hexanal. *Plant and cell physiology*, **1972**, *13*(3), 449-459.

Jenkins, J., Mackinney, G. Carotenoids of the apricot tomato and its hybrids with yellow and tangerine. *Genetics*, **1955**, 40, 715–720.

Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., Kaplan, M. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*, **2014**, 343(6176), 1247997.

Kachanovsky, D. E., Filler, S., Isaacson, T., Hirschberg, J. Epistasis in tomato color mutations involves regulation of phytoene synthase 1 expression by cis-carotenoids. *Proceedings of the National Academy of Sciences*, **2012**, 109(46), 19021-19026.

Kerr, E.A. Green flesh, gf. *Rpt Tomato Genet Coop*, **1956**, 6: 17.

Klee, H.J. Improving the flavor of fresh fruits: genomics, biochemistry, and biotechnology. *New Phytologist*, **2010**, 187, 44–56.

Klee, H. J., Giovannoni, J. J. Genetics and control of tomato fruit ripening and quality attributes. *Annual review of genetics*, **2011**, 45, 41-59.

Klein, B. P., Perry, A. K. Ascorbic acid and vitamin A activity in selected vegetables from different geographical areas of the United States. *Journal of Food Science*, **1982**, 47(3), 941-945.

Koeduka, T., Fridman, E., Gang, D. R., Vassao, D. G., Jackson, B. L., Kish, C. M., Baiga, T. J. Eugenol and isoeugenol, characteristic aromatic constituents of spices, are biosynthesized via reduction of a coniferyl alcohol ester. Proceedings of the National Academy of Sciences, **2006**, 103(26), 10128-10133.

Konsler, T.R. Three mutants appearing in 'Manapal' tomato. *Hort. Sci*. **1973**, 8:331–333

Kozukue, N., Han, J.S., Lee, K.R., Friedman, M. Dehydrotomatine and α-tomatine content in tomato fruits and vegetative plant tissues. *J. Agric. Food Chem*. **2004**, 52: 2079–2083.

Ksas, B., Becuwe, N., Chevalier, A., Havaux, M. Plant tolerance to excess light energy and photooxidative damage relies on plastoquinone biosynthesis. *Scientific reports*, **2015**, *5*(1), 1-16.

Laval-Martin, D., Quennemet, J., Monéger, R. Pigment evolution in Lycopersicon esculentum fruits during growth and ripening. *Phytochemistry*, **1975**, 14(11), 2357-2362.

Levin, I., De Vos, C.R., Tadmor, Y., Bovy, A., Lieberman, M., Oren-Shamir, M., Segev, O., Kolotilin, I., Keller, M., Ovadia, R., Meir, A., Bino, R.J. High pigment tomato mutants - more than just lycopene (a review). *Isr. J. Plant Sci*. **2006**, 54 (3), 179–190.

Lewinsohn, E., Sitrit, Y., Bar, E., Azulay, Y., Ibdah, M., Meir, A., Yoser, E., Zamir, D., Tadmor, Y. Carotenoid pigmentation affects the volatile composition of tomato and watermelon fruits, as revealed by comparative genetic analysis. *Journal of Agricultural and Food Chemistry,* **2005**, 53, 3142–3148.

Liavonchanka, A., Feussner, I. Lipoxygenases: occurrence, functions and catalysis. *Journal of Plant Physiology*, **2006**, 163, 348–357.

Lieberman, M., Sege, O., Gilboa, N., Lalazar, A., Levin, I. The tomato homolog of the gene encoding uv-damaged dna binding protein 1 (ddb1) underlined as the gene that causes the *high pigment-1* mutant. *Theor. Appl. Genet*. **2004**, 108, 1574–1581. https://doi.org/10.1007/s00122-004-1584-1.

Lincoln, R.E., Porter, J.W. Inheritance of β-carotene in tomatoes. *Genetics,* **1950**, 35:206-211.

Lindstrom, E.W. Inheritance in tomato. *Genetica*, **1925**, *10 (4)*, 305.

Liu, Y., Roof, S., Ye, Z., Barry, C., Van Tuinen, A., Vrebalov, J., Bowler, C., Giovannoni, J. Manipulation of light signal transduction as a means of modifying fruit nutritional quality in tomato. *Proc. Natl. Acad. Sci. USA*, **2004**, 101:9897–9902.

Liu, M., Lu, S. Plastoquinone and ubiquinone in plants: biosynthesis, physiological function and metabolic engineering. *Frontiers in plant science*, **2016**, *7*, 1898.

Luque, P., Bruque, S., Heredia, A. Water permeability of isolated cuticular membranes: a structural analysis. *Arch. Biochem. Biophys*. **1995**, 317: 417–422.

Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J., Moineau, S. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology*, **2019**, 1-17.

Martinez, E., Palhan, V.B., Tjernberg, A., Lymar, E.S., Gamper, A.M., Kundu, T.K., Chait, B.T., Roeder, R.G. Human STAGA complex is a chromatin-acetylating transcription coactivator that interacts with pre-mRNA splicing and DNA damage binding factors in vivo. *Mol. Cell. Biol.* **2001**, 21:6782–6795.

Martínez-Valverde, I., Periago, M. J., Provan, G., Chesson, A. Phenolic compounds, lycopene and antioxidant activity in commercial varieties of tomato (Lycopersicum esculentum). *Journal of the Science of Food and Agriculture*, **2002**, 82(3), 323-330.

Mazzucato, A., Papa, R., Bitocchi, E., Mosconi, P., Nanni, L., Negri, V., Veronesi, F. Genetic diversity, structure and marker-trait associations in a collection of Italian tomato (*Solanum lycopersicum L.*) landraces. *Theoretical and Applied Genetics*, **2008**, 116(5), 657-669.

Mène-Saffrané, L., DellaPenna, D. Biosynthesis, regulation and functions of tocochromanols in plants. *Plant Physiology and Biochemistry*, **2010**, 48(5), 301-309.

Mes, P.J., Boches, P., Myers, J. R., Durst, R. Characterization of tomatoes expressing anthocyanin in the fruit. *Journal of the American Society for Horticultural Science*, **2008**, 133(2), 262-269.

Minoggio, M., Bramati, L., Simonetti, P., Gardana, C., Iemoli, L., Santangelo, E., Pietta, P. G. Polyphenol pattern and antioxidant activity of different tomato lines and cultivars. *Annals of nutrition and metabolism*, **2003**, 47(2), 64-69.

Moco S., Capanoglu E., Tikunov Y., Bino R.J., Boyacioglu D., Hall R.D., Vervoort J., De Vos R.C. Tissue specialization at the metabolite level is perceived during the development of tomato fruit. *J. Exp. Bot*. **2007**, 58: 4131–4146.

Moco, S., Schneider, B., Vervoort, J. Plant micrometabolomics: the analysis of endogenous metabolites present in a plant cell or tissue. *Journal of Proteome Research*, **2009**, 8(4), 1694-1703.

Monti, L.M., Santangelo, E., Corrado, G., Rao, R., Soressi, G.P., Scarascia Mugnozza, G.T. Il San Marzano: problematiche e prospettive in relazione alla sua salvaguardia e alla necessita di interventi genetici. *Agroindustria*. **2004**, 3 (2), 161–170.

Mustilli, A.C., Fenzi, F., Ciliento, R., Alfano, F., Bowler, C. Phenotype of the tomato high pigment-2 mutant is caused by a mutation in the tomato homolog of DEETIOLATED1. *Plant Cell*, **1999**, 11 (2), 145–157 doi.org/10.1105/tpc.11.2.145.

Nakamura, A., Shimada, H., Masuda, T., Ohta, H. and Takamiya, K. Two distinct isopentenyl diphosphate isomerases in cytosol and plastid are differentially induced by environmental stresses in tobacco. *FEBS Lett.* **2001**, 506, 61–64.

Negoias, S., Visschers, R., Boelrijk, A., Hummel, T. New ways to understand aroma perception. *Food* Chemistry, **2008**, 108, 1247–1254.

Nisar, N., Li, L., Lu, S., Khin, N. C., & Pogson, B. J. Carotenoid metabolism in plants. *Molecular plant*, **2015**, 8(1), 68-82.

Noctor, G., Foyer, C. H. Ascorbate and glutathione: keeping actice oxygen species under control. Annu. Rev. *Plant Physiol. Plant Mol. Biol*. **1998**, 49, 249–279

Pan, C., Ye, L., Qin, L., Liu, X., He, Y., Wang, J., Lu, G. CRISPR/Cas9-mediated efficient and heritable targeted mutagenesis in tomato plants in the first and later generations. *Scientific reports*, **2016**, *6*, 24765.

Pankratov, I., McQuinn, R., Schwartz, J., Bar, E., Fei, Z., Lewinsohn, E., ... & Hirschberg, J. Fruit carotenoid-deficient mutants in tomato reveal a function of the plastidial isopentenyl diphosphate isomerase (IDI 1) in carotenoid biosynthesis. *The Plant Journal*, **2016**, 88(1), 82-94.

Peters, J.L., van Tuinen, A., Adamse, P., Kendrick, R.E., Koornneef, M. High pigment mutants of tomato exhibit high sensitivity for phytochrome action. *J. Plant. Physiol*. **1989**, 134:661–666

Philouze, J. Description of isogenic lines, except for one, or two, monogenically controlled morphological traits in tomato, *Lycopersicon esculentum* Mill. *Euphytica*, **1991**, 56 (2), 121–131.

Piironen, V., Lindsay, D. G., Miettinen, T. A., Toivo, J., Lampi, A. M. Plant sterols: biosynthesis, biological function and their importance to human nutrition. *Journal of the Science of Food and Agriculture*, **2000**, 80(7), 939-966.

Qi, B., Fraser, T., Mugford, S., Dobson, G., Sayanova, O., Butler, J., Lazarus, C. M. Production of very long chain polyunsaturated omega-3 and omega-6 fatty acids in plants. *Nature biotechnology*, **2004**, 22(6), 739-745.

Rahman, A. U., & Choudhary, M. I. Chemistry and biology of steroidal alkaloids. In *The Alkaloids: Chemistry and Biology*, **1998**, 50, 61-108, Academic Press.

Rai, V. K. Role of amino acids in plant responses to stresses. *Biologia plantarum*, **2002**, *45*(4), 481-487.

Raiola, A., Tenore, G. C., Barone, A., Frusciante, L., & Rigano, M. M. Vitamin E content and composition in tomato fruits: beneficial roles and bio-fortification. *International journal of molecular sciences*, **2015**, *16*(12), 29250-29264.

Ramos-Valdivia, A.C., VanderHeijden, R., Verpoorte, R. and Camara, B. Purification and characterization of two isoforms of isopentenyldiphosphate isomerase from elicitor-treated Cinchona robusta cells. *Eur. J. Biochem*. **1997**, 249, 161–170.

Rao, R., Corrado, G., Bianchi, M., Di Mauro, A. (GATA)4 DNA fingerprinting identifies morphologically characterized 'San Marzano' tomato plants. *Plant Breed.* **2006**, 125 (2), 173–176. https://doi.org/10.1111/j.1439-0523.2006.01183.x.

Rastogi, R., Davies, P. J. Polyamine metabolism in ripening tomato fruit: I. Identification of metabolites of putrescine and spermidine. *Plant physiology*, **1990**, *94*(3), 1449-1455.

Reuscher, S., Akiyama, M., Yasuda, T., Makino, H., Aoki, K., Shibata, D., & Shiratake, K. The sugar transporter inventory of tomato: genome-wide identification and expression analysis. *Plant and Cell Physiology*, **2014**, *55*(6), 1123-1141.

Reynard, G.B. Origin of Webb Special (Black Queen) in tomato. *Rep. Tomato. Genet. Coop.* **1956**, 40:44–64.

Rick, C.M., Reeves, A.F., Zobel, R.W. Inheritance and linkage relations of four newmutants. *Tomato Genet. Coop. Rep*. **1968**, 18:34–5

Rick, C. M., Simmond, N. Evolution of crop plants. **1976**.

Rick, C. M., Butler, L. Cytogenetics of the tomato. *Adv. Genet.* **1956***, 8*, 267-382.

Ronen, G., Cohen, M., Zamir, D., Hirschberg, J. Regulation of carotenoid biosynthesis during tomato fruit development: expression of the gene for lycopene epsilon-cyclase is down-regulated during ripening and is elevated in the mutant Delta. *Plant J*. **1999**, 17, 341–351.

Rosati, C., Aquilani, R., Dharmapuri, S., Pallara, P., Marusic, C., Tavazza, R., Giuliano, G. Metabolic engineering of beta-carotene and lycopene content in tomato fruit. *The Plant Journal*, **2000**, 24(3), 413-420.

Sanchez, G., Besada, C., Badenes, M.L., Monforte, A.J., Granell, A. A non-targeted approach unravels the volatile network in peach fruit. *PLoS One,* **2012**, 7, e38992.

Sapir, M., Oren-Shamir, M., Ovadia, R., Reuveni, M., Evenor, D., Tadmor, Y. Molecular aspects of Anthocyanin fruit tomato in relation to high pigment-I. *Journal of Heredity*, **2008**, 99:292–303

Savo Sardaro, M.L., Marmiroli, M., Maestri, E., Marmiroli, N. Genetic characterization of Italian tomato varieties and their traceability in tomato food products. *Food Sci. Nutr. 1*, **2013**, 54–62. https://doi.org/10.1002/fsn3.8.

Schauer, N., Semel, Y., Roessner, U., Gur, A., Balbo, I., Carrari, F., Willmitzer, L. Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nature biotechnology*, **2006**, 24(4), 447-454.

Schneider, C. Chemistry and biology of vitamin E. *Molecular nutrition & food research*, **2005**, *49*(1), 7-30.

Schroeder, D. F., Gahrtz, M., Maxwell, B. B., Cook, R. K., Kan, J. M., Alonso, J. M., Chory, J. De-etiolated 1 and damaged DNA binding protein 1 interact to regulate Arabidopsis photomorphogenesis. *Current Biology*, **2002**, 12(17), 1462-1472.

Schwartz, S.H., Tan, B.C., Gage, D.A., Zeevaart, J.A., McCarty, D.R. Specific oxidative cleavage of carotenoids by VP14 of maize. *Science*, **1997**, 276: 1872-1874.

Seymour, G. B., Taylor, J. E., & Tucker, G. A. Biochemistry of fruit ripening Chapman & Hall. *London, England.* **1993**, 327-341.

Simkin, A.J., Schwartz, S.H., Auldridge, M., Taylor, M.G., Klee, H.J. The tomato carotenoid cleavage dioxygenase 1 genes contribute to the formation of the flavor volatiles beta-ionone, pseudoionone, and geranylacetone. *The Plant Journal*, **2004**, 40, 882–892.

Singh, S. P., Zhou, X. R., Liu, Q., Stymne, S., Green, A. G. Metabolic engineering of new fatty acids in plants. *Current opinion in plant biology*, **2005**, *8*(2), 197-203.

Slimestad, R., Verheul, M. Review of flavonoids and other phenolics from fruits of different tomato (*Lycopersicon esculentum Mill*.) cultivars. *Journal of the Science of Food and Agriculture*, **2009**¸*89*(8), 1255-1270.

Smirnoff, N. Briefing botanico: funzione e metabolismo dell'acido ascorbico nelle piante. *Annali di botanica*, **1996**, *78* (6), 661-669.

Smith, J.M., Ritchie, D.B. A collection of near-isogenic lines of tomato: research tool of the future? *Plant Mol. Biol. Rep.* **1983**¸1 (1), 41–45.

Smith, Terence A. Recent advances in the biochemistry of plant amines. *Phytochemistry*, **1975**, 14.4, 865-890.

Soressi, G.P. Il pomodoro. *Ed Agricole*, Bologna, **1969**.

Soressi, G.P. New spontaneous or chemically-induced fruit ripening tomato mutants. Rep. *Tomato Genet. Coop*. **1975**, 25, 21–22.

Sun, Z., Cunningham, F.X. Jr and Gantt, E. Differential expression of two isopentenyl pyrophosphate isomerases and enhanced carotenoid accumulation in a unicellular chlorophyte. *Proc. Natl Acad. Sci. USA*, **1998**, 95, 11482–11488.

Swiezewska, E. Ubiquinone and plastoquinone metabolism in plants. *Methods Enzymol.* **2004**, 378, 124–131. doi: 10.1016/S0076-6879(04)78007-6

Swiezewska, E., Dallner, G., Andersson, B., Ernster, L. Biosynthesis of ubiquinone and plastoquinone in the endoplasmic reticulum-Golgi membranes of spinach leaves. *Journal of Biological Chemistry*, **1993**, *268*(2), 1494-1499.

Thomas, H., Howarth, C. J. Five ways to stay green. *Journal of experimental botany*, **2000**, *51*(suppl_1), 329-337.

Tieman, D.M., Loucas, H.M., Kim, J.Y., Clark, D.G., Klee, H.J. Tomato phenylacetaldehyde reductases catalyze the last step in the synthesis of the aroma volatile 2-phenylethanol. *Phytochemistry*, **2007**, 68, 2660–2669.

Tigchelaar, E., McGlasson, W., Buescher, R. Genetic regulation of tomato fruit ripening. *Hort. Science*, **1978**, 13, 508-513.

Tikunov, Y., Lommen, A., De Vos, C. R., Verhoeven, H. A., Bino, R. J., Hall, R. D., Bovy, A. G. A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. *Plant physiology*, **2005**, 139(3), 1125-1137.

Tomes, M. L., Quackenbush, F. W., Kargl, T. E. Action of the gene B in biosynthesis of carotenes in the tomato. *Botanical Gazette*, **1956**, 117(3), 248-253.

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

Tripathi, B. N., Bhatt, I., Dietz, K. J. Peroxiredoxins: a less studied component of hydrogen peroxide detoxification in photosynthetic organisms. *Protoplasma*, **2009**, 235, 2–15.

Van Tuinen, A., Cordonnier-Prat, M.M., Pratt, L.H., Verkerk, R., Zabel, P., Koornneef, M. The mapping of phytochrome genes and photomorphogenic mutants of tomato. *Theor Appl Genet*, **1997**, 94:115–122.

Verpoorte, R., van der Heijden, R., Memelink, J. Engineering the plant cell factory for secondary metabolite production. *Transgenic research*, **2000**, *9*(4-5), 323-343.

Ververidis, F., Trantas, E., Douglas, C., Vollmer, G., Kretzschmar, G., & Panopoulos, N. Biotechnology of flavonoids and other phenylpropanoid-derived natural products. Part II: Reconstruction of multienzyme pathways in plants and microbes. *Biotechnology Journal: Healthcare Nutrition Technology*, **2007**, *2*(10), 1235-1249.

Vogel, J.T., Tieman, D.M., Sims, C.A., Odabasi, A.Z., Clark, D.G., Klee, H.J. Carotenoid content impacts flavour acceptability in tomato (*Solanum lycopersicum*). *Journal of the Science of Food and Agriculture*, **2010**, 90, 2233–2240

Vranova, E., Coman, D. and Gruissem, W. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annu. Rev. Plant Biol.* **2013**, 64, 665–700.

Vrebalov, J., Ruezinsky, D., Padmanabhan, V., White, R., Medrano, D., Drake, R., Schuch, W., Giovannoni, J. A MADS-box gene necessary for ripening at the tomato ripening-inhibitor (*rin*) locus. *Science*, **2002**, 296, 343±346.

Whitaker, B. D. Changes in the steryl lipid content and composition of tomato fruit during ripening. *Phytochemistry*, **1988**, 27(11), 3411-3416.

Yu, Q.H., Wang, B., Li, N., Tang, Y., Yang, S., Yang, T., Asmutola, P. Mutagenesi mirata indotta da CRISPR / Cas9 e sostituzione genica per generare linee di pomodori a lunga conservazione. *Rapporti scientifici*, **2017**, *7* (1), 1-9.

Zanfini, A., Corbini, G., la Rosa, C., Dreass, E. Antioxidant activity of tomato lipophilic extracts and interactions between carotenoids and α-tocopherol in synthetic mixtures. *Food Sci. Technol*. **2010**, 43, 67–72.

Zhang, Y., Butelli, E., Alseekh, S., Tohge, T., Rallapalli, G., Luo, J., Martin, C. Multi-level engineering facilitates the production of phenylpropanoid compounds in tomato. *Nature communications*, **2015**, 6, 8635.

Zorrilla-Fontanesi, Y., Rambla, J.L., Cabeza, A., Medina, J.J., Sánchez- Sevilla, J.F., Valpuesta, V., Botella, M.A., Granell, A., Amaya, I. Genetic analysis of strawberry fruit aroma and identification of O-methyltransferase FaOMT as the locus controlling natural variation in mesifurane content. *Plant Physiology*, **2012**, 59, 851–870.

Chapter 2.

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

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_moB*), 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.

All results reported in this chapter have been published in the journal *Scientia Horticolturae* (Dono et al., 2020).

2.1. Introduction

In this chapter, 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.

2.2. Materials and methods

2.2.1. Plant material and growth conditions

Nineteen tomato lines with SM-type fruits have been studied (Table 2.1). 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_moB* line, obtained with the combination of the variant *High beta* (*B*) with its modifier *moB*, 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 S2.1.

To combine mutations, single mutants ILs were hand-crossed and the F1 generation grown to obtain F2 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 double mutants and therefore the degree of backcrossing of these lines was

estimated as the mean of the number of BCs of the two parent lines (Table 2.1). Two replicates of 20 seed for each line were germinated in Petri dishes with 3 ml of sterile water. Germination was monitored after four and ten days. Eight plants per accession at the 4–5th true leaf stage were transplanted in twin rows (100 cm between twins, 60 cm between

rows and 50 cm between plants within the row) in an unheated tunnel located at the University of Tuscia's Experimental Farm at Viterbo, Italy (42°260'N, 12°040'E). Plants were grown following standard cultural practices for indeterminate tomatoes, using tutors and weekly removal of lateral shoots. Daily temperature was maintained below 30 °C by a ventilation system and the plants were irrigated through a drop irrigation system.

The trial was repeated with identical materials and methods for two consecutive years (2017 and 2018). As outgroups for the genotypic analysis, plants of an IL of *S. chmielewskii* in VF145-22-8 background (LA1563) and of a *S. pimpinellifolium* accession (LA1589) were grown to extract DNA. The two accessions were obtained from TGRC.

Table 2.1. 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
Wild-type	- [a]	San Marzano	WT	-	-	-
San Marzano fruit variants	All pigments	*high pigment-1*	*hp-1*	2	5	0.024
		high pigment-2	*hp-2*	2	6	0.007
		pigment diluter	*pd*	*1*	4	0.044
	Carotenoids	*yellow flesh*	*r*	*5*	6	0.056
		tangerine	*t*	*4*	6	0.157
		apricot	*at*	1	4	0.049
		High Beta	*B*	1	2	0.067
		High Beta + Beta modifier	*B_moB*	*2*	4	0.062
	Chlorophyll	*green flesh*	*gf*	4	5	0.025
	Flavonoids	*colourless fruit epidermis*	*y*	3	2	0.018
	Ripening	*Never ripe*	*Nr*	4	5	0.015
		ripening inhibitor	*rin*	4	5	0.132
		Green ripe	*Gr*	2	5	0.035
	Double mutants	*yellow flesh + colourless fruit epidermis*	*r_y*	4[b]	4	0.022
		green flesh + colourless fruit epidermis	*gf_y*	3.5	4	0.048
		green flesh + yellow flesh	*gf_r*	4.5	4	0.052
		green flesh + high pigment-2	*gf_hp-2*	*3*	4	0.017
		Anthocyanin fruit + atroviolaceum	*Aft_atv*	*1*	3	0.268
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.

2.2.2. DNA extraction, GBS library preparation and genotyping

Genomic DNA was isolated from young leaves samples with SpeedTools Plant DNA Extraction kit (Biotools, Spain). The GBS was performed by LGC Genomics GmbH (Germany) following the procedure reported by Elshire et al. (2011). Briefly, DNA was digested with the restriction enzyme *Ape*KI and barcoded libraries were prepared by accession and sequenced on an Illumina HiSeq 2000 platform. A total of 3 million 75-bp reads per sample were generated. To obtain variant calls in form of VCF data, the FASTQ reads were trimmed and mapped to Heinz reference genome v2.5. Freebayes SNP caller (Garrison and Marth, 2012) was used to call the SNPs on the mapped sequence reads 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, 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).

2.2.3. 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/cm2 (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 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.

2.2.4. 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 S2.3 and S2.4). 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).

2.3. Results

The collection comprised 19 lines (Table 1; Fig. 2.1.). 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_moB*), 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 2.1). With the exception of *pd* and *moB*, the gene underlying all the mutations has been identified (Supplementary Table S2.1). 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 and Pandey, 2009; Mazzucato et al., 2013).

2.3.1. 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 S2.2). 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 2.1). 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. 2.2, right).

As the number of BCs was not significantly related to the genetic distance from the recurrent parent (P=0.53; Supplementary Fig. S2.1), 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 2.1). 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. S2.2). 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 S2.2).

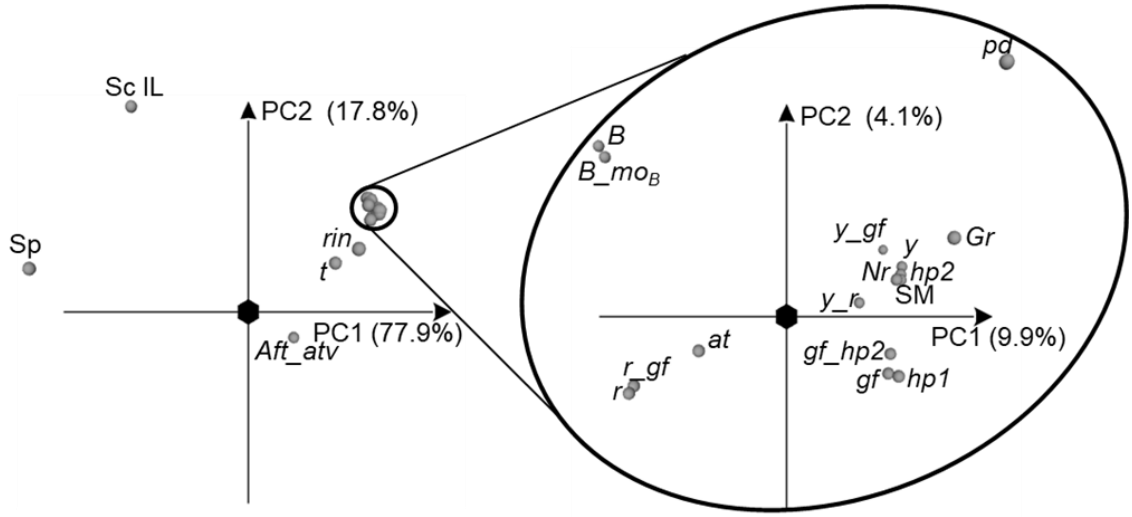


Figure 2.2. Distribution according to the first two principal components of the 21 lines studied (left) and of 16 clustered San Marzano lines (right) after GBS analysis at 1351 SNP markers (line symbols are reported in Table 2.1).

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 Fig. S2.2); 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 2.1), maintained big introgressions from parents different from AC (Supplementary Fig. S2.3A). 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. S2.3A). 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_moB*, *at*, *gf_r*), on Chr9 (*r*, *B*, *B_moB*, *y* and its double mutants) and on Chr10 (*hp-1*, *r*, *gf*, *gf_r*; Supplementary Fig. S2.3B). Describing haplotypes instead of single SNPs, 22 introgressions could be highlighted, with estimated size ranging from 0.05 to 64.74 Mbp (Supplementary Table S2.5). 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. S2.4).

2.3.2. 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. 2.3A). The G*Y interaction was significant for both CHL and FLAV (Supplementary Table S2.3); when data were mediated over years, a leaf CHL content of 32.4 µg/cm2 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. 2.3B). The average FLAV value measured in the WT was 0.69 µg/cm2.

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. 2.3C).

2.3.3. 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 2.2). However, considering single years, *B* and *hp-2* showed late flowering in 2017 and pd in 2018 (Supplementary Table S2.4). 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 2.2). 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 2.2). Interestingly, also the *hp- 1* mutant had value of PV lower than 90%, although not different from the WT (Table 2.2).

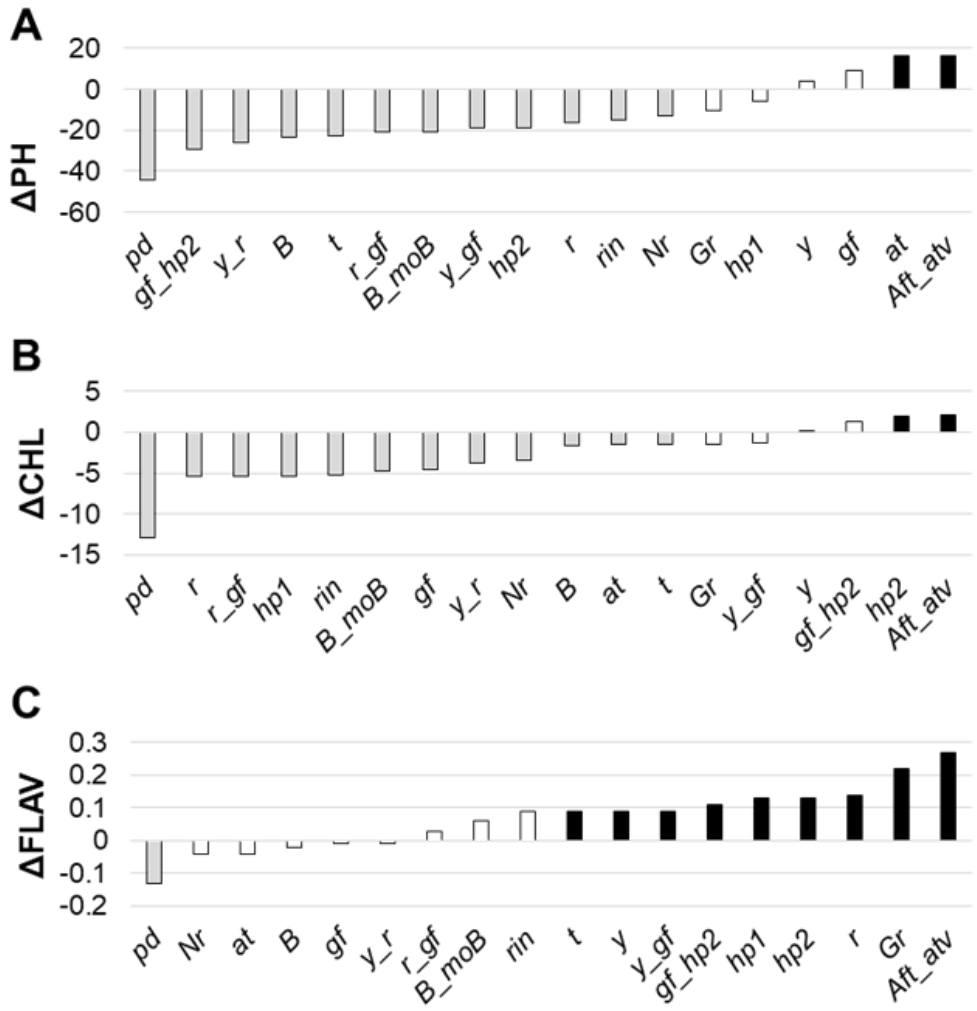


Figure 2.3. Absolute variation in (**A**) plant height (ΔPH, cm), (**B**) leaf chlorophyll (ΔCHL, µg/cm2) and (**C**) flavonoid content (ΔFLAV, µg/cm2) of 18 fruit mutant lines in San Marzano background compared with the recurrent parent. Line symbols are reported in Table 1. Bars coloured in grey and black indicate means significantly lower and higher than San Marzano for P≤0.05 after Student's *t*-test respectively.

2.3.4. Fruit traits

A significant G*Y interaction was detected also for FW and SI (Supplementary Table S2.3). 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 2.2). 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 2.2). 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 S2.3). Only *at* had a SxF significantly lower than SM and none had significantly higher values (Table 2.2).

Table 2.2. Flowering date (FLOW, d after transplant), number of flowers per inflorescence (NF), pollen viability (PV, %), fruit weight (FW, g), shape index (SI) and seeds per fruit (SxF) measured on plants of the San Marzano cv (WT) and of 18 fruit variant lines. Mean values significantly higher and lower than the WT for $P \leq 0.05$ after Student' s *t*-test are in bold and underlined 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

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 S2.3), they behaved consistently in the two years (not shown) and a PCA analysis clearly separated the 19 lines under study (Fig. 2.4). 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. 2.4). The pd mutant had an intermediate position on the upper part of the graph. Finally, the S-E quadrant also hosted the rin mutant, whose green fruits had the highest b value (Fig. 2.4).

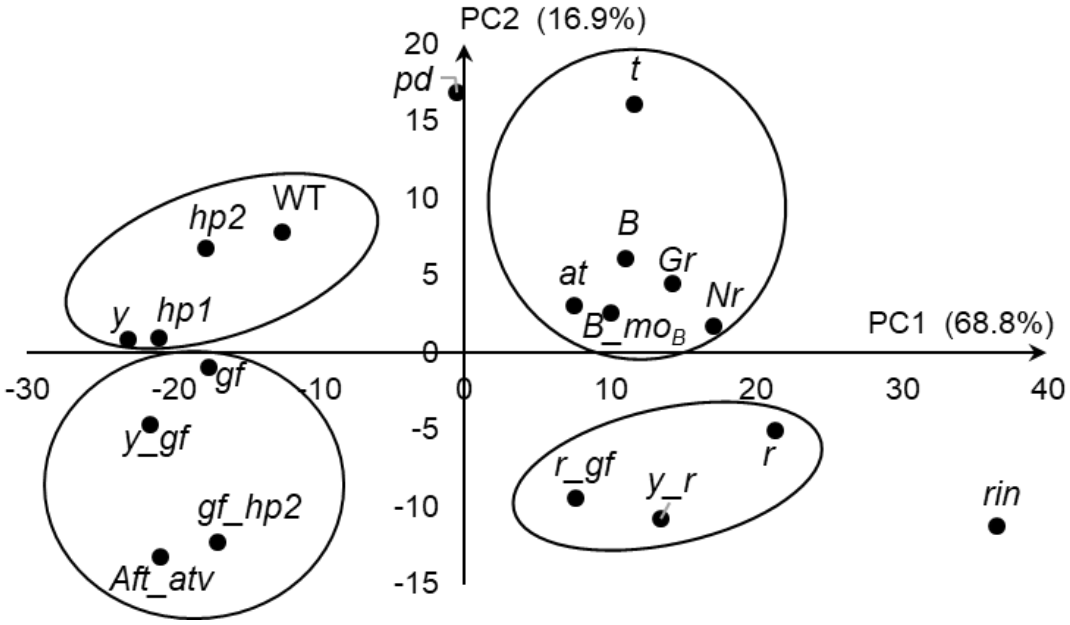


Figure 2.4. Distribution according to the first two principal components (PC) of the 19 San Marzano lines studied according to the chromameter parameters a, b and L (line symbols are reported in Table 2.1).

For Brix, the *B* line showed on average 0.5° Brix more than SM, although this difference was not significant; four lines had a value lower than SM (Fig. 2.5A). The shelf-life behaviour of the studied lines was investigated with a post-harvest experiment where FW losses and wrinkling were monitored for 20 and 40 days of storage respectively. The three ripening mutants plus *y_r* had higher SL than SM; in addition, also *y* and *y_gf* showed higher values although not significant. Four genotypes (the mutants involving

hp-2 plus *t* and *Aft_atv*) showed SL lower than SM (Fig. 2.5B). All the mutants with delayed ripening and those involving the *y* mutation also showed a resistance to wrinkling higher than the WT, whereas the high-pigment mutants, together with *B_moB*, *gf*, *pd* and *t* had lower resistance (Fig. 2.5C).

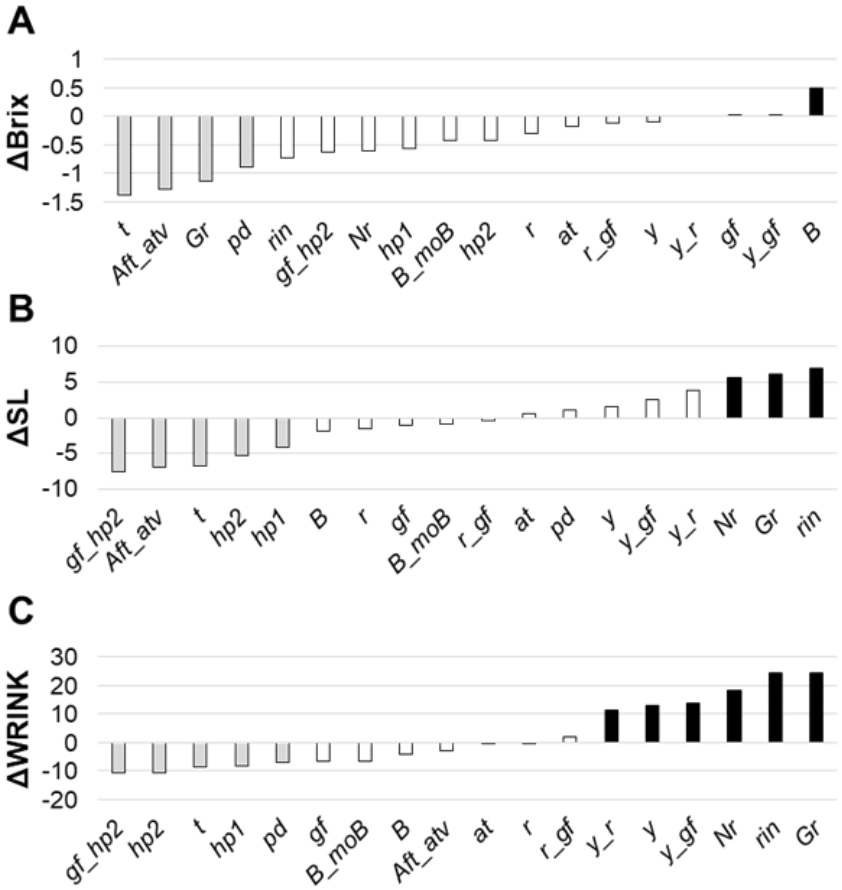


Figure 2.5. Variation in (A) soluble solids content (ΔBrix), (B) shelf-life after 20 d storage (ΔSL, % of initial FW) and (C) days to wrinkling (ΔWRINK) of 18 fruit mutant lines in San Marzano background compared with the recurrent parent. Line symbols are reported in Table 2.1. Bars coloured in grey and black indicate means significantly lower and higher than San Marzano for P≤0.05 after Student's *t*-test respectively.

2.4. Discussion

2.4.1. General features of the studied lines

This study addressed the analysis of genotypic and phenotypic variation in a repertoire of tomato lines harbouring different variants affecting fruit traits in the genetic background of the SM Italian landrace.

Each of these lines had undergone a variable number of BCs to the recurrent parent and of selfing generations in order to fix all SM typical traits, such as indeterminate habit, elongated fruit shape and strong green shoulder.

GBS analysis based on 1351 SNPs indicated that the number of BCs was not a good predictor of the genetic distance from the recurrent parent, as conspicuous regions from the donor parent remains introgressed also after several backcrosses in some genotypes. The *t* line despite being a BC4 generation, was genetically positioned rather far from the recurrent 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 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 (Casanas et al., 2017).

2.4.2. 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 (Lopez Camelo and Gomez, 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 semi-determinate habit (Minoggio et al., 2003) and the BC selection failed to fully recover the indeterminate phenotype after backcrossing.

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

2.4.3. Features of lines involving carotenoids

Three single and two double mutants involved in fruit carotenoid accumulation showed a leaf CHL content lower than the WT, whereas two had higher leaf FLAV content, indicating that the variation in one class of pigments may significantly impact the levels of other classes.

Apparently, a crosstalk among different metabolic pathways may be hypothesized and some findings seems to support this hypothesis (Minoggio et al., 2003; Pal et al., 2019). However, other evidences tend to limit the extent of the reciprocal influence between the phenylpropanoid/ flavonoid and carotenoid pathways (Long et al., 2006), thus giving an input for deeper studies.

Because the value of L* indicates the brightness and a decreasing L* value indicates the darkening of red color (Lopez Camelo and Gomez, 2004), the carotenoid mutant lines mapped in the N-E part of the PCA graph, characterized by positive and high L* values. The WT in our experiments scored a mean Brix value of 5.8, in line with values reported previously in SM (Ercolano et al., 2008; Baldina

et al., 2016). *B* was the only mutation to show a Brix value higher than the WT (6.24) although this difference was not significant. *B* was also characterized by an increased NF and by a delay in FLOW, a phenotype that was also observed in *B* introgressions in different genetic backgrounds (A. Mazzucato, 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 postharvest 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.

2.4.4. 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 (Hortensteiner, 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; Hortensteiner and Krautler, 2011). 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; Dominguez et al., 2009). Indeed, low levels of polyphenols induced by silencing of chalcone synthase (CHS) reduced the ability of the fruit to deform and decreased cuticle permeability (Espana et al., 2014). Thus, the *y* mutant phenocopies variants with delayed ripening for the resistance to wrinkling, although different underlying genetic mechanisms are responsible for this phenotype in different lines.

Among the double mutants, the *Aft_atv* combination, giving purple fruits, was the one with the highest departures from the original WT, differing for nearly all the traits that were taken into consideration. Part of this variation, e.g. that for SI, is likely due to the low level of backcrossing of this genetic combination and the genetic distance from SM

still inherent in this line. The extended SL of purple fruits is an interesting character that was reported in round-fruited backgrounds (Bassolino et al., 2013; Mazzucato et al., 2013; Borghesi et al., 2016). In SM, however, a lower resistance to dehydration was reported, in disagreement with previous data (Bassolino et al., 2013; Zhang et al., 2013). Therefore, further investigation is needed in order to assess if the better post-harvest performances of purple tomatoes can be generalized or if they are dependent on the fruit shape and, more generally, on the genetic background.

2.4.5. Features of lines with delayed ripening

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

Mutants for delayed ripening also showed the pleiotropic phenotype of an increased NF, due to the occurrence of compound inflorescences.

In the *rin* line, we observed large sepals and indeterminate inflorescences as expected because the original *rin* mutation also affects the *MACROCALYX* gene, a MADS-box transcription factor with a role in sepal size and inflorescence determinacy regulation (Vrebalov et al., 2002; Samach and Lotan, 2007). However, in the SM *rin* line the phenotype also included longer and bifurcated inflorescences that caused an increase in NF.

2.5. 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.

Bibliography

Adato, A., Mandel, T., Mintz-Oron, S., Venger, I., Levy, D., Yativ, M., Dominguez, E., Wang, Z., De Vos, R.C.H., Jetter, R., Schreiber, L., Heredia, A., Rogachev, I., Aharoni, A. Fruit-surface flavonoid accumulation in tomato is controlled by a SlMYB12- regulated transcriptional network. *PLoS Genet.* **2009**, 5 (12), e1000777. https://doi.org/10. 1371/journal.pgen.1000777.

Akhtar, M.S., Goldschmidt, E.E., John, I., Rodoni, S., Matile, P., Grierso, D. Altered patterns of senescence and ripening in gf, a stay-green mutant of tomato (*Lycopersicon esculentum* Mill.). *J. Exp. Bot.* **1999**, 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. Heatmapper: web-enabled heat mapping for all. *Nucl. Acids Res*. **2016**, 44, W147–W153. https://doi.org/10.1093/nar/gkw419. Web Server issue.

Baldina, S., Picarella, M.E., Troise, A.D., Pucci, A., Ruggieri, V., Ferracane, R., Barone, A., Fogliano, V., Mazzucato, A. Metabolite profiling of Italian tomato landraces with different fruit types. *Front. Plant Sci*. **2016**, 7, 664. https://doi.org/10.3389/fpls.2016. 00664.

Barry, C.S., Pandey, P.A. Survey of cultivated heirloom tomato varieties identifies four new mutant alleles at the *green-flesh* locus. *Mol. Breed.* **2009**, 24 (3), 269–276. https:// doi.org/10.1007/s11032-009-9289-4.

Bassolino, L., Zhang, Y., Schoonbeek, H.J., Kiferle, C., Perata, P., Martin, C. Accumulation of anthocyanins in tomato skin extends shelf life. *New Phytol.* **2013**, 200 (3), 650–655. https://doi.org/10.1093/jhered/esg093. http://doi:10.1111/nph.12524.

Bino, R.J., De Vos, C.H., Lieberman, M., Hall, R.D., Bovy, A., Jonker, H.H., Tikunov, Y., Lommen, A., Moco, S., Levin, I. The light-hyperresponsive *high pigment-2dg* mutation of tomato: alterations in the fruit metabolome. *New Phytol.* **2005**, 166, 427–438. https://doi.org/10.1111/j.1469-8137.2005.01362.

Borghesi, E., Ferrante, A., Gordillo, B., Rodriguez-Pulido, F.J., Cocetta, G., Trivellini, A., Mensuali-Sodi, A., Malorgio, F., Heredia, F.J. Comparative physiology during ripening in tomato rich-anthocyanins fruits. *Plant Growth Regul.* **2016**, 80 (2), 207–214. https://doi.org/10.1007/s10725-016-0158-y.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. **2007**, 23, 2633–2635.

Carvalho, R.F., Campos, M.L., Pino, L.E., Crestana, S.L., Zsogon, A., Lima, J.E., Vagner, A.B., Peres, L.E. Convergence of developmental mutants into a single tomato model system: Micro-Tom' as an effective toolkit for plant development research. *Plant Methods*, **2011**, *7*, 18. https://doi.org/10.1186/1746-4811-7-18.

Casanas, F., Simo, J., Casals, J., Prohens, J. Toward an evolved concept of landrace. *Front. Plant Sci*. **2017**, 8, 145 http://doi: 10.3389/fpls.2017.00145.

Cerovic, Z.G., Masdoumier, G., Ghozlen, N.B., Latouche, G. A new optical leaf-clip meter for simultaneous non-destructive assessment of leaf chlorophyll and epidermal flavonoids. *Physiol. Plant.* **2012**, 146, 251–260. http://doi:10.1111/j.1399-3054.2012. 01639.x.

Dahan-Meir, T., Filler-Hayut, S., Melamed-Bessudo, C., Bocobza, S., Czosnek, H., Aharoni, A., Levy, A.A. Efficient in planta gene targeting in tomato using geminiviral replicons and the CRISPR/Cas9 system. *Plant J.* **2018**, 95, 5–16 http://doi: 10.1111/ tpj.13932.

Darby, L.A. Isogenic lines of tomato fruit colour mutants. *Hortic. Res.* **1978**, 18, 73–84.

Deng, L., Wang, H., Sun, C., Li, Q., Jiang, H., Du, M., Li, C.B., Li, C. Efficient generation of pink-fruited tomatoes using CRISPR/Cas9 system. *J. Genet. Genom.* **2018**, 45, 51–54. https://doi.org/10.1016/j.jgg.2017.10.002. Epub 2017 Nov 6.

Dominguez, E., Lopez-Casado, G., Cuartero, J., Ramirez, L.E. Development of fruit cuticle in cherry tomato (*Solanum lycopersicum*). *Funct. Plant Biol.* **2009**, 36, 613–620.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E. A robust simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, **2011**, 6 (5), e19379. https://doi.org/10.1371/journal.pone. 0019379.

Ercolano, M.R., Carli, P., Soria, A., Cascone, A., Fogliano, V., Frusciante, L., Barone, A. Biochemical, sensorial and genomic profiling of traditional Italian tomato varieties. *Euphytica*, **2008**, 164 (2), 571–582.

Espana, L., Heredia-Guerrero, J.A., Reina-Pinto, J.J., Fernandez-Munoz, R., Heredia, A., Dominguez, E. Transient silencing of CHALCONE SYNTHASE during fruit ripening modifies tomato epidermal cells and cuticle properties. *Plant Physiol.* **2014**, 166 (3), 1371–1386. https://doi.org/10.1104/pp.114.246405.

Foolad, M.R. Genome mapping and molecular breeding of tomato. *Intl. J. Plant Genom*. **2007**, 64358. https://doi.org/10.1155/2007/64358. Article ID.

Garrison, E., Marth, G. Haplotype-based Variant Detection From Short-read Sequencing. *arXiv preprint arXiv*, **2012**, 1207.3907. .

Hortensteiner, S., Krautler, B. Chlorophyll breakdown in higher plants. *Biochim. Biophys. Acta*, **2011**, 1807 (8), 977–988 http://doi: 10.1016/j.bbabio.2010.12.007.

Jacobs, T.B., Zhang, N., Patel, D., Martin, G.B. Generation of a collection of mutant tomato lines using pooled CRISPR libraries. *Plant Physiol*. **2017**, 174 (4), 2023–2037 http://doi: 10.1104/pp.17.00489.

Levin, I., De Vos, C.R., Tadmor, Y., Bovy, A., Lieberman, M., Oren-Shamir, M., Segev, O., Kolotilin, I., Keller, M., Ovadia, R., Meir, A., Bino, R.J. High pigment tomato mutants - more than just lycopene (a review). *Isr. J. Plant Sci.* **2006**, 54 (3), 179–190.

Lieberman, M., Sege, O., Gilboa, N., Lalazar, A., Levin, I. The tomato homolog of the gene encoding uv-damaged dna binding protein 1 (ddb1) underlined as the gene that causes the *high pigment-1* mutant. *Theor. Appl. Genet.* **2004**, 108, 1574–1581. https:// doi.org/10.1007/s00122-004-1584-1.
Liu, K., Muse, S.V. PowerMarker: integrated analysis environment for genetic marker data. *Bioinformatics* **2005**, 21, 2128–2129. https://doi.org/10.1093/ bioinformatics/bti282.

Long, M., Millar, D.J., Kimura, Y., Donovan, G., Rees, J., Fraser, P.D., Bramley, P.M., Bolwell, G.P. Metabolite profiling of carotenoid and phenolic pathways in mutant and transgenic lines of tomato: identification of a high antioxidant fruit line. *Phytochemistry*, **2006**, 67, 1750–1757. https://doi.org/10.1016/j.phytochem.2006.02. 022.

Lopez Camelo, A.F., Gomez, P.A. Comparison of color indexes for tomato ripening. *Hort. Bras. Brasilia*. **2004**, 22 (3), 534–537.

Mazzucato, A., Willems, D., Bernini, R., Picarella, M.E., Santangelo, E., Ruiu, F., Tilesi, F., Soressi, G.P. Novel phenotypes related to the breeding of purple-fruited tomatoes and effect of peel extracts on human cancer cell proliferation. *Plant Physiol. Biochem*. **2013**, 72, 125–133 http://doi: 10.1016/j.plaphy.2013.05.012.

Minoggio, M., Bramati, L., Simonetti, P., Gardana, C., Iemoli, L., Santangelo, E., Mauri, P.L., Spigno, P., Soressi, G.P., Pietta, P.G. Polyphenol pattern and antioxidant activity of different tomato lines and cultivars. *Ann. Nutr. Metabol.* **2003**, 47 (2), 64–69. https://doi.org/10.1038/385718a0.

Moore, S., Vrebalov, J., Payton, P., Giovannoni, J. Use of genomics tools to isolate key ripening genes and analyse fruit maturation in tomato. *J. Exp. Bot.* **2002**, 53 (377), 2023–2030. https://doi.org/10.1093/jxb/erf057.

Mustilli, A.C., Fenzi, F., Ciliento, R., Alfano, F., Bowler, C. Phenotype of the tomato high pigment-2 mutant is caused by a mutation in the tomato homolog of DEETIOLATED1. *Plant Cell*, **1999**, 11 (2), 145–157 doi.org/10.1105/tpc.11.2.145.

Nei, M., Tajima, F., Tateno, Y. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J. Mol. Evol*. **1983**, 19, 153–170.

Pal, H., kundu, A., Sahu, R., Sethi, A., Hazra, P., Chatterjee, S. Unraveling the metabolic behavior in tomato high pigment mutants (*hp-1*, *hp-2*dg, *ogc*) and non ripening mutant (*rin*) during fruit ripening. *Sci. Hort.* **2019**, 246, 652–663.

Samach, A., Lotan, H. The transition to flowering in tomato. *Plant Biotechnol.* **2007**, 24 (1), 71–82.

Schroeder, D.F., Gahrtz, M., Maxwell, B.B., Cook, R.K., Kan, J.M., Alonso, J.M., Ecker, J.R., Chory, J. De-etiolated 1 and damaged DNA binding protein 1 interact to regulate Arabidopsis photomorphogenesis. *Curr. Biol*. **2002**, 12 (17), 1462–1472.

Soressi, G.P. New spontaneous or chemically-induced fruit ripening tomato mutants. *Rep. Tomato Genet. Coop.* **1975**, 25, 21–22.

Stam, P., Zeven, A.C. The theoretical proportion of the donor genome in nearisogenic lines of self-fertilizers bred by backcrossing. *Euphytica*, **1981**, 30, 227–238.

Tigchelaar, E.C., Tomes, M.L., Erickson, H.T., Graham, T.O., Barman, R.J. Pigment diluter" (*pd*), a new plant and fruit color mutant. *Rep. Tomato Genet. Coop*. **1970** , 20, 64.
Tomes, M.L. Flower color modification associated with the gene *t*. *Tom. Genet. Coop*. **1952**, 2, 12.

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

Vrebalov, J., Ruezinsky, D., Padmanabhan, V., White, R., Medrano, D., Drake, R., Schuch, W., Giovannoni, J.A. MADS-box gene necessary for fruit ripening at the tomato *ripening-inhibitor* (*rin*) locus. *Science*, **2002**, 296 (5566), 343–346. https://doi.org/10. 1126/science.1068181V.

Yen, H.C., Shelton, B.A., Howard, L.R., Lee, S., Vrebalov, J., Giovannoni, J.J. The tomato *high-pigment* (*hp*) locus maps to chromosome 2 and influences plastome copy number and fruit quality. *Theor. Appl. Genet*. **1997**, 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. CRISPR/Cas9-induced targeted mutagenesis and gene replacement to generate long-shelf life tomato lines. *Sci. Rep*. **2017**, 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. Anthocyanins double the shelf life of tomatoes by delaying overripening and reducing susceptibility to gray mold. *Curr. Biol*. **2013**, 23, 1094–1100 http://doi: 10.1016/j.cub.2013.04.072.

Supplementary Tables

Supplementary Table S2.1. 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.

Class of variation	Name	Genetic symbol	First description	Mapped/Cloned by	Chr.	Gene code	Type of variation[b]
All pigments	*high pigment-1*	*hp-1*	Reynard 1956	Lieberman et al. 2004	2	Solyc02g021650	Single base transversion in coding
	high pigment-2	*hp-2*	Soressi 1975	Mustilli et al. 1999	1	Solyc01g056340	Intronic SNP causing alternative splicing
	pigment diluter	*pd*	Tigchelaar et al. 1970	-[a]	-	-	-
Carotenoids	*yellow flesh*	*r*	Price and Drinkard 1908	Fray and Grierson 1993	3	Solyc03g031860	Large deletion in the coding
	tangerine	*t*	Tomes 1952	Isaacson et al. 2002	10	Solyc10g081650	348-bp deletion in the promoter
	apricot	*at*	Jenkins and Mackinney 1955	Pankratov et al. 2016	4	Solyc04g056390	Exonic SNP causing stop codon
	High Beta	*B*	Lincoln and Porter 1950	Ronen et al. 2000	6	Solyc06g074240	Rearrangements in the promoter region
	B-modifier	mo_B	Tomes et al. 1954	Zhang and Stommel 2000	6	-	-
Chlorophyll	*green flesh*	*gf*	Kerr 1956	Barry et al. 2008	8	Solyc08g080090	Missense transversion
Flavonoids	*colourless fruit epidermis*	*y*	Lindstrom 1925	Ballester et al. 2010 Lin et al. 2014	1	Solyc01g079620	603-bp deletion in the promoter
	Anthocyanin fruit	*Aft*	Georgiev 1972	Boches and Myers 2007 Schreiber et al. 2011	10	Solyc10g086250 Solyc10g086260	-
	atroviolaceum	*atv*	Rick et al. 1968	Colanero et al. 2018	7	Solyc07g052490	4-bp insertion leading to a stop codon
Ripening	*Never ripe*	*Nr*	Rick and Butler 1956	Wilkinson et al. 1995	9	Solyc09g075440	Missense mutation in sensor domain
	ripening inhibitor	*rin*	Tigchelaar 1978	Vrebalov et al. 2002	5	Solyc05g012020	Large deletion involving two MADS-box genes
	Green ripe	*Gr*	Kerr 1958	Barry and Giovannoni 2006	1	Solyc01g104340	334-bp deletion involving promoter and 1st exon

[a] not determined

[b] refers to the originally described allele, that was used in the introgressions described in this work.

References (only used in Table S1)

Ballester, A.R., Molthoff, J., de Vos, R., Hekkert, B.L., Orzaez, D., Fernández-Moreno, J.P., Tripodi, P., Grandillo, S., Martin, C., Heldens, J., Ykema, M., Granell A., Bovy A., 2010. Biochemical and Molecular Analysis of Pink Tomatoes: Deregulated Expression of the Gene Encoding Transcription Factor SlMYB12 Leads to Pink Tomato Fruit Color. Plant Physiology. 152, 71-84. https://doi.org/10.1104/pp.109.147322.

Barry, C.S. and Giovannoni, J. J., 2006. Ripening in the tomato *Green-ripe* mutant is inhibited by ectopic expression of a protein that disrupts ethylene signaling. Proceedings of the National Academy of Sciences. 103, 7923-7928.

Barry, C.S., McQuinn, R.P., Chung, M.Y., Besuden, A., Giovannoni J.J., 2008. Amino Acid Substitutions in Homologs of the STAY-GREEN Protein Are Responsible for the *green-flesh* and *chlorophyll retainer* mutations of Tomato and Pepper. Plant Physiology. 147, 179–187.

Boches, P., Myers, J., 2007. The anthocyanin fruit tomato gene (*Aft*) is associated with a DNA polymorphism in a MYB transcription factor. Hort Science. 42: 856.

Colanero, S., Perata, P., Gonzali, S., 2018. The atroviolacea Gene Encodes an R3-MYB Protein Repressing Anthocyanin Synthesis in Tomato Plants. Front. Plant Sci. 9:830. doi: 10.3389/fpls.2018.00830.

Fray, R. and Grierson, D. 1993. Identification and genetic analysis of normal and mutant phytoene synthase genes of tomato by sequencing, complementation and co-suppr. Plant Mol Biol. 22, 589-602.

Georgiev, C., 1972. Anthocyanin fruit (*Af*). Rep Tomato Genet Coop. 22: 10.

Isaacson, T., Ronen, G., Zamir, D. and Hirschberg, J. 2002. Cloning of *tangerine* from Tomato Reveals a Carotenoid Isomerase Essential for the Production of β-Carotene and Xanthophylls in Plants. Plant Cell. 14, 333-342.

Jenkins, J.A. and Mackinney, G. 1955. Carotenoids of the apricot tomato and its hybrids with yellow and tangerine. Genetics. 40, 715–720.

Kerr, E.A., 1956. Green flesh, gf. Rpt. Tomato Genet Coop. 6: 17.

Kerr, E.A., 1958. Mutations for chlorophyll retention in ripe fruit. Rpt Tomato Genet Coop. 8: 22.

Lin, T., Zhu, G., Zhang, J., Xu, X., Yu, Q., Zheng, Z., Zhang, Z. et al., 2014. Genomic analyses provide insights into the history of tomato breeding. Nature Genetics. 46, 1220–1226.

Lincoln, R.E. and Porter, J.W., 1950. Inheritance of β-carotene in tomatoes. Genetics. 35:206-211.

Lindstrom, E.W., 1925. Inheritance in Tomatoes. Genetics. 10(4), 305-17.

Pankratov, I. , McQuinn, R. , Schwartz, J. , Bar, E. , Fei, Z. , Lewinsohn, E. , Zamir, D. , Giovannoni, J. J. and Hirschberg, J., 2016. Fruit carotenoid-deficient mutants in tomato reveal a function of the plastidial isopentenyl diphosphate isomerase (IDI1) in carotenoid biosynthesis. Plant J. 88: 82-94. doi:10.1111/tpj.13232.

Price, H. L. and Drinkard, A. W., 1908. Inheritance in tomato hybrids. Virg. Agric. Exp. Stn. Bull. 177: 17–53.

Reynard, G.B., 1956. Origin of Webb Special (Black Queen) in tomato. Rep Tomato Genet Coop. 40:44–64.

Rick, C. and Butler, L., 1956. Cytogenetics of the tomato. Advan. Genet. 8, 267-382.

Rick, C. M., Reeves, A. F. and Zobel, R. W., 1968. Inheritance and linkage relations of four new mutants. Rep. Tomato Genet. Coop. 18, 34–35.

Ronen, G., Carmel-Goren, L., Zamir, D. and Hirschberg, J., 2000. An alternative pathway to β-carotene formation in plant chromoplasts discovered by map-based cloning of *beta* and *old-gold* color mutations in tomato. Proceedings of the National Academy of Sciences of the United States of America. 97, 11102-11107.

Schreiber, G., Reuveni, M., Evenor, D., Oren-Shamir, M., Ovadia, R., Sapir-Mir, R. et al., 2011. ANTHOCYANIN1 from Solanum chilense is more efficient in accumulating anthocyanin metabolites than its *Solanum lycopersicum* counterpart in association with the ANTHOCYANIN FRUIT phenotype of tomato. Theor Appl Genet. 124, 295–308.
Tigchelaar, E.C. Tomato ripening mutants. HortScience, 1978, 13: 502Zhang, Y. and Stommel, J.R., 2001. Development of SCAR and CAPS Markers Linked to the Beta Gene in Tomato. Crop Science. 41(5), 1602-1608.
Tomes, M.L., Quackenbush, F.W., McQuistan, M., 1954. Modification and dominance of the gene governing formation of high concentrations of β-carotene in the tomato. Genetics. 39:810-817.
Wilkinson, J.Q., Lanahan, M.B., Yen, H.C., Giovannoni, J.J., Klee, H.J., 1995. An ethylene inducible component of signal transduction encoded by Never-ripe. Science. 270, 1807–1809.

Supplementary Table S2.2. 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.

Type of filtering	Total No. of SNPs	No. of SNPs on chromosome												
		0[a]	1	2	3	4	5	6	7	8	9	10	11	12
Missing data < 30% Minimum allele frequency > 0.06	1351	23	39	155	64	118	443	202	78	21	35	67	73	33
Polymporphic within SM lines (*Aft_atv* excluded)	539	17	20	131	34	30	19	188	10	7	15	47	9	12
Polymorphic between SM and AC	129	8	1	2	29	5	5	41	4	3	9	16	3	3

[a] Sites not assigned to a specific chromosome.

Supplementary Table S2.3. 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.

Source of variation	PH	CHL	FLAV	FLOW	NF	PV	FW	SI	Brix	SxF	WRINK	SL	a	b	L
Genotype	15.9***	56.3***	12.5***	3.9***	3.9***	3.2***	10.7***	13.3***	4.3***	2.7**	31.9***	13.4***	82.9***	119.6***	97.9***
Year	254.1***	1897.7***	128.6***	2.2	0.6	2.3	42.6***	4.6**	113.2***	65.1***	26.8***	98.8***	13.3**	56.5***	21.9***
G*Y	2.6**	8.8***	5.2***	2.1*	1.9*	1.1	2.3 **	6.5***	1.1	1.6	4.1***	6.0***	9.2***	10.9***	5.2***

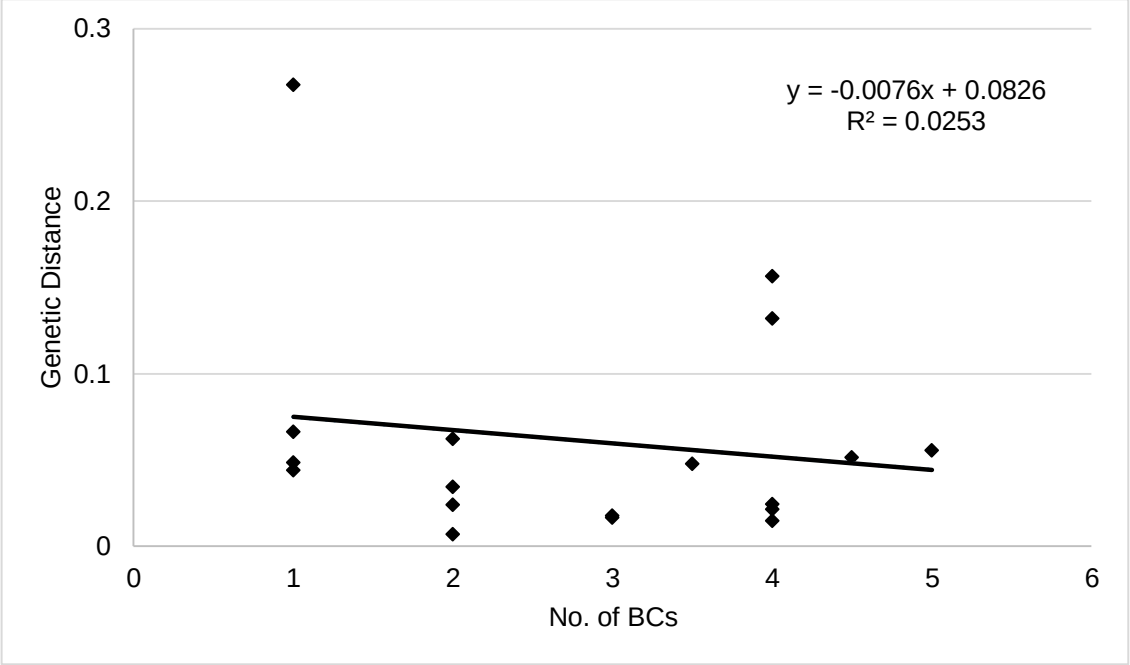
Supplementary Table S2.4. Mean values within levels of the main factor "Year" (1 and 2) for traits showing significant Genotype*Year interaction. Values are reported for plant height (PH, cm), chlorophyll (CHL, µg/cm^2) and flavonoid (FLAV, µg/cm^2) content, flowering date (FLOW, days from transplant), 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.

Geno-type	PH		CHL		FLAV		FLOW		NF		FW		SI		SL		WRINK		a		b		L	
	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
WT	116	85	38	27	0.5	0.7	31	30	7.5	6.8	52	69	1.8	2.0	89	84	17	15	28	27	26	25	39	39
hp-1	115	75	31	23	**0.8**	0.8	30	32	6.8	**14.3**	**77**	69	1.9	1.5	85	77	10	5	25	24	19	20	35	35
hp-2	91	73	41	28	**0.7**	0.9	**35**	32	7.3	7.3	35	53	-	2.2	-	84	-	5	28	**28**	23	21	**37**	**36**
pd	65	48	21	18	0.5	0.5	32	**40**	10.5	11.8	65	61	1.6	1.7	89	84	11	7	**32**	**28**	**38**	**30**	**46**	**42**
r	91	79	31	23	**0.8**	0.8	28	27	9.5	7.8	54	61	1.7	1.7	83	82	25	8	13	6	**40**	**43**	**47**	**50**
t	93	63	34	28	**0.7**	0.8	28	31	5.5	8.3	55	41	1.8	2.2	82	77	10	5	27	26	**44**	**38**	**49**	**46**
at	133	**102**	35	27	0.5	0.7	33	33	**11.0**	**10.3**	28	33	1.5	1.4	**92**	80	21	12	20	17	**36**	**35**	**45**	**44**
B	95	60	35	26	0.6	0.7	**35**	31	**11.5**	8.3	42	57	1.6	1.6	77	**87**	7	15	-	19	45	**45**	-	**51**
B_mo$_B$	98	62	31	24	**0.7**	0.7	32	35	9.8	8.5	58	50	1.9	1.8	97	79	13	5	20	16	**39**	**34**	**46**	**44**
gf	130	90	30	25	**0.7**	0.6	29	32	10.0	**13.0**	62	79	2.0	1.7	87	82	14	5	22	23	18	22	35	37
y	119	91	39	26	0.6	**0.9**	34	31	8.5	8.0	60	79	**2.1**	1.9	89	85	**36**	21	28	25	15	14	38	**41**
Nr	97	81	35	25	0.5	0.7	31	28	10.6	8.5	36	57	**2.2**	1.7	68	83	**40**	**28**	11	19	**49**	**29**	**55**	**42**
rin	100	76	35	21	**0.6**	0.8	30	29	**17.3**	11.8	24	54	1.9	2.0	**96**	**89**	**40**	**40**	2	1	**48**	**46**	**55**	**54**
Gr	118	55	36	25	**0.6**	**1.2**	32	32	**12.0**	8.7	67	75	1.8	1.6	**93**	86	**40**	**40**	13	21	**39**	**39**	**49**	**46**
r_y	90	59	35	23	0.6	0.7	30	31	9.0	8.3	36	61	1.7	1.6	**93**	85	**29**	**25**	16	3	**45**	20	**52**	**46**
gf_y	95	68	37	25	**0.6**	0.9	28	30	8.5	7.5	43	72	1.8	1.8	90	86	**36**	23	25	19	15	15	39	38
gf_r	98	62	32	22	**0.7**	0.7	31	30	8.3	**9.3**	55	64	1.7	1.8	79	88	8	24	-	3	-	**34**	-	**44**
gf_hp-2	86	56	39	28	**0.7**	0.9	31	36	6.8	5.8	37	51	-	1.9	-	84	-	5	17	13	19	17	35	34
Aft_atv	**148**	87	40	29	**0.8**	**1.1**	29	32	9.3	**9.3**	58	75	1.9	**2.4**	84	76	**21**	9	12	18	15	15	34	32

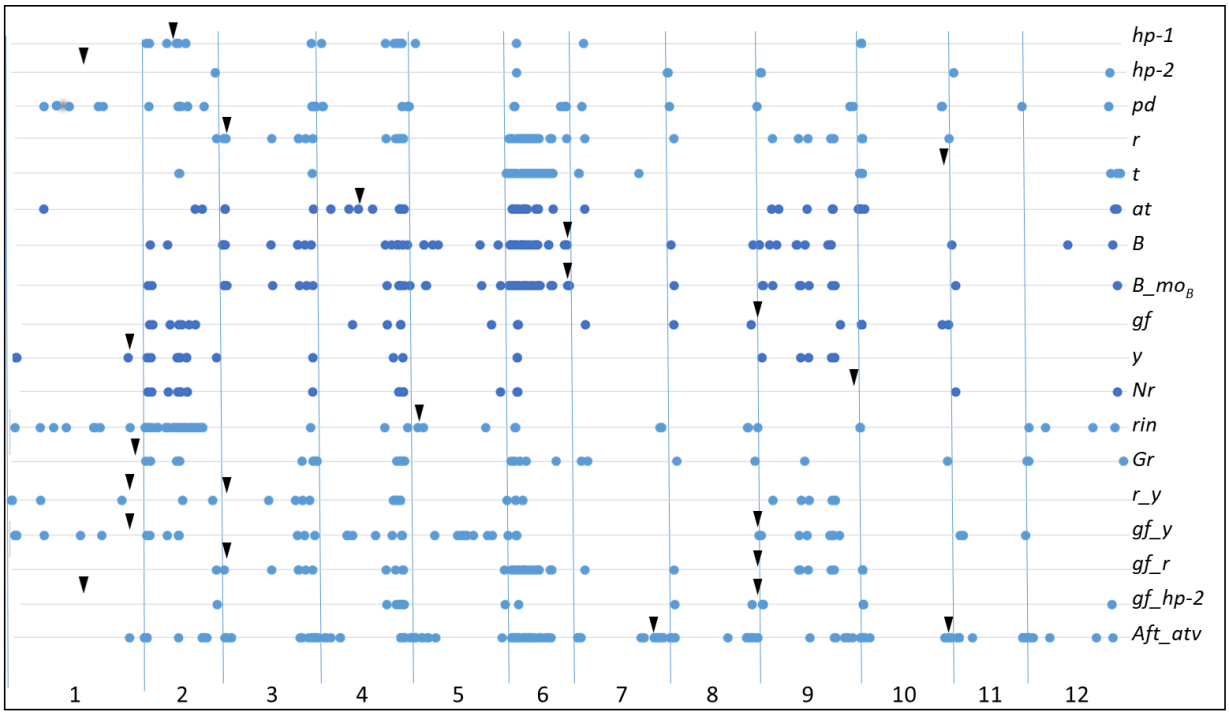
Supplementary Table S2.5. Chromosome, physical position and size of the 22 haplotypes detected and introgression lines harbouring each haplotype.

Haplotype	Chromosome	Physical position		Size	Lines
		Start	End		
Hap-1	01	22,494,789	66,634,654	44,139,865	*pd*, *rin*
Hap-2	01	3,781,812	21,990,401	18,208,589	*y*,*r_y*, *gf_y*
Hap-3	01	3,781,812	31,870,976	28,089,164	*y*,*r_y*, *gf_y*
Hap-4	02	592,814	41,603,179	41,010,365	*rin*
Hap-5	03	66,797,203	67,100,639	303,436	*pd*, *Gr*
Hap-6	03	66,008,766	67,161,958	1,153,192	*pd*
Hap-7	03	467,766	64,637,626	64,169,860	*r*, *B*, *B_mo$_B$*, *gf_r*
Hap-8	04	54,665,071	58,633,525	3,968,454	*hp-1*, *r*, *at*, *gf*, *Nr*, *Gr*, *r_y*, *gf_y*, *gf_r*, *gf_hp-2*
Hap-9	04	62,535,036	63,541,144	1,006,108	*pd*
Hap-10	04	18,216,956	55,414,200	37,197,244	*gf_y*
Hap-11	05	7,739,580	14,804,190	7,064,610	*B*, *B_mo$_B$*
Hap-12	05	15,417,226	56,957,323	41,540,097	*gf_y*
Hap-13	06	7,968,552	33,175,999	25,207,447	*r*, *B*, *B_mo$_B$*, *t*, *at*, *gf_r*
Hap-14	06	772,695	48,839,495	48,066,800	*t*
Hap-15	07	3,332,468	48,838,310	45,505,842	*t*
Hap-16	07	64,006,580	66,971,471	2,964,891	*rin*
Hap-17	08	240,661	64,982,331	64,741,670	*rin*
Hap-18	09	12,061,417	58,822,621	46,761,204	*r*, *B*, *B_mo$_B$*, *r_y*, *gf_r*
Hap-19	10	1,642,162	1,690,069	47,907	*hp-1*, *r*, *t*, *gf*, *rin*, *gf_r*
Hap-20	10	184,523	1,973,846	1,789,323	*t*
Hap-21	11	51,516,133	52,533,690	1,017,557	*Gr*
Hap-22	12	62,435,256	67,008,521	4,573,265	*t*

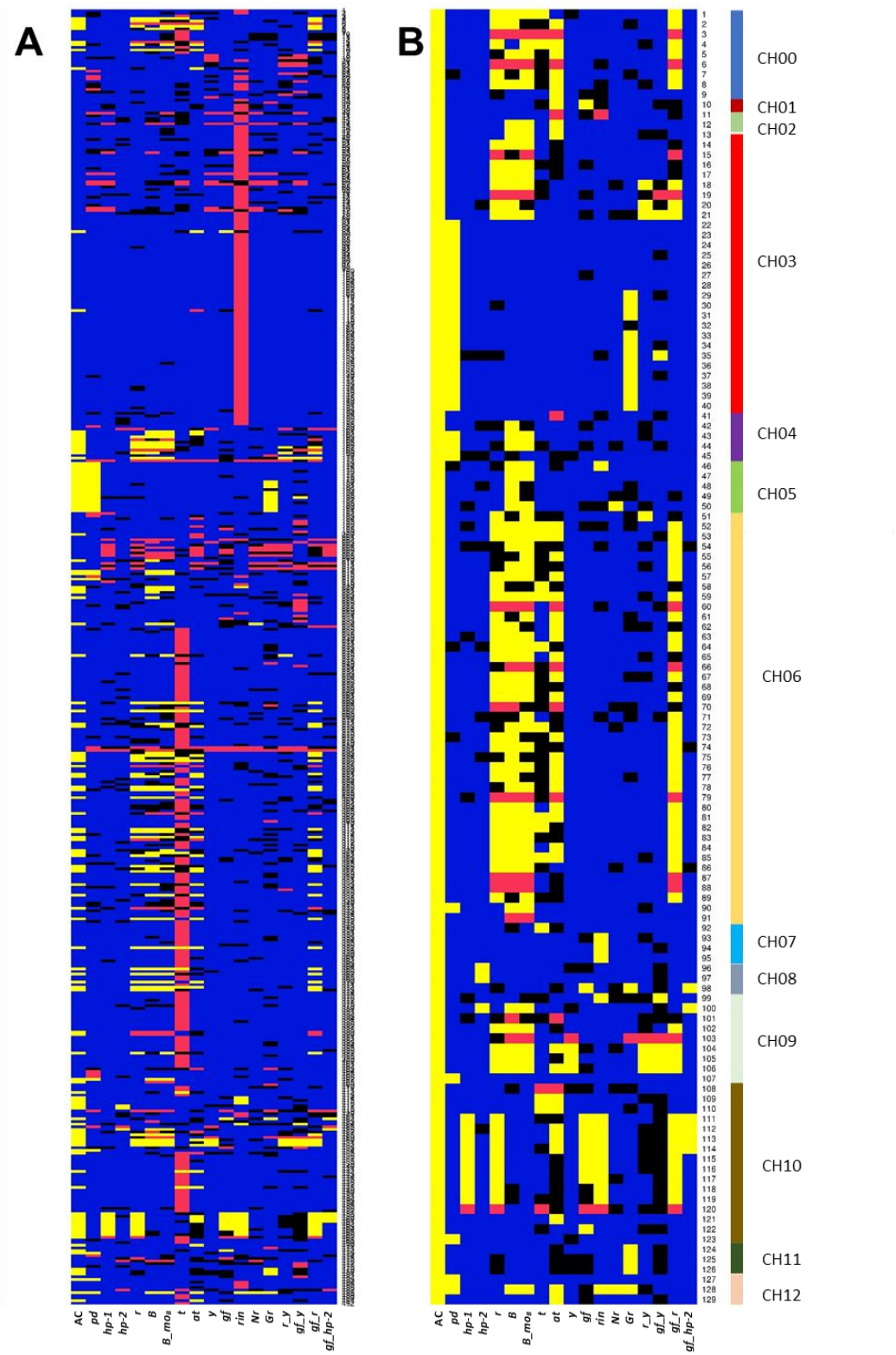
Supplementary Figures



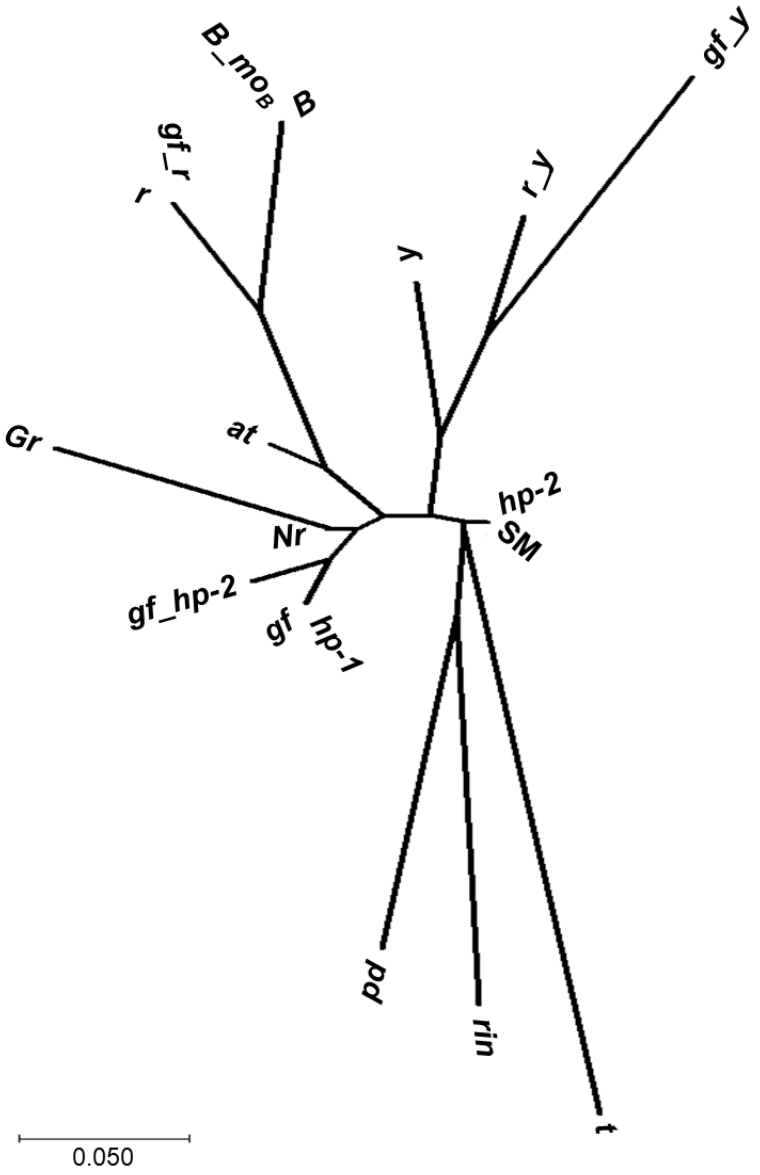
Supplementary Figure S2.1. Linear regression between the number of backcrosses (BCs) carried out for each line and the genetic distance from the San Marzano reference recurrent parent estimated by GBS analysis.



Supplementary Figure S2.2. Introgressions from the donor parent estimated in the 18 studied lines after GBS analysis at 1351 SNP markers. Blue dots 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 2.1 and Supplementary Table S2.1).



Supplementary Figure S2.3. 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 2.1.



Supplementary Figure S2.4. 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 2.1.

Chapter 3.

Color mutations alter the biochemical composition in the San Marzano tomato fruit

Abstract - San Marzano (SM) is a traditional Italian tomato landrace characterized by red elongated fruits, originated in the province of Naples (Italy) and cultivated worldwide. A preliminary metabolic analysis of the entire San Marzano collection provided details on the metabolic content of the lines, directing the analyzes towards a more specific subgroup of mutants. Herewith, three mutations, *yellow flesh* (*r*), *green flesh* (*gf*) and *colorless fruit epidermis* (*y*) and their double combinations, plus the SM reference were analysed for volatile (VOC), non-polar (NP) and polar (P) metabolites. Sixty-eight VOCs were identified, and several differences evidenced in the lines; overall *gf* showed epistasis over *r* and *y* and *r* over *y*. Analysis of the NP component identified 54 metabolites; variation in early carotenoids (up to lycopene) and chlorophylls characterized respectively the lines containing *r* and *gf*. In addition, compounds belonging to the quinone and xanthophyll classes were present in genotypes carrying the *r* mutation at levels higher than SM. Finally, the analysis of 125 P metabolites evidenced different levels of vitamins, amino acids, lipids and phenylpropanoids. Correlation network approach was used to investigate metabolite-metabolite relationships in the mutant lines. Altogether these differences potentially modified the hedonistic and nutritional value of the berry. In summary, single and combined mutations in *gf*, *r* and *y* generated interesting visual and compositional diversity in the SM landrace, while maintaining its original typology.

Results reported in this chapter as a focus on seven selected lines have been published in the journal *Metabolites* (Dono et al., 2020).

3.1. Introduction

In tomato (*Solanum lycopersicum* L.), fruit color is one of the most important traits affecting consumer liking, and it is the result of the combined effect of carotenoids, flavonoids and eventually chlorophylls. The red color of ripe fruit comes from the accumulation of all-trans-lycopene; mutants affected in the carotenoid pathway have an altered carotenoid composition, resulting in different fruit colors (Lewinsohn et al., 2005; Gascuel et al., 2017). Besides carotenoids, flavonoids play a role in determining the color of tomato fruit, particularly at the epidermal level (Bovy et al., 2010). Lastly, chlorophylls can eventually have a role in defining the color of the fruit.

San Marzano is considered an important model for fruit quality parameters, since it revealed peculiar sensory profiles (Ercolano et al., 2008; D'Esposito et al., 2017). The entire collection of mutants has undergone a wide metabolic analyses, including volatile compounds and both the non polar and polar metabolite.

Then the work focused on three tomato mutants, representative of the main classes of color pigments, to study the effects of color mutations at the metabolite level.

Yellow-fruited tomatoes have been documented since the first introduction in Europe (Hunt et al., 1980; Jenkins et al., 1955). The yellow color is due to the *yellow flesh* (*r*) mutation, represented by loss-of-function alleles of *Phytoene synthase 1* (*Psy-1*); PSY-1 catalyzes the first rate-limiting step in the carotenoid pathway, the condensation of two molecules of geranylgeranyl diphosphate in phytoene (Rodríguez-Villalón et al., 2009). Loss-of-function mutations at the *Psy-1* locus result in the inhibition of the whole carotenoid biosynthesis (Kachanovsky et al., 2012; Rocha et al., 2013). Recently, *r* tomatoes have been meeting with an increasing success, for the color novelty and the peculiar organoleptic qualities.

The *colorless fruit epidermis* (*y*) fruit mutant was originally described as a monogenic recessive variant leading to the formation of a colorless fruit peel (Lindstrom et al., 1925). The mutation, mapped on the short arm of Chr1 Rick and Butler 1956), involves the *SlMYB12* transcription factor, causing the lack of naringenin chalcone, one of the major flavonoids in tomato fruit peel, which gives the yellow color and has been proposed to influence the characteristics and function of the cuticular layer (Adato et al., 2009; Ballester

et al., 2010). The pink *y*-type fruit mutation has been identified in numerous cultivated varieties that are highly consumed in Asian countries.

Fruits of the *green flesh* (*gf*) tomato mutant were described for their characteristic muddy brown color, resulting from the accumulation of lycopene coupled with the heterochronic presence of chlorophyll in the ripen fruit due to a lack of chlorophyll degradation (Kerr 1956). In *S. lycopersicum*, the *Gf locus* maps on the long arm of Chr8. Further studies indicated that *gf* is a member of the SGR gene family, *SlSGR1*, a Mg-decheletase gene needed for chlorophyll catabolism (Barry et al., 2008; Barry et al., 2009). Nowadays, many cultivated tomatoes, heirloom varieties but also modern hybrids, exhibiting the *gf* phenotype are commercially available; indeed, these varieties are appreciated and have been reported to be superior for taste-related compounds (Bortolotti et al., 2003; Cocaliadis et al., 2014).

In this work, we analyzed the effect of introducing mutations affectingdifferent aspects of the fruit development in the metabolite complement of the SM fruit. Namely we compared with the ripe fruit of the SM genetic background a comprehensive set of fruit volatile compounds (VOCs), including those involved in flavor, and both non-polar (NP) and polar (P) specialized metabolites, including those involved in health properties, firstly in the entire SM collection, described in Chapter 2 and finally focusing on a specific core of the mutant collection, comprising *r*, *y* and *gf* single mutant lines and their respective double mutants. The double mutants expanded the effect on metabolism of the introduced single mutations by revealing additional additive or epistatic effects that could be further exploited for the improvement and diversification of the SM landrace by introducing innovation, while maintaining the characteristic SM typology.

3.2. Materials and methods

3.2.1. Plant material and growth conditions

The 18 lines described in Chapter 2, introgressing in single or double combination genetic variants affecting the fruit phenotype, were studied. After a preliminary metabolic analysis of the entire San Marzano collection, three single mutant lines harbouring *r*, *y* and *gf* and the double mutants representing all the possible combinations were studied in more detail in comparison with the corresponding wild-type (Table 3.1). Details on the backcross scheme used to obtain these introgression lines, as well as the growth conditions used were reported before (Dono et al., 2010). Briefly, eight plants per accession were transplanted and cultivated in twin rows in an unheated tunnel following standard cultural practices for indeterminate tomatoes, using tutors and weekly removal of lateral shoots. Daily temperature was controlled by a ventilation system and plants were irrigated through a drop system. The trial was repeated with identical materials and methods for two consecutive years (2017 and 2018).

Table 3.1. Extended names, genetic symbols and description of the berry color of the seven genotypes studied in detail, including San Marzano (SM), three single mutant and their respective double mutants.

Class of material	Class of variation	Name	Genetic symbol	Fruit color
Wild-type	_	San Marzano	SM	Red
San Marzano fruit variants	Chlorophyll	*green flesh*	*gf*	Muddy brown
	Carotenoids	*yellow flesh*	*r*	Yellow
	Flavonoids	*colourless fruit epidermis*	*y*	Pink
	Double mutants	*yellow flesh + green flesh*	*r_gf*	Light yellow
		colourless fruit epidermis + green flesh	*y_gf*	Wine-coloured
		colourless fruit epidermis + yellow flesh	*y_r*	Green

3.2.2. Fruit sampling

Before sampling, berries were visually inspected and only intact and healthy tomatoes were collected. Two biological replicates for genotype, each represented by four fully ripe berries, were harvested over a period of three days, during the first week of August, for each year. A longitudinal pericarp wedge was excised from each of the four appropriately washed berries and cut into cubes; each replica, consisting of about 30 g of fresh material, was immediately frozen in liquid nitrogen and homogenized until a fine powder was obtained. Aliquots of about 10 g of this material were freeze-dried for the analysis of non-volatile secondary metabolites. All samples, both frozen and freeze-dried, were stored at −80°C until analysis.

3.2.3. Volatile detection and quantification

For volatile analysis, two biological replicates and two technical replicates were processed and analyzed independently for the two year experiments. Prior to the analysis, frozen fruit powder (0.5 g fresh weight) from each sample was weighed in a 15 mL vial, closed and incubated at 37°C for 10 min. Then, 1.1 g of $CaCl_2.2H_2O$ and 500 µL of EDTA 100 mM (pH 7.5) were added; samples were then gently shaken and sonicated for 5 min. One mL of the homogenized mixture was transferred into a 10 mL screw cap headspace vial, where volatiles were collected by head space solid-phase microextraction as previously described (Rambla et a., 2017). A 65 µm PDMS/DVB SPME fiber (Supelco) was used for all the analysis. Pre-incubation and extraction were performed at 50°C for 10 and 20 min respectively. Desorption was performed for 1 min at 250°C in splitless mode. Volatile extraction and injection were performed by means of a CombiPAL autosampler (CTC Analytics). Separation and detection were performed by a 6890N gas chromatograph coupled to a 5975B mass spectrometer (Agilent Technologies) with DB-5ms fused silica capillary column (60 m, 0.25 mm, 1 µm) (J&W Scientific). Oven temperature conditions were 40°C for 2 min, 5°C/min ramp until 250°C and then held isothermally at 250°C for 5 min. Helium was used as carrier gas at 1.2 mL/min constant flow. Ionization was performed by electron impact (ionization energy, 70 eV; source temperature 230°C). Data acquisition was performed in scan mode (mass range *m/z* 35–250; 6.2 scans per second). Chromatograms and spectra were recorded and processed using the Enhanced ChemStation software (Agilent). Untargeted

analysis of all the compounds in the chromatogram was performed by means of the MetAlign 3 software (Tikunov et al., 2005). Compounds were unequivocally identified by comparison of both mass spectra and retention time to those of pure standards (SIGMA-Aldrich). For quantification, peak areas of selected specific ions were integrated for each compound and normalized by comparison with the peak area of the same compound in a reference sample injected regularly, in order to correct for variations in detector sensitivity and fiber aging. This reference sample consisted of a homogeneous mixture of all the samples analyzed. Data for each sample were expressed as the relative content of each etable compared to those in the SM reference.

3.2.4. Non-volatile detection and quantification

For both non-volatile P and NP metabolites, two biological replicates and two technical replicates, for two years, were processed and analyzed independently. Prior to analysis, 10 mg of freeze-dried fruit powder from each sample were weighed and extracted (i) with 0.75 mL cold 75% (v/v) methanol with for 0.5mg/l formononetin as internal standard (IS) for P metabolites, as previously described (Fasano et al., 2016); and with (ii) 0.25 mL cold 100% (v/v), 1 mL of chloroform spiked with 25 mg/l α-tocopherol acetate as internal standard and 0.25 mL 50 mM Tris buffer (pH 7.5, containing 1 M NaCl) as described in (Rambla et al., 2016). Liquid chromatography coupled to high-resolution mass spectrometry (LC-HRMS) conditions were as previously reported for, respectively, polar (Diretto et al., 2017) and non-polar and (Sulli et al., 2017) metabolomes.

Metabolite identification was performed by comparing chromatographic and spectral/MS properties with authentic standards, if available, and reference spectra, and based on the m/z accurate masses as found in the Pubchem database for monoisotopic masses, or in the Metabolomics Fiehn Lab Mass Spectrometry Adduct Calculator for adduct ions. Quantification of each metabolite was carried by calculating the relative contents to the formononetin (P) and α-tocopherol acetate (NP) IS levels.

3.2.5. Statistical and bioinformatics analyses

Raw data were firstly inspected and manually curated for the presence of outliers (e.g. when % st.dev./avg exceed 30%). For Principal Component Analysis (PCA), the complete dataset after log_2 transformation and including all replicates was considered. Untargeted analysis of VOC, and NP and P metabolomes was carried out as previously reported using, respectively, MetAlign and the SIEVE software (Thermofisher Scientific; Cappelli et al., 2018).

In the seven lines studied in detail, untargeted analysis revealed a consistent Year effect. The "Gen*Year" interaction was investigated by two-way multivariate analysis of variance (MANOVA) on those metabolites than presented less than 30% missing values. The analysis was performed with the PROC GLM procedure and the MANOVA statement implemented in the SAS software package (SAS® University Edition). Since "Gen*Year" interaction was found to be the less consistent source of variation, allowance was made for the existing interaction, data were mediated over the two years and all genotypes were presented with a single mean value. PCA was performed with SIMCA-P version 11 (Umetrics) with Unit Variance normalization. The differences between each line and the SM reference were assessed using Student's *t*-test at the 5% significance level ($p<0.05$). Graphs were elaborated with Excel (Microsoft Office 2013).

Venn diagrams were generated using Venny 2.1 software (https://bioinfogp.cnb.csic.es/tools/venny/index.html). Correlation networks were generated using average values over the two years under study, as previously described (Ahrazem et al., 2018; Diretto et al., 2019). To better evaluate most robust metabolite-metabolite associations (e.g. significant correlations), the MCODE Cytoscape plugin was used (Bader et al., 2003).

3.3. Results

3.3.1. Targeted analysis of volatile, non-polar and polar compounds of the entire collection

In order to give a characterization of the flavor, volatile composition of each of the mutated lines in comparison with the wild-type SM was analyzed. The selected analytical strategy

allowed the relative quantification of 68 VOCs unequivocally identified by both mass spectra and retention index with those of authentic standards. Overall, eight compounds were related to benzenoids (B), ten to branched-chain amino acid-relatives (BCAA), nine to apocarotenoids (C), two to esters (E), twenty-four to fatty acids derivatives (L), four to phenylalanine derivatives (Phe), two to sulphur compounds (S) and six to monoterpenoids (T).

PCA of the volatile composition revealed that the first two components explained about the 45% of the total variance; the b-plot revealed the position of each line, with respect to SM. Indeed PC1 kept the delayed ripening mutants, *rin*, *Gr* and *Nr*, far from all the other lines, grouped together with SM (Figure 3.1a). Excluding them, a second view of the distribution of the mutants appeared, with PC1 clearly clustering all carotenoid lines plus *y_r*, separated from *t* (Figure 3.1b). Furthermore the same component also divided *hp2*, *gf_hp2* and *Aft_atv*, three mutants affecting all pigments under distinct points of view, from SM and its neighbor lines, *pd*, *hp-1*, *gf*, *y* and *y_gf*.

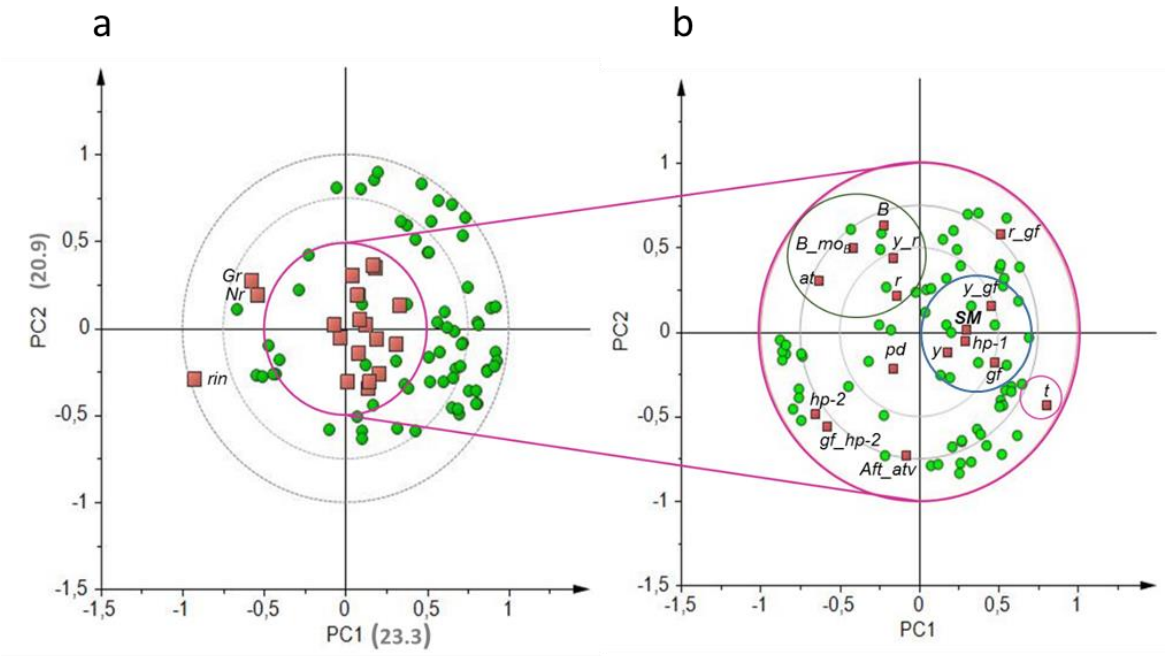


Figure 3.1. PCA of log2 values of 68 volatile compounds measured by a Solid-Phase Micro-Extraction Gas-Chromatography coupled to Mass Spectrometry (HS-SPME/GC-MS). (**a**) PC1 X PC2 b-plot of the 19 mutated lines plus San Marzano (SM). (**b**) PC1 X PC2 b-plot of the collection, *rin*, *Gr* and *Nr* excluded. Line symbols in the score plot are explained in Table 2.1 (chapter 2) and color symbols indicate respectively volatiles in green circles and mutated lines in pink squares.

To investigate changes at the NP specialized metabolome, LC-HRMS was used to determine the level of 54 known and previously validated compounds. They were divided in different metabolic classes, including 14 fatty acids (FA), one phospholipid (PHO), two sterols (STE), 15 carotenoids (CAR), eight chlorophylls (CHL), six quinones (QUI), and five tocochromanols (TOC).

The score plot of the first two PCA components explained about the 40% of the total variance, with PC1 grouping the majority of the lines near SM, except *t* and *r_gf* , located on their own and *Aft_atv*, *hp-2* and *gf_hp-2*, **s**haring a similar behaviour (Figure 3.2a). Looking at a more restricted zone around SM (Figure 3.2b), the b-plot identified once again the same groups of Figure 3.1b (VOCs), respectively, the first one with carotenoid mutants plus *y_r*, and the second including SM with the adjacent *pd*, *hp-1*, *gf* and *y*. Finally *rin*, *Nr* and *Gr* were also closely grouped (Figure 3.2b).

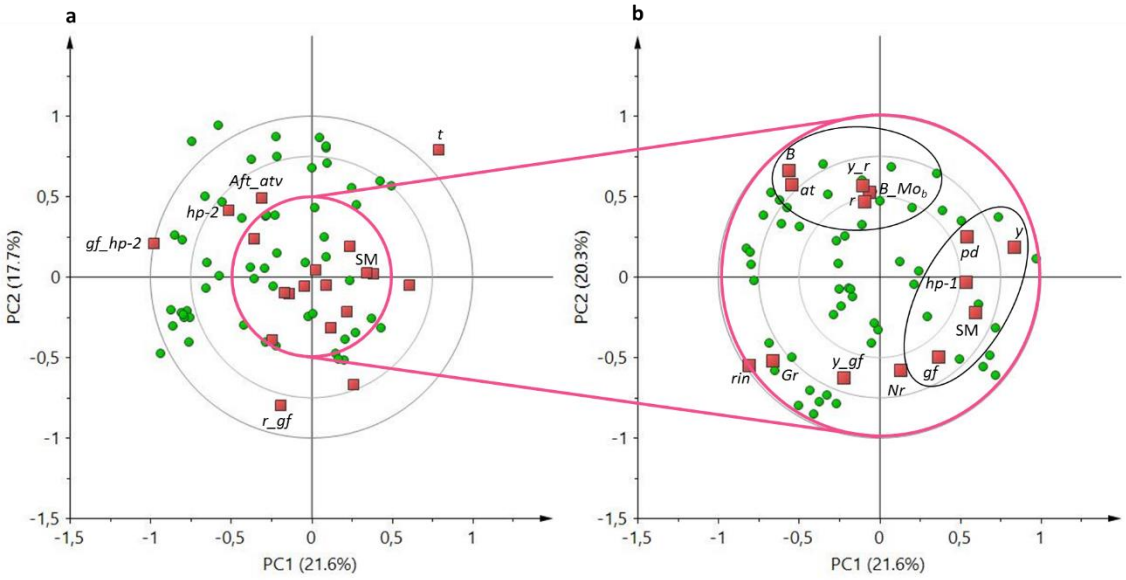


Figure 3.2. PCA of log2 values of 54 NP metabolites measured by LC-HRMS. (**a**) PC1 X PC2 b-plot of the nineteen mutated lines plus San Marzano; (**b**) PC1 X PC2 b-plot of thirteen lines, *t*, *Aft_atv*, *hp-2*, *gf-hp-2* and *r_gf* excluded. Line symbols in score plots are explained in Table 2.1, and color symbols indicate respectively volatiles in green circles and mutated lines in pink squares.

Finally the relative quantification of 128 polar metabolites allowed to complete the metabolomics characterization. P metabolites were divided into different metabolic classes, including 19 amino acids (AA), 17 acids (AC), four amines (AM), two lipids (LI), one nucleic acid (NU), 15 sugars and polyols (SAP), 11 alkaloids (ALK), 55 phenylpropanoids

(PHE) and three vitamins (VIT). The score plot of the first two components explained about the 42% of the total variance, with PC1 that particularly separated SM from the delayed ripening mutants and *Aft_atv*; interestingly, *t*, *pd* and *hp-2* also co-located remotely to the parental line (Figure 3.3a). Focusing on a restricted number of mutated lines (Figure 3.3b), PC1 separated again SM from delayed ripening, all pigment and two carotenoids mutants (*at* and *B*), while PC2 grouped together mutants carrying *y*.

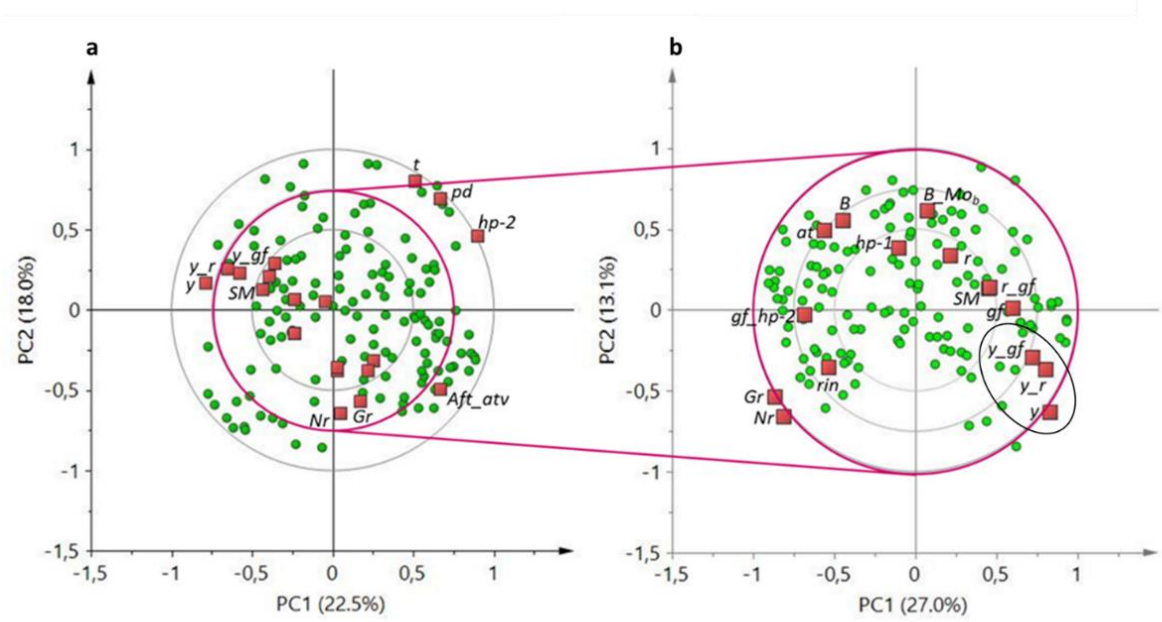


Figure 3.3. PCA of log2 values of 128 P metabolites measured by LC-HRMS. (**a**) PC1 X PC2 b-plot of the nineteen mutated lines plus San Marzano; (**b**) PC1 X PC2 b-plot of fourteen lines, *t*, *Aft_atv*, *hp-2*, and *pd* excluded. Line symbols in score plots are explained in Table 2.1, and color symbols indicate respectively volatiles in green circles and mutated lines in pink squares

In the following part of the chapter we focused on six mutated lines, *gf*, *r* and *y*, plus their double mutants. As described before, the choice fell on them firstly because of the growing market interest in these color mutants, which represent a new source of variability in terms of nutraceutical and organoleptic properties. Furthermore, even if they share the same genetic background and number of backcrosses they showed contrasting behaviour, highlighted from the metabolic analysis of the entire collection; all these conditions set the scene for a more detailed metabolic analysis of the lines under consideration.

3.3.2. Untargeted analysis of volatile, non-polar and polar metabolites

Untargeted metabolomics aims to gather information on global metabolic profiles by retrieving, in an unsupervised way, as many metabolites are detectable in a GC-MS/LC-HRMS chromatograms. The comparison of the entire VOC, NP and P metabolome between the SM control and the six mutated lines detected the total features of their metabolic profile differences, setting the stage for a more specific targeted metabolomics study later in this paper. By using this approach, 263 VOCs, 746 NP and 110 P compounds were identified in the samples, many of which were differentially accumulated in, at least, one pairwise comparison (Table S3.1). For VOCs, the first two principal components explained over 51% of total variance; PC1 separated *gf* and *y* from the other lines, while PC2 kept *r* plus all the combinations harbouring the *y* mutation distinct from the others (Figure 3.4a). In Figure 1b, PCA of the first two components for the NP compounds explained almost 53% of the total variance; SM and two *green flesh* genotypes were clearly separated from *y* mutants plus *r* by the PC1, and *r* mutants were grouped in the lower quadrants by the PC2. For P metabolites, the first two PCA components explained the over 55% of total variance (Figure 3.4c); PC1 kept more clearly SM and the lines carrying *r* separated from *y* mutants, and *y_r* was positioned exactly halfway between its parental lines. PC2 separated lines carrying *yellow flesh*, plus *y* from SM, which was in the upper side of the graph together with *gf* and *y_gf*. Overall, the untargeted metabolomic analysis revealed that the mutations mostly affected NP rather than VOC and P metabolome, with the *r* mutant showing the highest extent of changes, including 112 NP compounds (106 down- and 6 up-regulated; Table S3.1). In addition, *r* also displayed the highest epistatic attitude towards *y* (in the VOC and NP fractions) and *gf* (P metabolome), while, notably, *y* was epistatic to *gf* in the NP untargeted metabolome.

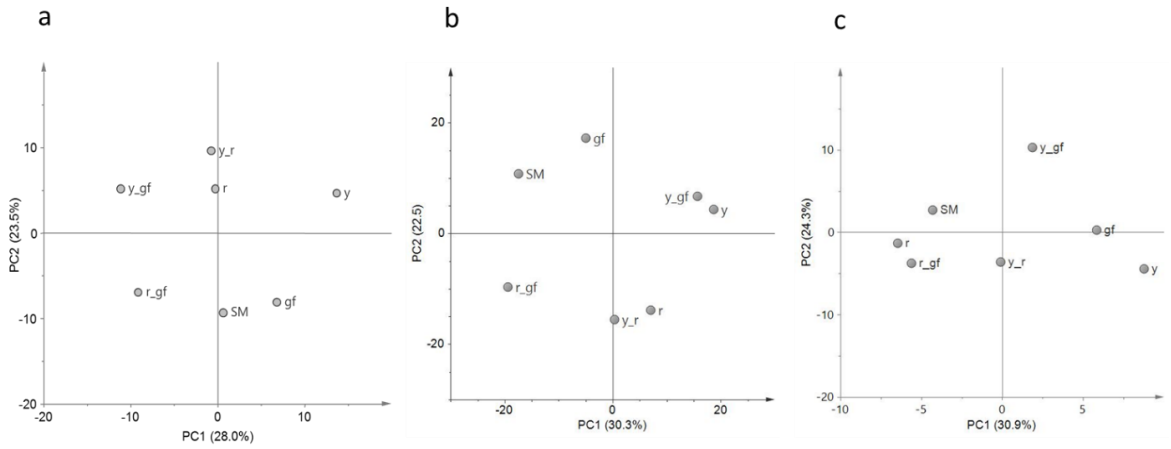


Figure 3.4. PC1 X PC2 score plots of the six mutated lines plus San Marzano (SM) according to relative values of 263 VOCs (a), 746 NP (b) and 110 P (c) metabolites measured by GC-LS and LC-HRMS. Line symbols are explained in Table 3.1.

3.3.3. Estimation of "gen*year" interaction in the quantification of targeted metabolites

Analysis of untargeted metabolites revealed the presence of a consistent Year effect (Figure S3.1). Seventy-eight VOCs, 33 NP and 69 P metabolites were independently subjected to multivariate ANOVA; the interaction was not significant for VOCs, but it was highly significant for NP and significant for P compounds (Table S3.2). "Gen*Year" interaction was found to be the least consistent source of variation; therefore, data were mediated over the two years and all genotypes were presented with a single mean value in targeted analyses.

3.3.4. Targeted analysis of volatile compounds

In order to give a more specific characterization of the flavor, volatile composition of each of the mutated lines in comparison with the wild-type SM was carried out. The selected analytical strategy allowed the relative quantification of 68 VOCs unequivocally identified by both mass spectra and retention index with those of authentic standards. Overall, eight compounds were related to benzenoids (B), ten to branched-chain amino acid-relatives (BCAA), nine to apocarotenoids (C), two to esters (E), twenty-four to fatty acids derivatives (L), four to phenylalanine derivatives (Phe), two to sulphur compounds (S) and six to monoterpenoids (T).

PCA of the volatile composition revealed that the first two components explained about the 54% of the total variance; the score plot showed the position of the double mutants, related to their parental lines, and with respect to SM (Figure 3.5a). Indeed, *r* and *y_r* were co-located in the same dial, in agreement with the VOC untargeted metabolome plot; conversely, *r_gf* placed halfway between *r* and *gf* according to PC1 (Figure 3.5a). Moreover, PC1 kept the mutants *r* and *y_r* separated from the other lines. PC2 placed *y_r* and its parental lines in the upper side of the graph, together with SM. The corresponding loading plot was able to identify groups of metabolites, often belonging to the same metabolic pathways, as apocarotenoids (C, in red) and lipids (L, in light blue), terpenoids (T, in green blue) and branched-chain amino acid derivatives (BCAA, in blue; Figure 3.5b). A comparison between the score and the loading plots revealed the overall compositional differences between the mutated lines.

Indeed, one of the most obvious features was that all the lines harbouring the *yellow flesh* mutation (*r*, *y_r* and *r_gf*) and, in a lesser extent, also the double mutant *y_gf,* were characterized by producing lower levels of volatile apocarotenoids, which was particularly dramatic in the case of some linear apocarotenoids such as 6-methyl-5-hepten-2-one (Figure 3.5). In the case of the mutants for *yellow flesh* this was in accordance with the scarcity of their carotenoid precursors. These lines also showed lower levels of several phenylalanine derivatives. A characteristic feature of the lines *y*, *r* and *y_r* was the high production of fatty acid derivatives together with low levels of branched-chain amino acid-related volatiles, conversely to double mutant lines *y_gf* and *r_gf*, which showed the opposite pattern. Finally, *y* and *gf* lines were characterised by higher levels of apocarotenoids and terpenoids (Figure 3.5b).

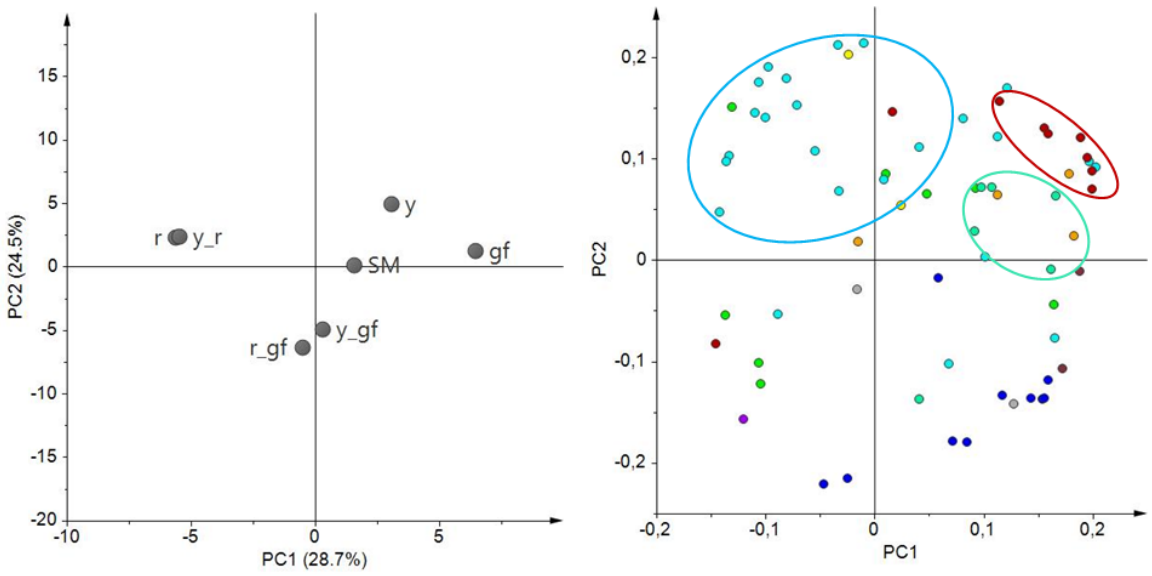


Figure 3.5. PCA of log2 values of 68 volatile compounds measured by a Solid-Phase Micro-Extraction Gas-Chromatography coupled to Mass Spectrometry (HS-SPME/GC-MS). (**a**) PC1 X PC2 score plot of the six mutated lines plus San Marzano (SM). (**b**) PC1 X PC2 loading plot. Line symbols in the score plot (**a)** are explained in Table 1. Color symbols in the loading plot (**b**) correspond to benzenoids (B, green), branched-chain amino acid-related (BCAA, blue), apocarotenoids (C, red), esters (E, yellow), fatty acid derivatives (L, light blue), phenylalanine derivatives (Phe, orange), sulphur compounds (S, brown), monoterpenoids (T, water green).

To further investigate the volatile compounds content of each of the mutated line in comparison with their original parental SM, a t-test analysis was performed. Out of 68 VOCs identified, the line with the highest number of compounds significantly different from SM was *y*, mainly because of differences in fatty acid derivatives, among other VOCs (Table 3.2). The lines containing the *r* mutation strongly differed for apocarotenoid volatiles, a group of metabolites considered to be involved in tomato flavour (Tieman et al., 2017), representing the metabolic pathways most notably altered in these ILs (Figure 3.6). Some volatiles, such as the benzenoid eugenol, had higher levels in all mutant lines (Figure 3.6a). Mutants carrying *yellow flesh* had lower levels of the apocarotenoid β-ionone (Figure 3.6b), as well as 6-methyl-5-hepten-2-one, where *y_gf* acquired lower levels too (Figure 3.6c). Furthermore, *r*, *y* and their double mutant had higher levels of many fatty acid derivatives, such as *(E)*-2-pentenal (Figure 3.6d) and 1-penten-3-one (Figure 3.6e). Lastly, *y* was enriched in phenylalanine derivatives, such as 2-phenylethanol (Figure 3.6f).

Table 3.2. Number of compounds in the different categories and classes of metabolites that are significantly different from San Marzano (SM) in each of the six lines under study.

Metabolomics fraction	Metabolic Class	Abbreviation name	Compounds No	Differentially accumulated compounds over SM (No)					
				gf	*r*	*y*	*r_gf*	*y_gf*	*y_r*
Volatile	Benzenoids	B	8	2	2	3	2		3
	Branched-chain amino acid derivatives	BCAA	10	3	3	2	4		1
	Apocarotenoids	C	9	3	5	3	6	3	5
	Esters	E	2						
	Fatty acid derivatives	L	24	3	4	6	1		4
	Others	Phe, S, T, N	15	6	4	8	4		6
Total		**VOCs**	**68**	**17**	**18**	**22**	**17**	**3**	**19**
Non-polar	Carotenoids	CAR	15	2	5	3	7	5	7
	Chlorophylls	CHL	8	2	1		2	3	
	Fatty acids	FA	14	1				1	
	Phospholipids	PHO	1			1	1		
	Quinones	QUI	6	1	2		2		3
	Tocopherols	TOC	5	1	1	1		1	2
	Others	(ND*, STE)	5	1	1		1		
Total		**NP**	**54**	**8**	**10**	**5**	**13**	**10**	**12**
Polar	Amino acids	AA	19	2	6	3	4	3	6
	Acids	AC	17	5	5	6	5	4	4
	Amines	AM	4			1	1		1
	Others	A, NU, LI	1			1		2	1
	Sugars and polyols	SAP	15	3	3	3	2	4	2
	Alkaloids	ALK	11	1	4	2	3	1	2
	Phenylpropanoids	PHE	55	6	13	24	18	18	16
	Vitamins	VIT	3	1	2				1
Total		**P**	**125**	**18**	**33**	**40**	**33**	**32**	**33**
Gran total			**247**	**43**	**61**	**67**	**63**	**45**	**64**

*Undefined compounds

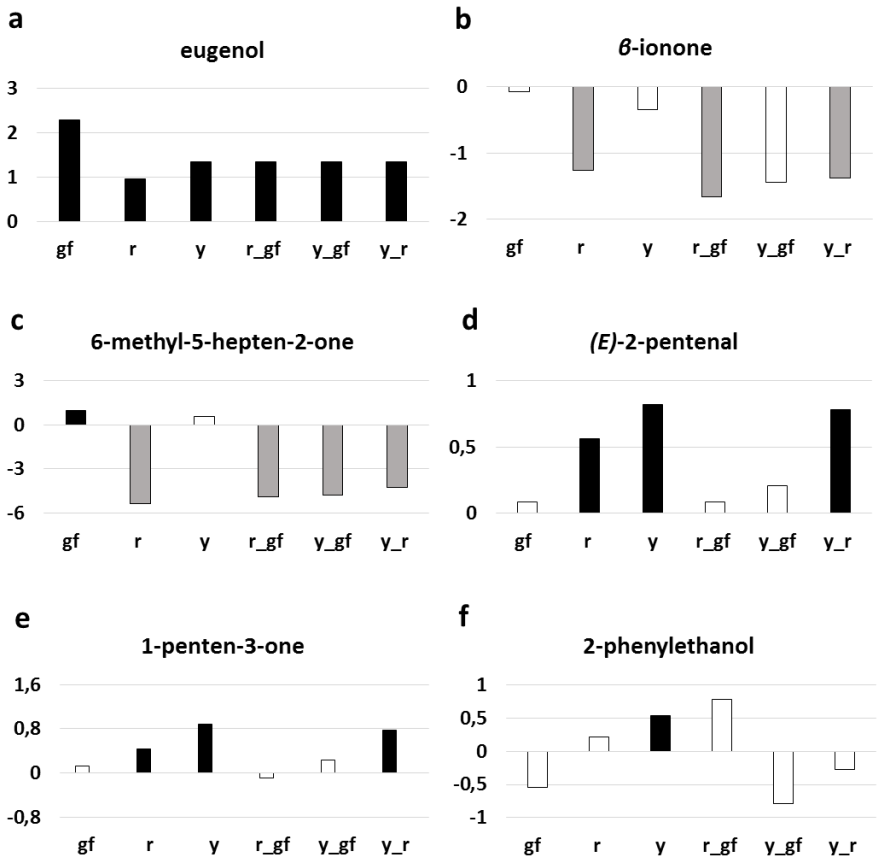


Figure 3.6. Relative $\log_2$ variation in selected volatiles involved in tomato flavour: (a) eugenol, (b) *β*-ionone, (c) 6-methyl-5-hepten-2-one, (d) *(E)*-2-pentenal, (e) 1-penten-3-one, (f) 2-phenylethanol in the six fruit mutant lines in San Marzano (SM) background compared with the recurrent parent. Line symbols are reported in Table 3.1. Bars coloured in grey and black indicate means significantly lower and higher than SM for $P \leq 0.05$ after Student's *t* test, respectively.

3.3.5. Targeted analysis of non-polar metabolites

To investigate changes at the NP specialized metabolome, LC-HRMS was used to determine the level of 54 known and previously validated compounds. They were divided in different metabolic classes, including 14 fatty acids (FA), one phospholipid (PHO), two sterols (STE), 15 carotenoids (CAR), eight chlorophylls (CHL), six quinones (QUI), and five tocochromanols (TOC). The score plot of the first two PCA components explained about the 57% of the total variance, with double mutants differently spaced from their parental lines (Figure 3.7a). Indeed, PC1 kept the mutants carrying *r* separated from SM and the other lines. PC2 clearly separated *y_r* and its parental lines from the three mutants carrying *gf* plus SM, partially confirming untargeted metabolomics results. The loading plot grouped metabolites belonging to the same metabolic pathways (Figure 3.7b). A comparison between the score and the loading plots revealed the compositional differences between the mutated lines.

Indeed, the chlorophyll group in the upper side of PC2 characterized the *green flesh* genotypes, while two quinones were in the lower side of PC2, in correspondence of the *yellow flesh* genotypes. The carotenoid group was split into early carotenoids (up to lycopene) and xanthophylls at the opposite sides of PC1; they were respectively decreased and increased in the *yellow flesh* mutant lines (Figure 3.7b).

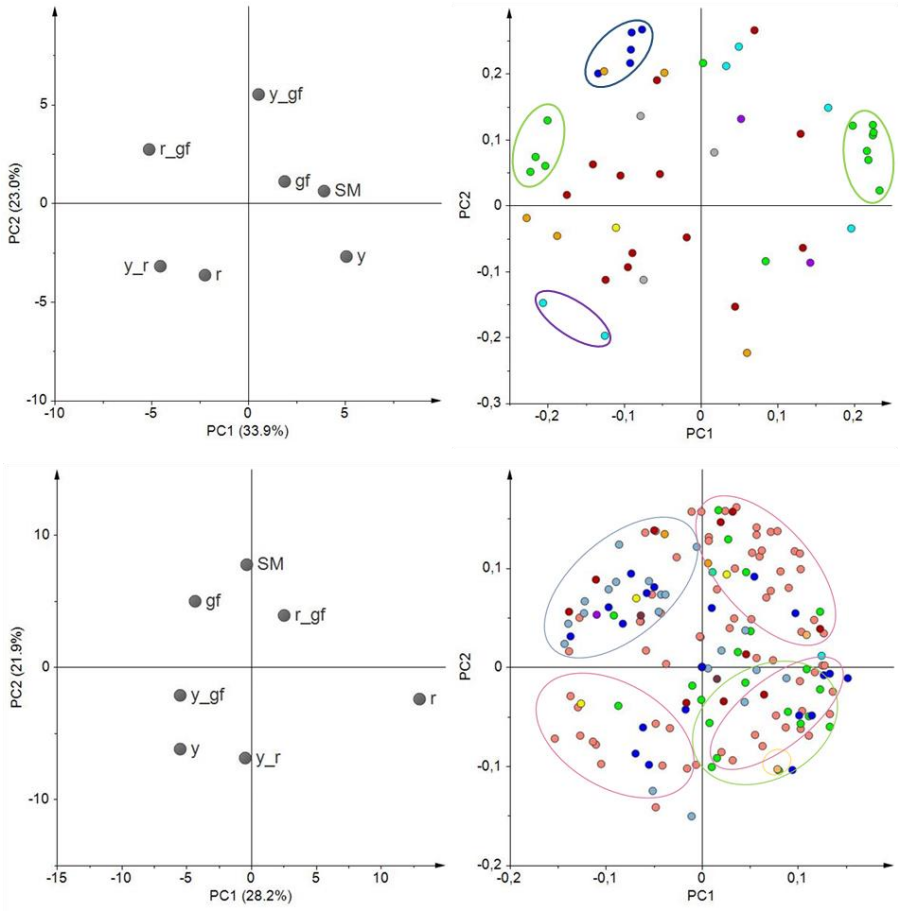


Figure 3.7. PCA of log2 values of 54 NP (a and b) and 128 P (c and d) metabolites measured by LC-HRMS. (**a-c**) PC1 X PC2 score plot of the six mutated lines plus San Marzano (SM). (**b-d**) PC1 X PC2 corresponding loading plots. Line symbols in score plots are explained in Table 3.1. Color symbols in loading plot (**b)** correspond to carotenoids (CAR, green), chlorophylls (CHL, blue), fatty acids (FA, red), phospholipids (PHO, yellow), sterols (STE, purple), tocopherols (TOC, orange) and quinones (QUI, light blue). Color symbols in **d** correspond to aminoacids (AA, green), amides (A, green water), acids (A, blue), amines (AM, yellow), alkaloids (ALK, red), lipids (LI, orange), nucleic acids (NU, brown), sugars and polyols (SAP, light blue), phenylpropanoids (PHE, pink), vitamins (VIT, light pink).

To further investigate the metabolic changes of each mutated line in comparison with the original parental SM, a t-test analysis was performed (Table 3.2), and we particularly focused on metabolites with sensorial (color, taste) and health-related properties. The lines carrying *r* showed the highest number of NP compounds different from SM, mainly due, as

expected, to differences in carotenoids (Table 3.2). The lower number of differences was shown by the single mutant *y*, indicating that this genotype is more similar to SM for NP targeted compounds.

r, *r_gf* and *y_r* were characterized by levels of phytoene (Figure 3.8a), β-carotene (Figure 3.8b) and lycopene (Figure 3.8c) lower than SM and the other lines, in agreement with previous reports; *r*, *r_gf*, *y_gf*, and *y_r* reported higher levels of the xanthophylls all-trans-neoxanthin (Figure 3.8d) and luteoxanthin (Table S4). At chlorophyll metabolism level, *gf* was characterized by higher contents of both chlorophyll a (Table S3.4) and b, with the latter also higher in *y_gf* and *r_gf* (Figure 3.8e). Moreover, *r_gf* showed higher levels of pheophytin a and pheophorbide a (Table S3.4). Drawing the attention on quinones, plastoquinone increased in lines carrying *r* (Figure 3.8f) and plastoquinol-9 in *r* and *y_r*. Lastly, □-tochopherol amount was higher than SM only in *r_gf*, while γ-tocopherol and β-tocopherol enhanced in *r_gf* and *y_gf* (Figure 3.8 g,h).

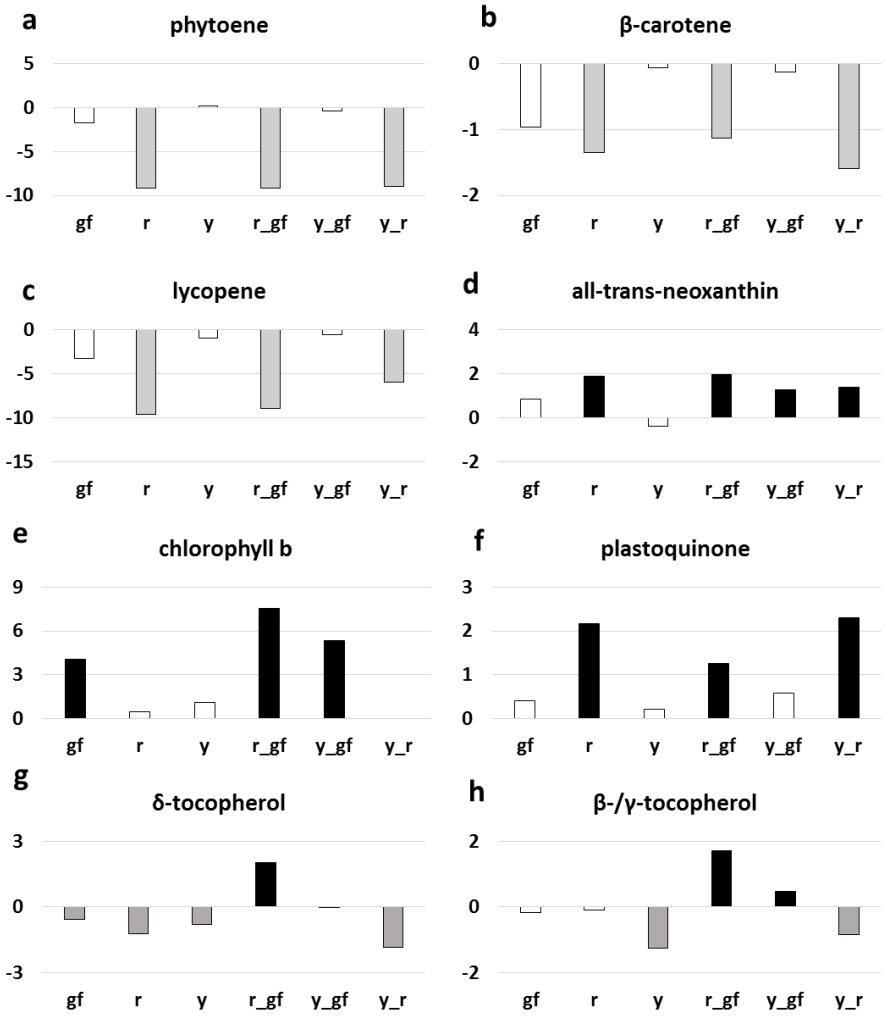


Figure 3.8. Relative $\log_2$ variation in (a) phytoene, (b) β-carotene, (c) lycopene, (d) all-trans-neoxanthin, (e) chlorophyll b, (f) plastoquinone, (g) δ-tocopherol, (h) β-/γ-tocopherol, of six fruit mutant lines in San Marzano (SM) background compared with the recurrent parent. Line symbols are reported in Table 3.1. Bars coloured in grey and black indicate means significantly lower and higher than SM for $P \leq 0.05$ after Student's t test, respectively.

3.3.6. Targeted analysis of polar metabolites

The relative quantification of 128 polar metabolites allowed to complete the metabolomics characterization of the six mutants under study. P metabolites were divided into different metabolic classes, including 19 amino acids (AA), 17 acids (AC), four amines (AM), two lipids (LI), one nucleic acid (NU), 15 sugars and polyols (SAP), 11 alkaloids (ALK), 55 phenylpropanoids (PHE) and three vitamins (VIT). The score plot of the first two components explained about the 52% of the total variance, with PC1 that particularly separated *r* and *r_gf* from all the other lines, and PC2 that identified a group including *gf* and *r_gf* together with SM (Figure 3.7c). Interestingly, P untargeted and targeted

metabolomes differently separated the mutants under study, providing clues about a large extent of distinct metabolic components contributing to their chemical profiles. The loading plot grouped metabolites belonging to the same metabolic pathway (Figure 3.7d). By the comparison of the score and the loading plots, the position of some metabolites in relation to the lines studied was highlighted; indeed, many kaempferols and quercetins were in the PHE group corresponding to the area of the *y* mutants. On the contrary, many naringenins grouped in the opposite side. Most AAs were grouped together, matching with *r* and in opposition to the SAP group (Figure 3.7d).

To further investigate the P metabolite content of each of the mutated lines in comparison with SM, a t-test analysis was performed (Table 3.2; Table S3.5), giving emphasis to nutritional- and sensorial attribute-related molecules. The line with the highest number of differentially accumulated polar compounds was the *y* single mutant, with a preponderance of down-regulated metabolites, as expected; the PHE group mostly contributed to this diversity (Table 3.2). On the contrary, *gf* was the line more similar to SM. Notably, lines containing *r* showed a higher number of AA over SM (Table 3.2).

As already highlighted in the corresponding loading plot (Figure 3.7d), lines carrying *y* had lower levels of the PHE naringenin chalcone glucoside (Figure 3.9a), and conversely higher levels of kaempferol-hexose deoxyhexose-pentose compared to SM, a biochemical phenotype also observed in *gf* and *r_gf* too for the latter (Figure 3.9b). Regarding the ALK group, calystegine B1 resulted statistically lower in all the lines, with exception of *gf* and *y_gf*, which however displayed reduced amounts compared to SM (Figure 3.9c). In addition, a series of primary metabolites were characterized by higher levels in the mutants under study: for example, the AA proline in *r* and *y_r* (Figure 3.9d) and the VIT nicotinamide in *gf*, *r* and *y_r* (Figure 3.9e). Similarly, the SAP glucoheptulose was higher in *gf*, whereas all the other lines were more similar to SM (Figure 3.9f). The LI phosphocoline displayed lower levels in lines carrying the *y* mutation (Figure 3.9g), while the AC sinapinic acid was higher in all lines, with the only exception of *r* (Figure 3.9h).

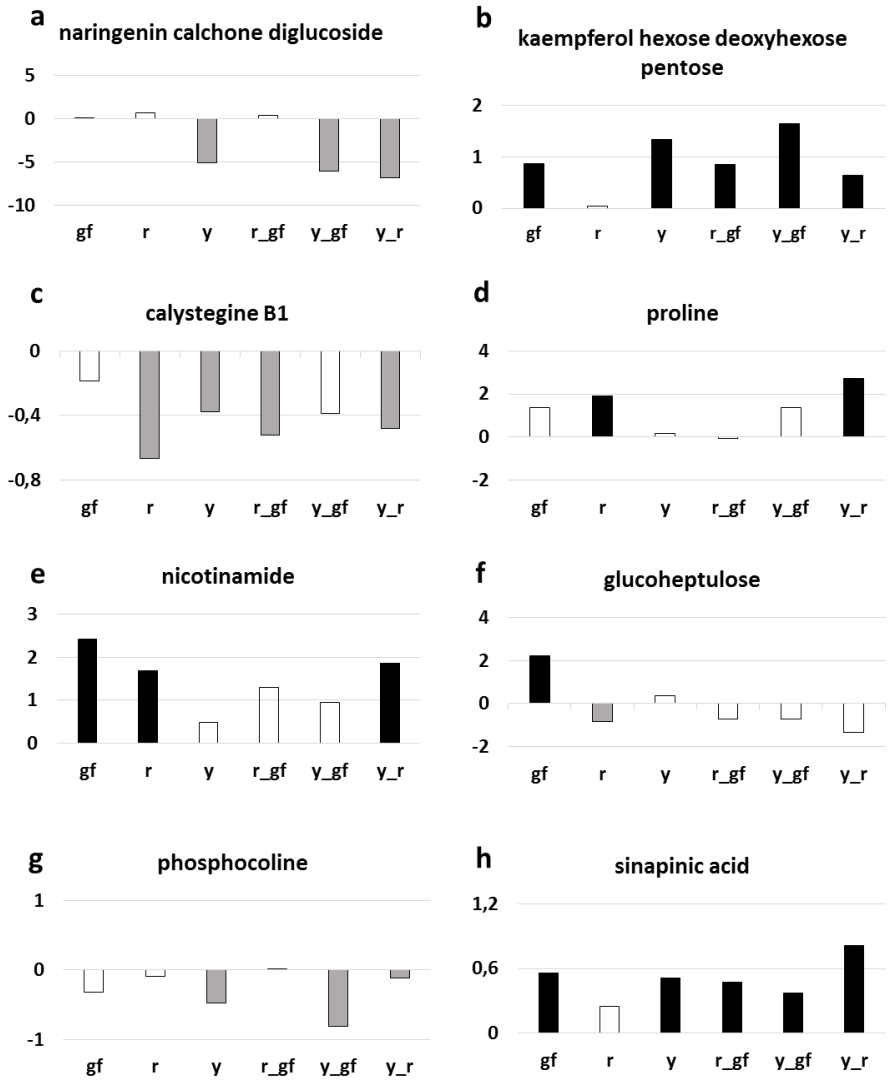


Figure 3.9. Relative $\log_2$ variation in (a) naringenin chalcone glucoside, (b) kaempferol-hexose-deoxyhexose-pentose, (c) calystegine B1, (d) proline, (e) nicotinamide, (f) glucoheptulose, (g) phosphocoline, (h) sinapinic acid, of six fruit mutant lines in San Marzano (SM) background compared with the recurrent parent. Line symbols are reported in Table 3.1. Bars coloured in grey and black indicate means significantly lower and higher than SM for $P \leq 0.05$ after Student's t-test, respectively.

3.3.7. Bioinformatics to investigate metabolite-metabolite relationships

Bioinformatic approaches, including Venn diagram visualization and correlation network analysis, were used in order to achieve a deeper understanding of the biochemical perturbations and relationships occurring in the SM mutants under study. Venn diagrams showed the degree of overlap for VOC, NP and P metabolites in each double mutant and its two parental single mutants. For line-specific metabolites in each group, all lines, except *y_gf* and *y_r*, showed a higher number of significantly up-regulated metabolites (Figure 3.10). When the overlaps between single mutants and the respective double combinations were considered, the epistasis of *r* over *gf* (29 metabolites were in common between *r* and *r_gf*) and *y* (28 metabolites in common between *r* and *y_r*) and of *y* over *gf* (21 metabolites in common between *y* and *y_gf*) were found (Figure 3.10).

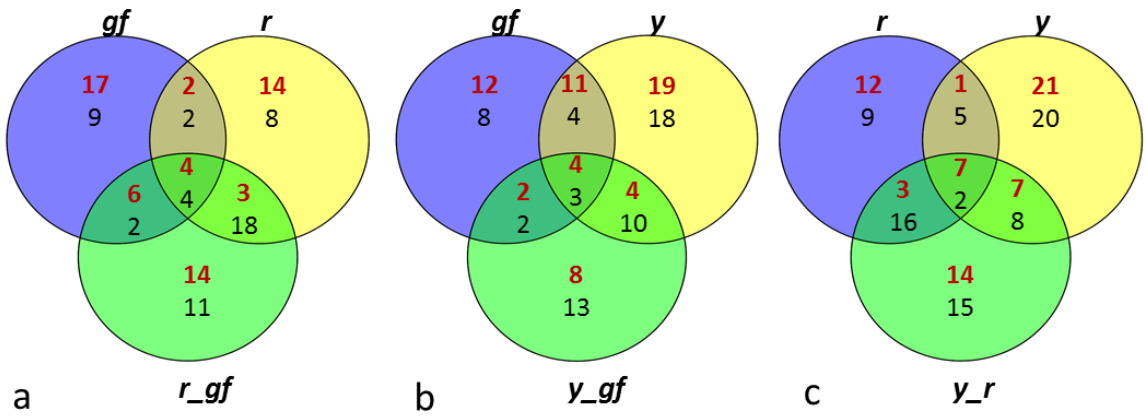


Figure 3.10. Venn diagrams for (a) *gf*, *r* and *r_gf*; (b) *gf*, *y* and *y_gf*; (c) *r*, *y* and *y_r*. Red and black numbers correspond to up- and down-regulated metabolites, respectively.

Furthermore, we used correlation network approach to investigate mutation-induced alteration at VOC, NP and P metabolome levels. To this purpose, three networks were built by integrating all differentially accumulated metabolites in, at least, each single mutant and the corresponding double mutants (Figure 3.11). Overall, the three force-directed networks allowed the achievement of specific topologies according to the distribution of the significant correlation networks existing in each metabolite-metabolite interaction; notably, a high extent of conservation was observed in each network, either in the direct

(PHE in the *y*- and CAR in the *r*-yielding networks) or not direct (L volatiles in *r, y* and *r_y;* and AA in *r*, *gf* and *r_gf*) targets of the mutations. In order to evidence the more robust and strongest correlations, the MCODE Cytoscape plugin was applied to each of the three networks (Figures S3.2-S3.4). In this way, it was possible to identify the highly interconnected regions, which resulted particularly abundant in the *r*/*gf* mutants, with a lower number of greatly dense clusters including primary (SAP, AA) and secondary (isoprenoids as CAR, CHL, QUI, T and PHE) compounds; on the contrary, the *y*/*r* mutants were characterized by a higher number of clusters with lower density, and with a high representation of amino and organic acids, besides CAR and PHE.

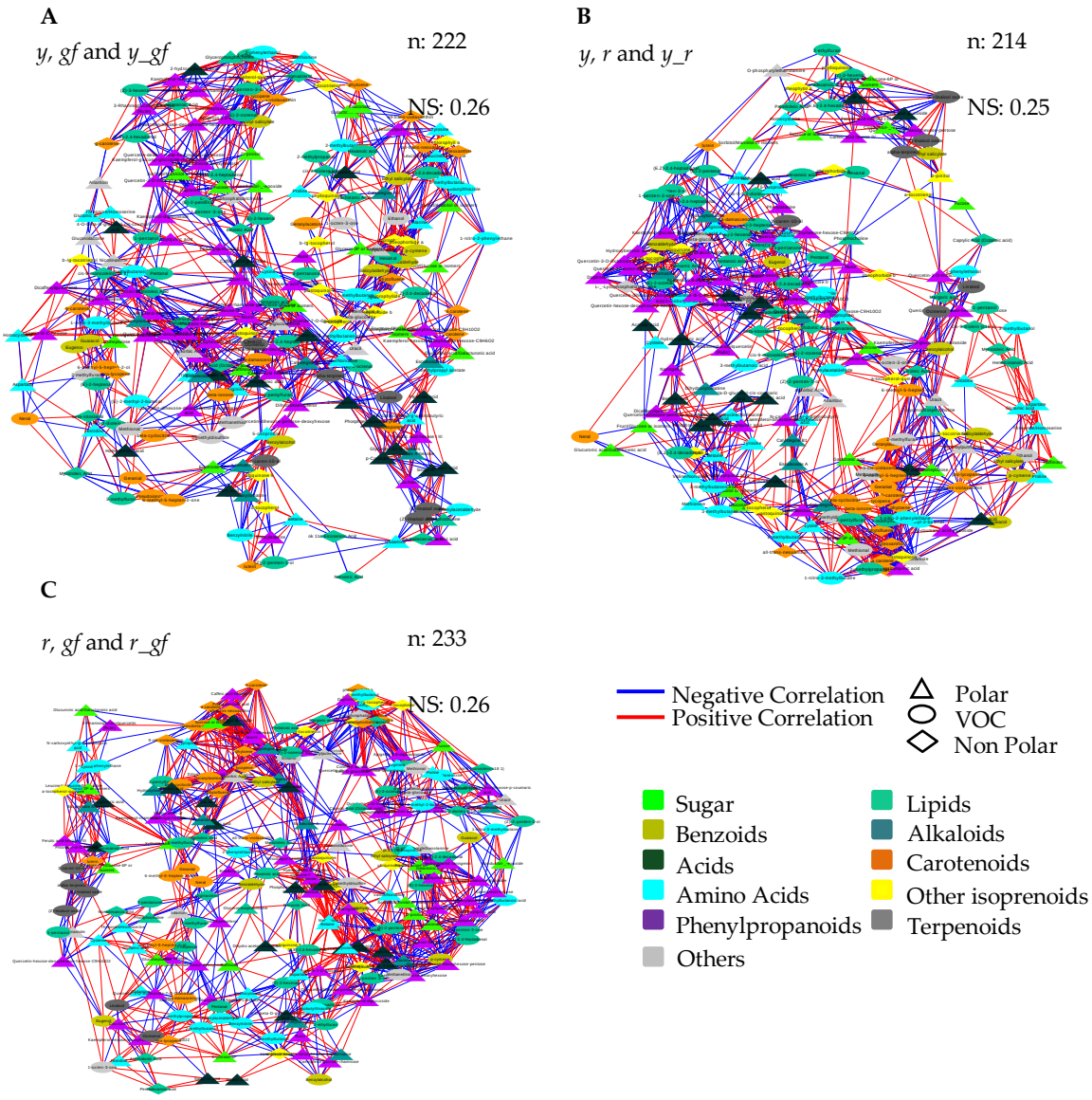


Figure 3.11. Correlation networks of metabolomics data in (a) *y, gf* and *y_gf*; (b) *y, r* and *y_r*; (c) *r, gf* and *r_gf* mutants. All the data from volatile (VOC), non-polar (NP) and polar (P) analyses were used as fold change to the San Marzano (SM) reference. Network topology was directed by the force of Pearson correlation coefficient index. Each node represents VOC (turquoise circle), a NP (diamond) or a P (triangle) metabolite. Lines joining the nodes represent positive (red) or negative (blue) correlations. Node sizes are proportional to the respective node strengths, which are shown in Tables S3.7-S3.9. Node color is depending on the metabolic class of each compound as indicated in the figure. Number of nodes (n) and network strength (NS) are shown at the top of each network. Only correlations with $|\rho|$ >0.95 are shown (*p-value* 0.05).

3.4. Discussion

This study, carried out over a 2-year period, focused on the analysis of the volatile and non-volatile compounds of the San Marzano collection, with a more detailed and specific focus on six tomato lines introgressing three tomato mutations, either in single or double combination, that have been selected as representative of the main classes of fruit pigments. The characterization of the biochemical effects of these mutations was based on the comparison with the SM original parent, with the final aim of obtaining new mutants with diversified nutritional, aesthetic and flavor characteristics. The focus on *gf*, *r* and *y* took advantage of having mutations introgressed in SM background with the same number of backcrosses (four), which guarantees a more accurate comparison. HS-SPME/GC-MS and LC-HRMS techniques were chosen for the acquisition of metabolite data due to their high sensitivity and low sample manipulation required.

3.4.1. Differences in volatile and non volatile compounds of the entire collection predict new insights

Regarding the metabolic analysis of the entire collection, the first aspect discussed concerns *Gr*, *Nr* and *rin* behaviour, that was consistent with their nature of delayed ripening mutants in all the three analysis performed. Indeed they resulted to be poor of VOCs, as expected, since these compounds are typically produced and released at the end of the maturation process, but conversely they are properly stocked as a strategy of discouraging predators from feeding on the fruit before the maturation of the seed (Rambla et al., 2014). Both non-polar and polar compound analyses maintained them grouped together, in correspondence of some metabolites (green spots); indeed, presumably, they are characterized by an accumulation of the non-polar chlorophylls.

Another interesting aspect regards the carotenoid mutants; indeed they shared an analogous trend, with *t* located on its own, in the opposite position with respect to *r*, *at*, *B*, *B_Mo$_b$* and *y_r* group. Indeed carotenoid mutants shared the same donor parent background (Ailsa craig; Dono et al., 2020), but *t* showed the highest genetic distance from SM, which could result in wider variations not due to the mutation itself, as deepened more in details in the next chapter, specifically for the volatile compounds. Similar variations in

both the non-polar carotenoids and their derivatives volatile apocarotenoids in *t*, could finally explain the analogous behavior of this line, in opposition with the others.

A group of mutants, *hp-1*, *pd*, *y* and *gf*, were located close to SM, thanks to their smaller genetic distance and their little differences in both non-polar and volatile compounds; the exception come with the polar analysis, in which *y*, its double mutants and *hp-1* were able to differentiate. Doubtless, the phenylpropanoid class of compounds, mainly affected by the *colorless fruit epidermis* mutation, provided the cause of this variation (Adato et al., 2009), as well as flavonoid accumulation in *hp-1* (Calvenzani et al., 2010) and chlorophyll variations and polyphenols increases reported in *pd* by Minoggio (2002). Finally, *hp-2*, along or coupled with *gf*, plus *Aft_atv* always co-located far from SM; in the first case the main reason has to be attributed to the variations in the content of secondary metabolites strictly due to the *hp-2* mutation, which had already been widely described in literature (Enfissi et al., 2010). This changes resulted epistatic on the *gf* mutation, which on the contrary was always located near SM. *Aft_atv* was the mutation with the highest genetic distance, with the resultant biggest difference in the metabolic content, a part from the effective changed content in anthocyanins provided by the combined mutations (Mes et al., 2008).

3.4.2. Differences in volatile compounds in six selected lines

Both untargeted and targeted PC1 of volatile compounds maintained SM close to *gf* and *y*; *r* and *y_r* were located in the same dial, as well as *y_gf* and *r_gf*, separated from SM by PC1 and PC2 in both analyses.

A significant number of VOCs considered to have an influence on tomato flavour and liking are derived from amino acids. These include some C_6-C_3 volatiles derived from a phenylpropanoid branch of phenylalanine catabolism; some of these compounds would be substrates for benzenoic volatiles, such as guaiacol and eugenol. The latter was higher in all the lines, notably in those carrying *y* (even if not significant for *y_gf*) since this mutation has notoriously alterations in the phenylpropanoid branch of phenylalanine catabolism (Koeduka et al., 2006). Another compound derived from the amino acid phenylalanine, 2-phenylethanol, was higher in *y*, as well as phenylacetaldehyde (even if not significant); their synthesis implies an initial phenylalanine decarboxylation, whose levels, however,

remained unchanged in *y*, leading to the conclusion that their variations occur subsequently in the biosynthetic pathway. Regarding the impact of these class of volatiles on the aroma, 2-phenylethanol had been previously described as the main contributor to the flavor, increasing floral aroma and the perception of sweetness (Tieman et al., 2006; Baldwin et al., 2008).

Although the question has been debated for long, carotenoid-derived volatiles have proved to have an important role in tomato flavour, as their levels positively correlate with tomato acceptability (Vogel et al., 2010). Loss of function of Psy1 in *yellow flesh* mutants leads to the lack of substrates for apocarotenoid production (Fray et al., 1993), thus justifying the low amount of this class of compounds (Figure 3.6 b-c). In particular, the enzymes involved in the production of at least some cyclic apocarotenoids, the carotenoid cleavage dioxygenases (CCDs) cleave multiple carotenoids, both linear and cyclic, explaining the reduced levels of *β*-ionone in lines carrying the *r* mutation (Vogel et al., 2008). Likewise, the lower level of the linear apocarotenoid 6-methyl-5-hepten-2-one in the same mutated lines strongly depended on the absence of the open-chain carotenoids phytoene, phytofluene, and lycopene, which can be used as precursors (Lewinsohn et al., 2005). Considering that some of these apocarotenoids provide floral or fruity notes to the aroma, tomato lines harboring the *r* mutation will probably show a lower score for this trait.

Shifting the attention on the *y* mutants, they were characterized by higher production of many fatty acid derivatives, which are the most abundant volatiles produced in the tomato fruit (Rambla et al., 2014). These include several C_5 compounds such as 1-penten-3-one or (*E*)-2-pentenal, and C_6 volatiles such as 1-hexanol, (*Z*)-3-hexenal, (*E*)-2-hexenal, or hexanal, among others. These compounds are classified as “green leaf” volatiles due to their ‘green’ characteristic, with a fresh aroma of cut grass. In the tomato fruit, the production of these compounds is increased at ripening, probably due to the loss of integrity of cellular membranes (Klee et al., 2010), and high amounts are typically released from both vegetative and fruit tissues when disrupted (Rambla et al., 2014). Previous studies suggested a reduced impact on tomato flavor and no effect on consumer preference (Chen et al., 2004; Tieman et al., 2012), although others still claim that several of these compounds have an impact on overall flavour intensity and liking (Tieman et al., 2017).

The 1-penten-3-one and *(E)*-2-pentenal increased in *r*, *y* and *r_y*, in contrast with their lower levels of C (Figure 3.6 b-c), with an evident synergistic effect of the double mutant. This pushes to further investigate the activity of the lipoxygenase–linoleate (LOX) enzyme that catalyses the oxidation of polyunsaturated fatty acids by molecular oxygen with the formation of unstable hydroperoxides which in turn oxidize carotenoid pigments (Borrelli et al., 1999).

3.4.3. Differences in non-polar compounds in six selected lines

PCA plots of the untargeted and targeted analyses showed a similar distribution of the six studied lines, with *y* and *y_r* in the same quadrant, *gf* always next to SM, *r_gf* located alone. The only difference consisted in the two mutants *y_gf* and *y*, kept separate in the targeted analysis, which could be due to the greater ability of the targeted analysis to explain their alterations, although the presence in the untargeted metabolome, of additional unknown components characterized by a differential accumulation among the mutants under study cannot be excluded (Figure 3.7 a; Figure S3.1 b).

Non-polar targeted metabolites clearly separated the group of mutants containing *r* from the group containing *gf*, mainly due to the compositional differences of carotenoids and chlorophylls (Figure 3.7 a). The double mutant *r_gf* mapped in a separate position, indicating partially additive and partially synergistic effects of this combination. The blocking effect in the carotenoid pathway of the *yellow flesh* mutation strongly reduced the colorless phytoene, and the colored carotenoids as β-carotene and lycopene (LeRosen et al., 1941); Figure 3.8 a-c). Phytofluene was also reduced, as previously described in tomato accessions (Kang et al., 2014) or, although at a stronger extent, in ripening fruits of *Psy1* knockout lines (Gady et al., 2012), in which a concomitant decrease in the volatile apocarotenoids was also observed. In literature, the study of *r* metabolites was often limited to the analysis of carotenoids, including the xanthophyll lutein, whose levels, in line with our results, did not vary (Kachanovsky et al., 2012). Few other isoprenoids as abscisic acid (ABA), gibberellins (GAs) and sterols were studied in *r* mutants (Fraser et al., 1995). To get further and general insights on the biochemical changes occurring in the mutant lines, we measured the levels of several metabolic classes in the "local" (carotenoids, and specifically the xanthophyll all-trans-neoxanthin) and global (lipids) NP metabolism. All-

trans-neoxanthin, the last xanthophyll in the carotenoid pathway, showed higher levels in the *r* mutant, with respect to SM (Figure 3.8 d). We hypothesize that this increase, consistent for all analyzed xanthophylls, could be explained as a strategy implemented by the cell to compensate for the decrease of other carotenoids, whose presence plays a crucial role. Indeed, xanthophylls act in flowing the energy through the photosynthetic apparatus and protecting organisms against damage caused by photosynthesis itself (Demmig-Adams et al., 1992). In addition, both violaxanthin and neoxanthin are key-substrates of the nine-cis-epoxycarotenoid dioxygenase (NCED), the first enzyme in ABA biosynthesis, which can influence fruit ripening and the attitude towards abiotic stresses (Parry et al., 1990). In this context, enhanced levels of xanthophylls could be interpreted as a way to provide sufficient metabolic flux to guarantee an adequate ABA production. Interestingly, recent studies have investigated the positive contribution of violaxanthin and neoxanthin towards total dietary carotenoid intake (Biehler et al., 2012). The same hypothesis can be extended to the analysis of quinones, in particular to plastoquinone, whose levels increased in the *yellow flesh* mutants. As a matter of fact, the importance of quinones in basic metabolic processes such as respiration and photosynthesis has been well established (Nohl et al., 1986), so that an increase in their contents is plausible, in a scenario where it is necessary to compensate for the lack of carotenoids. Interestingly, plastoquinol-9, an additional compound in the quinone pathway with high antioxidant properties and being involved in plant high-light acclimation (Kruk et al., 1994; Szymańska et al., 2010), increased in *r* and *y_r*.

The group of *green flesh* mutants showed no changes in carotenoids, but, as expected, was strongly affected in the content of chlorophylls (Figure 4a-b). Breakdown of chlorophyll starts in senescing chloroplasts and ends with the deposition of NCC-type catabolites inside the vacuole (Hörtensteiner et al., 2011), with the crucial enzymatic reduction of chlorophyll b to chlorophyll a by the chlorophyll b reductase enzyme (Tanaka et al., 2006). SGR mutants have been well described in several plant species to retain substantial amount of chlorophylls in fruits during ripening, maintaining other ripening-related metabolites, such as lycopene, unchanged (Figure 3.8c,e; Akhtar et al., 1999). Fruits of the *green-flesh* tomato mutant ripen to a muddy brown color (Kerr, 1956), as we confirmed with higher amount of chlorophyll b in mutants carrying *gf* (Figure 3.8e). From a nutritional point of view, a series of studies have highlighted the positive effects of chlorophylls and chlorophyll-

related metabolites (e.g. pheophytin a) on cellular inflammation and as anti-mutagen and anti-carcinogen agent (Lin et al., 2014) and reviewed in (Mishra et al., 2011), thus increasing the potential commercial value of the mutants under study.

No difference in carotenoids between *y* and SM was found, apart from the higher all-trans-neoxanthin levels in *y_gf* (Figure 3.8 d), with a similar trend of xanthophylls also in *y_r*; on the other hand, tocochromanols metabolism was strongly altered in the mutant berries: indeed, there was an almost constant decrease in tocopherols (Figure 3.8 g-h), mainly affecting *y* and *y_r*, confirming previous reports (Adato et al., 2009; Fernandez-Moreno et al., 2016), and likely associated to endogenous mechanisms regulating the synthesis of different metabolic classes with defence activities; while, several forms of tocopherols increased in *gf* single and double mutants, an interesting feature considering their importance in the human diet as Vitamin E-producing metabolites. However, this finding confirms previous data on these mutants (Almeida et al., 2016); A. Mazzucato and G.P. Soressi unpublished results).

3.4.4. Differences in polar compounds in six selected lines

For polar compounds, *y* and *y_r* remained close, both in the untargeted and targeted PCA; the latter clearly identified the group of *y* mutants (Figure 4c-d), characterized by substantial modifications of the class of phenylpropanoids, mainly due to the lack of yellow pigment naringenin chalcone. As also observed in the case of the NP fraction, untargeted and targeted metabolomes did not perfectly matched, suggesting the existence of larger, still unexplored, metabolic changes. A detailed characterization of the *y* mutation revealed extensive alterations in transcripts and metabolites associated with the phenylpropanoid/flavonoid pathway (Adato et al., 2009). *SlMYB12* is the transcription factor underlying the *y* phenotype, leading to several changes in phenylpropanoids, first the lower levels of naringenin chalcone glucoside, in line with previous results (Adato et al., 2009); Figure 3.9a). Conversely, we found higher kaempferol-hexose-deoxyhexose-pentose in the *y* mutants with respect to SM (Figure 3.9b), even if many other kaempferols showed no differences. The decrease or absence of naringenins characterized *y* and its two double mutants, leading to the conclusion that *y* is epistatic on *gf* and *r* for this class of compounds. Another aspect regarded the modifications of the quercetins, half of which showed higher

levels in lines carrying *y*, while the remaining displayed lower levels compared to SM, thus confirming the great extent of metabolic changes in the flavonoid group. Being flavonoids part of the cuticular structure, this mutation affects fruit cuticle composition, thickness and elasticity, although it has gained increasing popularity due to the characteristic colorless peel of the berry. Mintz-Oron et al. (2008) further demonstrated that increased activity of pathways generating cuticular lipids in tomato fruit peel precedes that of phenylpropanoid and flavonoid biosynthesis pathways. This finding could explain the reduced phosphocholine levels detected in the three *y* mutants.

Additional interesting metabolic changes occurred in the mutants under study: for instance, several amino acids positively varied, as proline in the *r* mutant, alone or with *y* (Figure 3.9d), and valine/norvaline, alone or combined with *gf*. Proline and valine, as free amino acids, have significant functions in plant cells including plant stress sensor, through a tight interaction with ABA and polyamines, and are precursors of BCAA-derived volatiles, respectively (Ghorbanli et al., 2013; Pál et al., 2018). Similarly, the vitamin nicotinamide, playing a primary role as catabolite of nicotinamide adenine nucleotide and, thus, of pyridine metabolism as well defence responses (Alferez et al., 2018), also increased in *r*, alone or with *y* (Figure 3.9e). All these changes provided clue about a deeper remodeling in fruit metabolism over the mutation gene function, with consequences at fruit physiology level (e.g. stress responses, fitness and ripening), which will be the subject of future studies. Remarkably, the increase in proline, valine and nicotinamide guarantee added value to the *r* mutant, due to their nutraceutical properties.

Regarding *gf*, in general it was found to be similar to SM for many classes of polar compounds, with few notable exceptions: *gf* was superior for some sugars, such as glucoheptulose, which also changed, in a negative way, in almost all the other lines, unluckily losing the trait in the *gf* double mutants.

3.4.5. Correlation and network analyses

When the overlaps of all the studied compounds between the single and double mutants were considered, Venn diagrams indicated that *r* was epistatic on both *gf* and *y*, as previously highlighted by the t-test analysis. Most of the variation was due to down-

regulated metabolites. Furthermore, *y* showed again the highest level of variation (both up- and down-regulated metabolites), as reported in Table 3.2.

Interestingly, when integrated and subjected to a correlation network analysis, differentially accumulated metabolites in any of the six single and double mutant ILs exhibited a strong level of coordination: indeed, irrespectively to the metabolic class object of the mutation, metabolites acting in the same pathway clustered together, thus indicating a great conservative capacity of the fruit metabolism in its mutation-derived reorganization. This finding is consistent with previous reports showing, either in tomato and grape, a general phenotypic-metabolic plasticity in response to genetic or environmental changes (D'Esposito et al., 2017; Rambla et al., 2016; Dal Santo et al., 2016). However, looking at the sub-clusters generated by each network, distinct and specific relationships were unravelled, with the mutants carrying the *y* and *r* mutations involving the largest number of metabolites belonging to highly diversified pathways.

3.5. Conclusions

This work was based on the study of the effects of mutations with an emerging commercial interest, compared within the traditional Italian tomato variety San Marzano. The entire collection was analysed under the biochemical point of view, helping us to select the six lines chosen to be more in-depth characterized and at the same time setting the stage for subsequent future analysis. This biochemical and bioinformatics characterization has given further insights on the effect of each mutation on fruit aesthetic, flavor and nutritional composition. The analysis of the respective three double mutants offered an added value, making it possible to establish epistatic or synergistic effects between each pair of mutations and represented a starting point for breeding new tomato lines with different phenotypes. Although *r* and *y* cause the decrement of the most important classes of health-related pigments (carotenoids and flavonoids), the compensating increase of other metabolites with nutraceutical (xanthophylls, tocopherols, amino acids) or flavour-related (phenylalanine and fatty acid derivatives) positive properties make the studied lines worthy of attention for breeding novel and better tomatoes. Novel phenotypes could take advantage of new plant breeding techniques (NPBT) as genome editing to recapitulate the original mutations in different tomato backgrounds, or in different species, and link the variations obtained directly to the effect of the mutation itself. In addition, advanced breeding programs will convert the new lines into novel elite varieties of commercial interest.

Bibliografia

Adato, A., Mandel, T., Mintz-Oron, S., Venger, I., Levy, D., Yativ, M., et al. Fruit-surface flavonoid accumulation in tomato is controlled by a SlMYB12-regulated transcriptional network. *PLoS Genet*. **2009**, 5 (12), doi: 10.1371/journal.pgen.1000777.

Ahrazem, O., Argandoña, J., Fiore, A., Aguado, C., Luján, R., Rubio-Moraga, Á., Marro, M., Araujo-Andrade, C., Loza-Alvarez, P., Diretto, G., Gòmez-Gòmez, L. Transcriptome analysis in tissue sectors with contrasting crocins accumulation provides novel insights into apocarotenoid biosynthesis and regulation during chromoplast biogenesis. *Sci. Rep*. **2018**, 8(1), 1-17, doi.org/10.1038/s41598-018-21225-z.

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

Alferez, F. M., Gerberich, K. M., Li, J. L., Zhang, Y., Graham, J. H., Mou, Z. Exogenous nicotinamide adenine dinucleotide induces resistance to citrus canker in citrus. Front. *Plant Sci.* **2018**, 9, 1472.

Almeida, J., Azevedo, M. D. S., Spicher, L., Glauser, G., vom Dorp, K., Guyer, L., ... & Demarco, D. Down-regulation of tomato PHYTOL KINASE strongly impairs tocopherol biosynthesis and affects prenyllipid metabolism in an organ-specific manner. *J. Exp. Bot*. **2016**, 67(3), 919-934, doi.org/10.1093/jxb/erv504.

Bader, G. D., Hogue, C. W. An automated method for finding molecular complexes in large protein interaction networks. *BMC bioinformatics*, **2003**, 4(1), 2.

Baldwin, E.A., Goodner, K., Plotto, A. Interactions of volatiles, sugars, and acids on perception of tomato aroma and flavor descriptors. *J. Food Sci.* **2008**, *73*, S294–S307, DOI: 10.1111/j.1750-3841.2008.00825.x

Ballester, A. R., Molthoff, J., de Vos, R., te Lintel Hekkert, B., Orzaez, D., Fernández-Moreno, J. P., et al. Biochemical and molecular analysis of pink tomatoes: deregulated expression of the gene encoding transcription factor SlMYB12 leads to pink tomato fruit color. *Plant physiol.* **2010**, 152(1), 71-84.

Barry, C. S., McQuinn, R. P., Chung, M. Y., Besuden, A., Giovannoni, J. J. Amino acid substitutions in homologs of the STAY-GREEN protein are responsible for the green-flesh and chlorophyll retainer mutations of tomato and pepper. *Plant Physiol.* **2008**, 147(1), 179-187, doi: 10.1104/pp.108.118430.

Barry, C.S., Cornelius S., Pandey. P. A survey of cultivated heirloom tomato varieties identifies four new mutant alleles at the green-flesh locus. *Mol.* **2009**, 24(3), 269-276, doi: 10.1007/s11032-009-9289-4.

Biehler, E., Alkerwi, A. A., Hoffmann, L., Krause, E., Guillaume, M., Lair, M. L., Bohn, T. Contribution of violaxanthin, neoxanthin, phytoene and phytofluene to total carotenoid intake: Assessment in Luxembourg. *J. Food Compos. Anal*. **2012**, 25(1), 56-65, doi.org/10.1016/j.jfca.2011.07.005.

Borrelli, G. M., Troccoli, A., Di Fonzo, N., Fares, C. Durum wheat lipoxygenase activity and other quality parameters that affect pasta color. *Cereal Chemistry*, **1999**, 76(3), 335-340.

Bortolotti, S., Boggio, S. B., Delgado, L., Orellano, E. G., Valle, E. M. Different induction patterns of glutamate metabolizing enzymes in ripening fruits of the tomato mutant green flesh. *Physiologia Plantarum*, **2003**, 119(3), 384-391.

Bovy, A. G., Gómez-Roldán, V., Hall, R. D. Strategies to optimize the flavonoid content of tomato fruit. *Recent advances in polyphenol research,* **2010***, 2*, 138-162.

Calvenzani, V., Martinelli, M., Lazzeri, V., Giuntini, D., Dall'Asta, C., Galaverna, G., ... & Petroni, K. Response of wild-type and high pigment-1 tomato fruit to UV-B depletion: flavonoid profiling and gene expression. *Planta*, **2010**, *231*(3), 755-765.

Cappelli, G., Giovannini, D., Basso, A. L., Demurtas, O. C., Diretto, G., Santi, C., Girelli, G., Bacchetta, L., Mariani, F. A Corylus avellana L. extract enhances human macrophage bactericidal response against Staphylococcus aureus by increasing the expression of anti-inflammatory and iron metabolism genes. *J. Funct. Foods*, **2018**, 45, 499-511, doi.org/10.1016/j.jff.2018.04.007.

Chen, G., Hackett, R., Walker, D., Taylor, A., Lin, Z., Grierson, D. Identification of a specific isoform of tomato lipoxygenase (TomloxC) involved in the generation of fatty acid-derived flavor compounds. *Plant physiol*. **2004**, 136(1), 2641-2651, doi.org/10.1104/pp.104.041608.

Cocaliadis, M. F., Fernández-Muñoz, R., Pons, C., Orzaez, D., Granell, A. Increasing tomato fruit quality by enhancing fruit chloroplast function. A double-edged sword? *J. Exp. Bot.* **2014**, 65(16), 4589-4598, doi.org/10.1034/j.1399-3054.2003.00184.x

D'Esposito, D., Ferriello, F., Dal Molin, A., Diretto, G., Sacco, A., Minio, A., et al. Unraveling the complexity of transcriptomic, metabolomic and quality environmental response of tomato fruit. *BMC plant biology*, **2017**, 17(1), 66, doi.org/10.1186/s12870-017-1008-4.

Dal Santo, S., Fasoli, M., Negri, S., D'Incà, E., Vicenzi, N., Guzzo, F. Plasticity of the berry ripening program in a white grape variety. *Front. Plant Sci*. **2016**, 7, 970, doi.org/10.3389/fpls.2016.00970.

Demmig-Adams, B., Adams Iii, W. W. Photoprotection and other responses of plants to high light stress. *Annual review of plant biology*, **1992**, 43(1), 599-626.

Diretto, G., Frusciante, S., Fabbri, C., Schauer, N., Busta, L., Wang, Z., Matas, A.J., Fiore, A., Rose, J.K.C., Fernie, R.A. Manipulation of β-carotene levels in tomato fruits results in increased ABA content and extended shelf life. *Plant biotechnol. J.* **2019**, doi.org/10.1111/pbi.13283

Diretto, G., Rubio-Moraga, A., Argandoña, J., Castillo, P., Gómez-Gómez, L., Ahrazem, O. Tissue-specific accumulation of sulfur compounds and saponins in different parts of garlic cloves from purple and white ecotypes. *Molecules*, **2017**, 22(8), 1359, doi.org/10.3389/fpls.2014.00170.

Ercolano, M. R., Carli, P., Soria, A., Cascone, A., Fogliano, V., Frusciante, L., Barone, A. Biochemical, sensorial and genomic profiling of traditional Italian tomato varieties. *Euphytica*, **2008**, 164(2), 571-582, doi: 10.1007/s10681-008-9768-4.

Fasano, C., Diretto, G., Aversano, R., D'Agostino, N., Di Matteo, A., Frusciante, L., Carputo, D. Transcriptome and metabolome of synthetic Solanum autotetraploids reveal key genomic stress events following polyploidization. *New Phytologist*, **2016**, 210(4), 1382-1394, doi.org/10.1111/nph.13878.

Fernandez-Moreno, J. P., Tzfadia, O., Forment, J., Presa, S., Rogachev, I., Meir, S., Orzaez, D., Aharoni, A., Granell, A. Characterization of a new pink-fruited tomato mutant results in the identification of a null allele of the SlMYB12 transcription factor. *Plant Physiol*. **2016**, 171(3), 1821-1836, doi.org/10.1104/pp.16.00282.

Fraser, P., Hedden, P., Cooke, D., Bird, C., Schuch, W., Bramley, P. The effect of reduced activity of phytoene synthase on isoprenoid levels in tomato pericarp during fruit development and ripening. *Planta*, **1995**, 196(2), 321-326.

Fray, R. G., Grierson, D. Identification and genetic analysis of normal and mutant phytoene synthase genes of tomato by sequencing, complementation and co-suppression. *Plant mol. biol*. **1993**, 22(4), 589-602.

Gady, A. L., Vriezen, W. H., Van de Wal, M. H., Huang, P., Bovy, A. G., Visser, R. G., Bachem, C. W. Induced point mutations in the phytoene synthase 1 gene cause differences in carotenoid content during tomato fruit ripening. *Mol*. **2012**, 29(3), 801-812, doi: 10.1007/s11032-011-9591-9.

Gascuel, Q., Diretto, G., Monforte, A. J., Fortes, A. M., & Granell, A. Use of natural diversity and biotechnology to increase the quality and nutritional content of tomato and grape. *Front. Plant Sci.* **2017**, *8*, 652, doi.org/10.3389/fpls.2017.00652.

Ghorbanli, M., Gafarabad, M., Amirkian, T. A. N. N. A. Z., Allahverdi, M. B. Investigation of proline, total protein, chlorophyll, ascorbate and dehydroascorbate changes under drought stress in Akria and Mobil tomato cultivars. *Iran. J. Plant Physiol*. **2013**, 3(2), 651-658, doi: 10.22034/IJPP.2013.540675.

Hörtensteiner, S., Kräutler, B. Chlorophyll breakdown in higher plants. (*BBA)-Bioenergetics*, **2011**, 1807(8), 977-988, doi: 10.1016/j.bbabio.2010.12.007.

Hunt, G. M., Baker, E. A. Phenolic constituents of tomato fruit cuticles. *Phytochemistry*, **1980**, 19(7), 1415-1419, doi.org/10.1016/0031-9422(80)80185-3.

Jenkins, J. A., G. Mackinney. Carotenoids of the apricot tomato and its hybrids with yellow and tangerine. *Genetics*, **1955**, 40(5), 715.

Kachanovsky, D. E., Filler, S., Isaacson, T., Hirschberg, J. Epistasis in tomato color mutations involves regulation of phytoene synthase 1 expression by cis-carotenoids. *PNAS*. **2012**, 109(46), 19021-19026, doi: 10.1073/pnas.1214808109.

Kachanovsky, D. E., Filler, S., Isaacson, T., Hirschberg, J. Epistasis in tomato color mutations involves regulation of phytoene synthase 1 expression by cis-carotenoids. *PNAS*. **2012**, 109(46), 19021-19026, doi: 10.1073/pnas.1214808109.

Kang, B., Gu, Q., Tian, P., Xiao, L., Cao, H., & Yang, W. A chimeric transcript containing Psy1 and a potential mRNA is associated with yellow flesh color in tomato accession PI 114490. *Planta*, **2014**, 240(5), 1011-1021, doi: 10.1007/s00425-014-2052-z.

Kerr, E. A. Green flesh, gf. *Rpt Tomato Genet. Coop*. **1956**, 6, 17.

Klee, H. J. Improving the flavor of fresh fruits: genomics, biochemistry, and biotechnology. *New Phytol*. **2010**, 187(1), 44-56, doi.org/10.1111/j.1469-8137.2010.03281.

Koeduka, T., Fridman, E., Gang, D.R., Vassão, D.G., Jackson, B.L., Kish, C.M., Orlova, I., Spassova, S.M., Lewis, N.G., Noel, J.P., Baiga, T.J., Dudareva, N., Pichersky, E. Eugenol and isoeugenol, characteristic aromatic constituents of spices, are biosynthesized via reduction of a coniferyl alcohol ester. *PNAS*, **2006***, 103*, 10128–10133, DOI: 10.1073/pnas.0603732103.

Kruk, J., Schmid, G. H., Strzalka, K. Antioxidant properties of plastoquinol and other biological prenylquinols in liposomes and solution. *Free Radic. l Res*. **1994**, 21(6), 409-416.

LeRosen, A. L., Went, F. W., Zechmeister, L. Relation between genes and carotenoids of the tomato. P*NAS of USA.* **1941**, 27(5), 236, doi: 10.1073/pnas.27.5.236.

Lewinsohn, E., Sitrit, Y., Bar, E., Azulay, Y., Ibdah, M., Meir, A., Yosef, E., Zamir, D., Tadmor, Y. Not just colors—carotenoid degradation as a link between pigmentation and aroma in tomato and watermelon fruit. *Trends Food Sci. Technol.* **2005**, *16*(9), 407-415, doi.org/10.1016/j.tifs.2005.04.004.

Lewinsohn, E., Sitrit, Y., Bar, E., Azulay, Y., Meir, A., Zamir, D., Tadmor, Y. Carotenoid pigmentation affects the volatile composition of tomato and watermelon fruits, as revealed by comparative genetic analyses. *J. Agr. Food Chem.* **2005**, 53(8), 3142-3148, doi.org/10.1021/jf047927t.

Lin, C. Y., Lee, C. H., Chang, Y. W., Wang, H. M., Chen, C. Y., Chen, Y. H. Pheophytin a inhibits inflammation via suppression of LPS-induced nitric oxide synthase-2, prostaglandin E2, and interleukin-1β of macrophages. *Int. J. Mol*. **2014**, 15(12), 22819-22834, doi.org/10.3390/ijms151222819.

Lindstrom, E.W. Inheritance in tomato. *Genetica*, **1925**, 10 (4), 305.

Mintz-Oron, S., Mandel, T., Rogachev, I., Feldberg, L., Lotan, O., Yativ, M., Jetter, R., Venger, I., Adato, A., Aharoni, A. Gene expression and metabolism in tomato fruit surface tissues. *Plant Physiol*. **2008**, 147(2), 823-851, doi.org/10.1104/pp.108.116004.

Mishra, V. K., Bachet, R. K., Azamal Husen. Medicinal uses of chlorophyll: a critical overview. Chlorophyll: Structure, Function and Medicinal, Le, H., Salcedo, E., Eds **2011**

Nohl, H., Jordan, W., Youngman, R. J. Quinones in biology: functions in electron transfer and oxygen activation. *Free Radic Biol Med*. **1986**, 2(1), 211-279, doi.org/10.1016/S8755-9668(86)80030-8.

Pál, M., Tajti, J., Szalai, G., Peeva, V., Végh, B., Janda, T. Interaction of polyamines, abscisic acid and proline under osmotic stress in the leaves of wheat plants. *Sci. Rep*. **2018**, 8(1), 1-12, doi.org/10.1038/s41598-018-31297-6.

Parry, A. D., Babiano, M. J., Horgan, R. The role of cis-carotenoids in abscisic acid biosynthesis. *Planta*, **1990**, 182.1, 118-128, doi: 10.1007/BF00239993.

Rambla, J. L., Medina, A., Fernández-del-Carmen, A., Barrantes, W., Grandillo, S., Cammareri, M., Lopez-Casado, G., Rodrigo, G., Alonso, A., Garcìa-Martinez, S. Identification, introgression, and validation of fruit volatile QTLs from a red-fruited wild tomato species. *J. Exp. Bot*. **2017**, 68(3), 429-442, doi.org/10.1093/jxb/erw455, doi.org/10.1186/s12870-017-1008-4.

Rambla, J. L., Tikunov, Y. M., Monforte, A. J., Bovy, A. G., Granell, A. The expanded tomato fruit volatile landscape. *J. Exp. Bot*. **2014**, 65(16), 4613-4623, doi: 10.1093/jxb/eru128.

Rambla, J. L., Trapero-Mozos, A., Diretto, G., Rubio-Moraga, A., Granell, A., Gómez-Gómez, L., Ahrazem, O. Gene-metabolite networks of volatile metabolism in airen and tempranillo grape cultivars revealed a distinct mechanism of aroma bouquet production. *Front. Plant sci*. **2016**, 7, 1619, DOI: 10.3389/fpls.2016.01619.

Rick, C. M., Butler, L. Cytogenetics of the tomato. *Adv. Genet*. **1956**, 8, 267-382.

Rocha, M. D. C., Deliza, R., Corrêa, F. M., do Carmo, M. G., Abboud, A. C. A study to guide breeding of new cultivars of organic cherry tomato following a consumer-driven approach. *Food Res.* **2013**, 51(1), 265-273, doi.org/10.1016/j.foodres.2012.12.019.

Rodríguez-Villalón, A., Gas, E., Rodríguez-Concepción, M. Phytoene synthase activity controls the biosynthesis of carotenoids and the supply of their metabolic precursors in dark-grown Arabidopsis seedlings. *The Plant J.* **2009**, 60(3), 424-435, doi: 10.1111/j.1365-313X.2009.03966.

Sulli, M., Mandolino, G., Sturaro, M., Onofri, C., Diretto, G., Parisi, B., Giuliano, G. Molecular and biochemical characterization of a potato collection with contrasting tuber carotenoid content. *PloS one* **2017**, 12(9), doi.org/10.1371/journal.pone.0184143.

Szymańska, R., Kruk, J. Plastoquinol is the main prenyllipid synthesized during acclimation to high light conditions in Arabidopsis and is converted to plastochromanol by tocopherol cyclase. *Plant Cell Physiol*. **2010**, 51(4), 537-545.

Tanaka, A., Tanaka, R. Chlorophyll metabolism. *Curr. Opin. Plant biol*. **2006**, 9(3), 248-255, doi: 10.1016/j.pbi.2006.03.011.

Tieman, D., Bliss, P., McIntyre, L. M., Blandon-Ubeda, A., Bies, D., Odabasi, A. Z., Rodriguez, G.R., van der Knaap, E., Taylor, M.G., Goulet, C. The chemical interactions underlying tomato flavor preferences. *Curr. Biol*. **2012**, 22(11), 1035-1039, doi: 10.1016/j.cub.2012.04.016.

Tieman, D., Taylor, M., Schauer, N., Fernie, A. R., Hanson, A. D., Klee, H. J. Tomato aromatic amino acid decarboxylases participate in synthesis of the flavor volatiles 2-phenylethanol and 2-phenylacetaldehyde. *PNAS.* **2006**, *103*(21), 8287-8292, doi.org/10.1073/pnas.0602469103.

Tikunov, Y., Lommen, A., De Vos, C. R., Verhoeven, H. A., Bino, R. J., Hall, R. D., Bovy, A. G. A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. *Plant physiol.* **2005**, 139(3), 1125-1137, doi.org/10.1104/pp.105.068130

Vogel, J. T., Tan, B. C., McCarty, D. R., Klee, H. J. The carotenoid cleavage dioxygenase 1 enzyme has broad substrate specificity, cleaving multiple carotenoids at two different bond positions. *J. Biol. Chem*. **2008**, 283(17), 11364-11373, doi: 10.1074/jbc.M710106200.

Vogel, J. T., Tieman, D. M., Sims, C. A., Odabasi, A. Z., Clark, D. G., Klee, H. J. Carotenoid content impacts flavor acceptability in tomato (Solanum lycopersicum). *J. Sci. Food Agr.* **2010**, 90(13), 2233-2240, doi: 10.1002/jsfa.4076.

Supplementary tables

Table S3.1. Number of volatile (VOC), polar (P) and non-polar (NP) untargeted metabolites significantly different from Sam Marzano (SM) in the six introgression lines. Line symbols are explained in Table 1.

Genotypes	No. of untargeted metabolites						
	VOC		P		NP		Total
	Up	Down	Up	Down	Up	Down	
gf	30	19	6	9	4	84	152
r	29	42	1	10	6	106	194
y	52	12	8	11	4	68	155
r_gf	5	52	2	6	1	46	112
y_gf	8	55	7	9	2	94	175
y_r	26	47	3	13	9	57	155

Table S3.2. Fisher's F test (F) and relative degree of significance (P) after two-way multivariate analysis of variance (MANOVA) for three main categories of metabolites analyzed.

Category of metabolites	No. of compounds	MANOVA sources of variation					
		Gen		Year		Gen*Year	
		F	P	F	P	F	P
Volatile	78	11.31	<.0001	29.43	<.0001	1.34	0.2645
Non polar	33	37.97	<.0001	527.29	<.0001	5.84	0.0005
Polar	69	4.1	0.0045	26.58	<.0001	3.52	0.0101

Supplementary figures

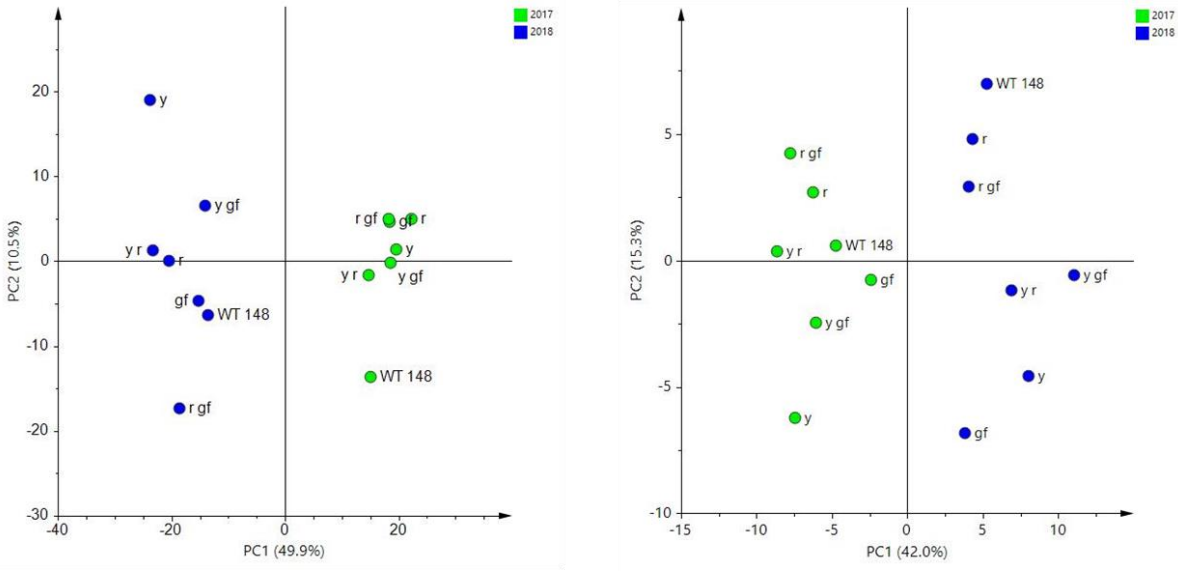


Figure S3.1. PC1 X PC2 score plots of the six mutated lines plus San Marzano (SM), sampled in two different years, according to relative values of 263 volatile (a), 746 non-polar (b) and 110 polar (c) metabolites measured by a Solid-phase micro-extraction Gas-Cromatography coupled with a mass spectrometry instrument (HS-SPME-GC-MS). Line symbols are explained in Table 3.1.

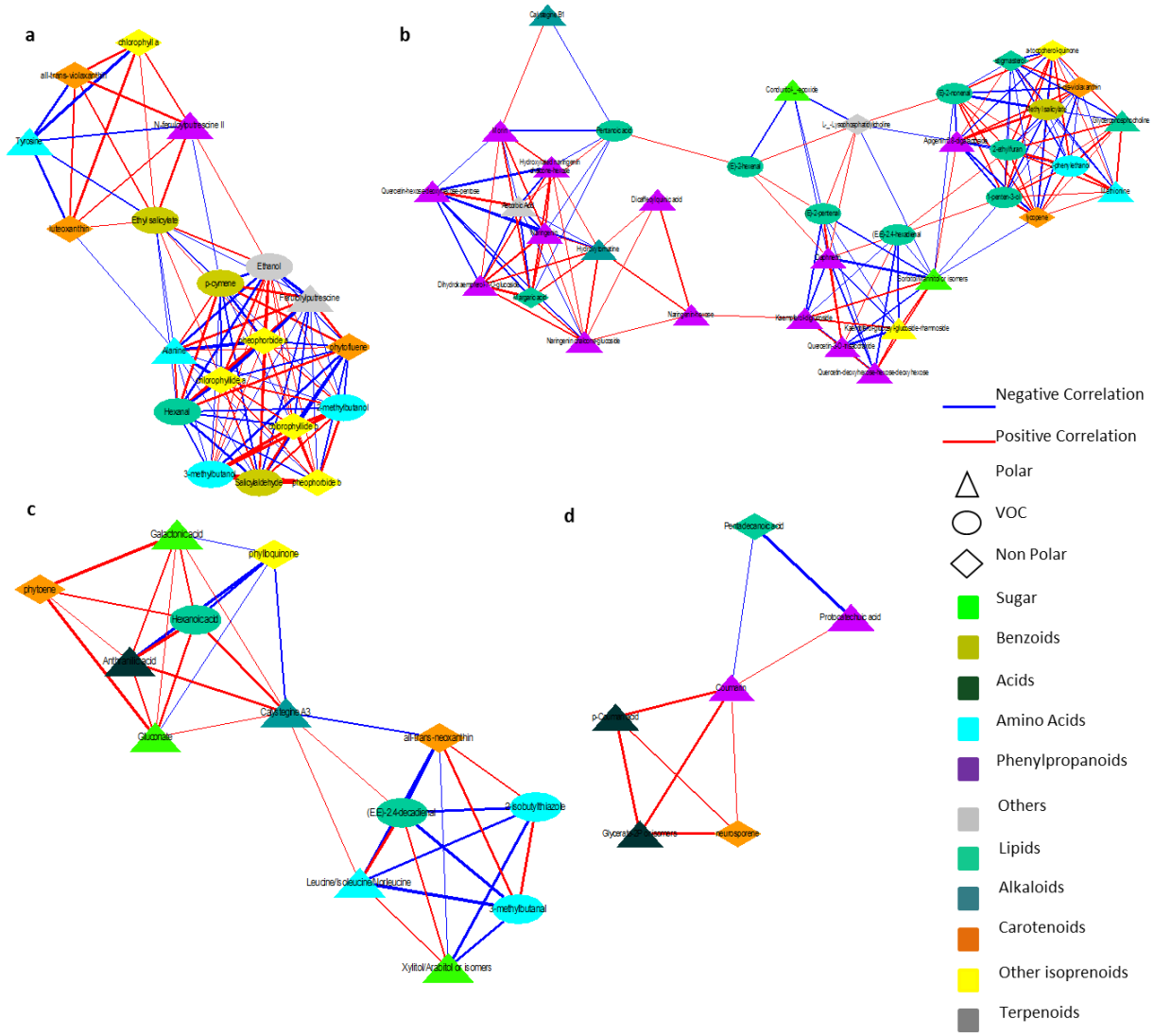


Figure S3.2. Subnetworks related to *y, gf* and *y_gf* and generated from the Global Network present in Fig. 11 using MCODE (Cytoscape plugin; see text for details). Each subnetwork highlights the high densely connected regions present in the global network. Each node representing a polar metabolite (triangle), a non-polar metabolite (diamond) or a VOC metabolite (turquoise circle). Lines joining the nodes represent correlations; direct correlations are shown in red, while inverse correlations are in blue. Node sizes are proportional to the respective node strengths. The color of the node is depending from the membership metabolic class as reported on the top. Number of nodes (n) and network strength (NS) are shown on top of each network. Only correlations whit $|\rho|$ >0.95 are shown (*p-value* 0.05).

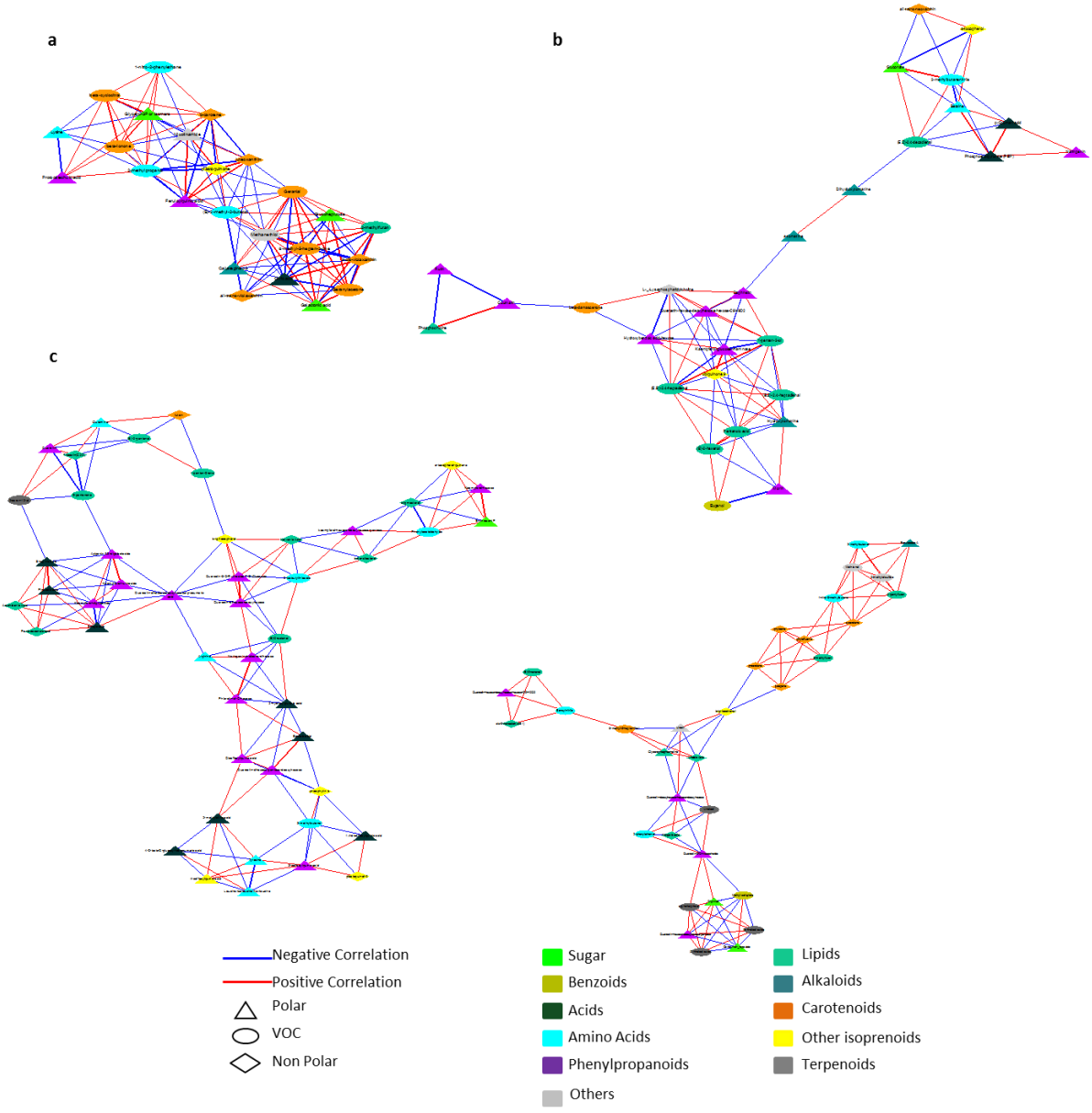


Figure S3.3. Subnetworks related to *y, r* and *y_r* and generated from the Global Network present in Fig. 8 using MCODE (Cytoscape plugin; see text for details). Each subnetwork highlights the high densely connected regions present in the global network. Each node representing a polar metabolite (triangle), a non-polar metabolite (diamond) or a VOC metabolite (turquoise circle). Lines joining the nodes represent correlations; direct correlations are shown in red, while inverse correlations are in blue. Node sizes are proportional to the respective node strengths, which are shown in Table S8. The color of the node is depending from the membership metabolic class as reported on the top. Number of nodes (n) and network strength (NS) are shown on top of each network. Only correlations whit $|\rho|$ >0.95 are shown (*p-value* 0.05).

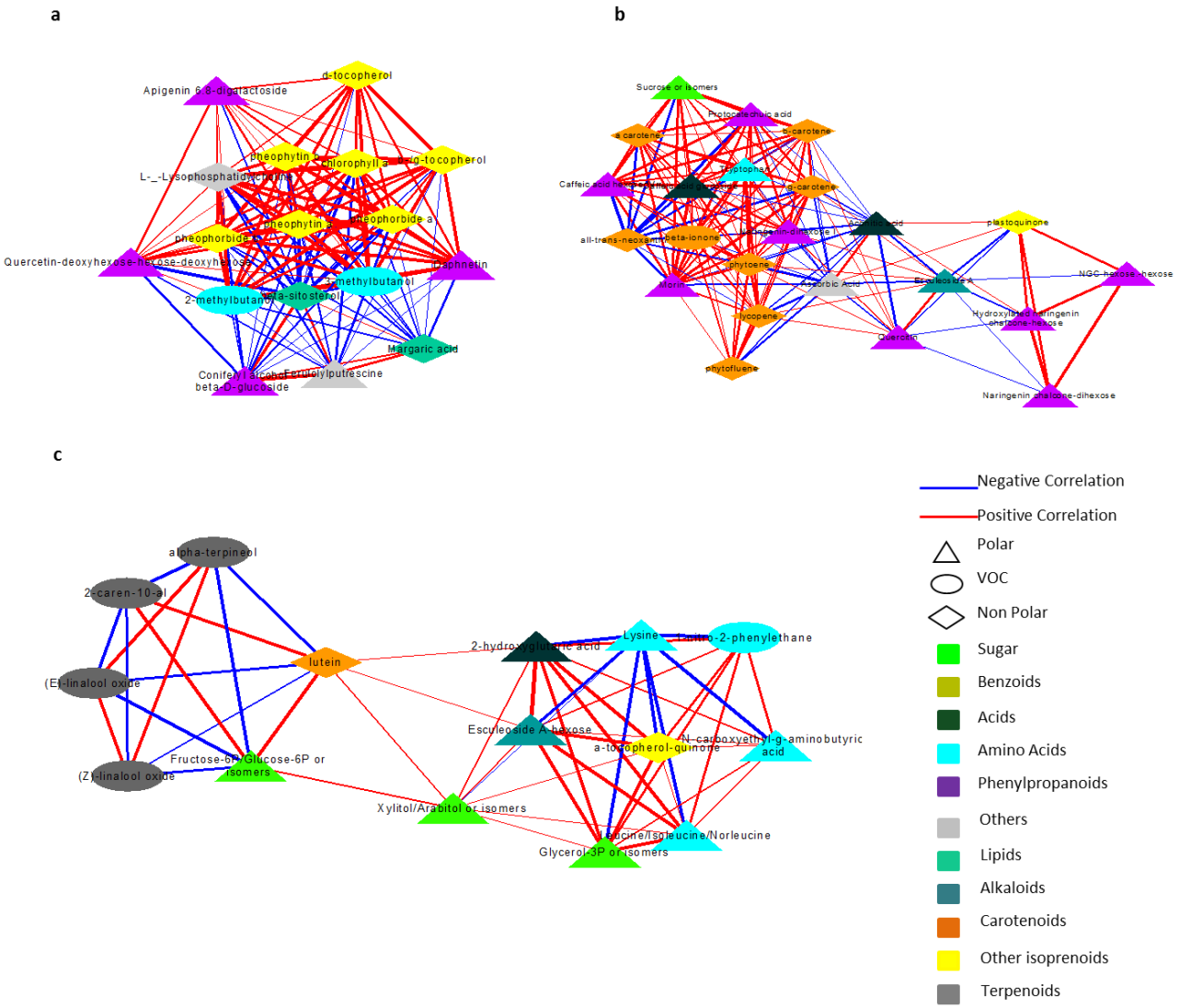


Figure S3.4. Subnetworks related to *r, gf* and *r_gf* and generated from the Global Network present in Fig. 8 using MCODE (Cytoscape plugin; see text for details). Each subnetwork highlights the high densely connected regions present in the global network. Each node representing a polar metabolite (triangle), a non-polar metabolite (diamond) or a VOC metabolite (turquoise circle). Lines joining the nodes represent correlations; direct correlations are shown in red, while inverse correlations are in blue. Node sizes are proportional to the respective node strengths, which are shown in Table S9. The color of the node is depending from the membership metabolic class as reported on the top. Number of nodes (n) and network strength (NS) are shown on top of each network. Only correlations whit $|\rho|$ >0.95 are shown (*p-value* 0.05)

Chapter 4.

Application of the CRISPR-Cas9 tool to reproduce *tangerine* and *green flesh* mutations in San Marzano

Abstract - San Marzano (SM) is a traditional Italian tomato landrace characterized by red elongated fruits, originated in the province of Naples (Italy) and cultivated worldwide. Here two tomato genes involved in two of the mutations studied previously in the SM background, *Stay Green Protein* (*SGR*, Solyc08g080090) and *Carotenoid Isomerase* (*CRTISO*, Solyc10g081650), were edited with the CRISPR/Cas9 system, with the final aim to compare the phenotypic and metabolic effects of these two mutations, with those obtained by introducing them through the classic backcross breeding scheme, into the Italian traditional variety San Marzano (SM). A high editing efficiency was obtained for both genes (67,8% and 71,4%) and the types of mutations associated with the desired phenotypes were screened in T0 generation. Three edited lines for each gene were chosen for growing the T1 generation. The phenotypic analysis allowed to recognize the essential vegetative, flower and fruit traits of the *gf* and *t* mutation, that were properly recapitulated by the edited mutations fixed in T1. A HS-SPME/GC-MS analysis of the volatile fraction of the tomato pericarps was carried out to explore the volatile compound profile of the mutations under study; the introgressed *t* line showed the most distinct profile from SM, characterized by higher production of several benzenoids, terpenoids and a few apocarotenoids. On the other hand, both the introgressed *t* line and T6 and T13 edited lines showed a depleted production of many fatty acid derivatives, when compared to the wild type San Marzano variety. On the contrary, *gf* edited lines did not show detailed differences under the point of view of the volatile profile.

4.1. Introduction

One of the greatest CRISPR-Cas9 mediated gene editing applications concerns its use on vegetable species. Indeed effective genome editing has been achieved in numerous species including *Arabidopsis thaliana*, *Oryza sativa*, *Nicotiana benthamiana*, *Solanum lycopersicum*, maize, soybean, sorghum, *Citrus sinensis*, *Triticum aestivum*, *Marchantia polymorpha* and Populus, with a wide and detailed investigation of both the occurred mutations and the inheritance pattern (Brooks et al., 2014; Pan et al., 2016).

In this work, two tomato genes, *Stay Green Protein* (*SGR*, Solyc08g080090) and *Carotenoid Isomerase* (*CRTISO*, Solyc10g081650), were targeted using computationally designed gRNAs with the stable transformed CRISPR/Cas9 system, with the final aim to reproduce the phenotypic and metabolic effects of these two mutations into the Italian traditional variety San Marzano (SM).

Indeed fruits of *green flesh* (*gf)* mutant of tomato (*Solanum lycopersicum*) ripen to a muddy brown color due to the accumulation of lycopene coupled with a lack of chlorophyll degradation (Kerr, 1956), a process that normally precedes or occurs concomitantly with *de* novo synthesis and accumulation of other pigments, such as carotenoids, anthocyanins, and flavonoids, at the onset of ripening (Seymour et al., 1993). On the other hand *CRTISO* encodes a redox-type enzyme that convertes cis-lycopene to trans-lycopene; the recessive mutations of *tangerine* (*t*) abolishes its expression in fruits, which turnes orange due to the accumulation prolycopene instead of all-*trans*-lycopene, which is normally synthesized in wild-type fruit (Isaacson et al., 2002).

These two aesthetically novel phenotypes represent an alternative to the red tomato also because of their different nutraceutical and organoleptic qualities. In this regard recent reports have revealed that the enhancement of chloroplast development in the fruit may result in higher contents not only of tomato fruit-specialized metabolites but also of sugars, re-evaluating the importance of the contribution of chloroplasts/photosynthesis in fruit development (Cocaliadis et al., 2014). At the same time a comparison of red and *tangerine* tomatoes allowed to expose the enanched bioavailability of lycopene when consumed as *cis*-isomers from *tangerine* compared to red tomato juice (Cooperstone et al., 2014), justifying the use of *tangerine* tomatoes as a lycopene source with new potential health benefits.

Both these mutations had been previously backcrossed into SM, with four backcross cycles, however our previous study highlighted their different genetic distance from SM, due to the presence of conspicuous regions from the donor parents that has not been possible to recover yet (Chap. 2; Dono et al., 2020). In this work aspects of genotypic, phenotypic and organoleptic qualities were collected in plants of both the *gf* and *t* introgressed and edited lines, compared to SM, with the aim to bring the observed effects back to the mutations themselves.

4.2. Materials and methods

4.2.1. Vector design and construction

Two single guides RNA were designed to target the coding sequence of respectively the *gf1*/*sgr* gene (Solyc08g080090) and *SlCRTISO* (Solyc10g081650) using the design tools on Benchling (https://benchling.com).

The choice of the gRNA was made following a series of requirements:

1. the gRNA would start with a G, a requirement associated with the use of the U6 26 Arabidopsis promoter

2. it would be located in a structurally important region of the cds

3. it had a satisfactory on-target activity and, likely, no off-target effects

The selected sequence was domesticated for cloning into GoldenBraid pDGB3 alpha1 vectors by adding adapters for the insertion of the gRNA between the U6-26 promoter and the scaffold RNA sequences in the assembly, done using the GB CRISPR domesticator available at https://gbcloning.upv.es/. Oligonucleotides were diluted to a final concentration of 1 µM; 5 µl of each were mixed and let anneal at room temperature for 30 min before setting up the multipartite restriction-ligation reaction. All other GB parts used in restriction-ligation reactions were diluted to a concentration of 75 ng µl-1. Such reactions, and the following steps were performed as detailed in Gianoglio (2016). Three successive cloning steps which incorporated the gRNA cassette, the *hCas9* transcriptional unit (TU) and the *NptII* TU.

The two final pDGB3 alpha2 construct Tnos:*NptII*:Pnos - U6-26:gRNA:scaffold - 35S:*hCas9*:Tnos were transformed into the rifampicin-resistant *Agrobacterium*

tumefaciens LBA4404 electrocompetent cells by electroporation. Cells were then resuspended in 500 µl SOC *medium* and let grow for 2 h at 28°C with stirring, then plated on LB-agar *medium* supplemented with 50 µg ml-1 rifampicin and 50 µg ml-1 kanamycin, as specified by vector resistance, and incubated for 3 d at 28°C. Colonies were grown in LB liquid *medium* supplemented with rifampicin and kanamycin, with agitation for 2 days at 28°C. The presence and identity of the plasmid were confirmed by restriction using at least two restriction enzymes.

4.2.2. Plant material

Tomato 'San Marzano' seeds were sterilized by washing for 30 min in a 2.5% sodium hypochlorite solution, then rinsed for three times in sterile water, with each rinse lasting for 5, 10 and >15 minutes, respectively. Sterile, clean seeds were transferred to a solid germination *medium* (2.5 g l^{-1} MS vitamins, 10 g l^{-1} sucrose and 10 g l^{-1} phytoagar, pH 5.8) in sterile cups and kept in the dark for 3 days before being exposed to light. Growth chamber conditions provided a 16 hours day light cycle (250 µmol photons m^{-2} s^{-1}), a relative humidity of 60-70% and 25°C. Ten days after sowing, cotyledons were used for *Agrobacterium*-mediated transformation. Cotyledon development was considered optimal for transformation when the first true leaves started to emerge. 100 seeds for each transformation were used.

4.2.3. Agrobacterium-mediated transformation

Two *Agrobacterium* LBA4404 pre-culture, containing the two CRISPR/Cas9 constructs, were set up 48 hours before transformation in MGL liquid *medium* pH 7 supplemented with antibiotics (Table 4.1) and incubated overnight at 28°C with agitation. Twenty-four hours before transformation, from this pre-culture a second culture was set up in TY liquid *medium* pH 5.8 supplemented with 200 µM acetosyringone with no antibiotics (Table 4.1) and incubated overnight in the dark at 28°C with agitation. Before transformation, the optical density (OD) of the bacterial culture was measured at 600 nm and the culture diluted to a final OD of 0.10-0.12 in TY *medium* with 200 µM acetosyringone. Explants of about 5 mm in length were cut from the cotyledons, dipped in the bacterial culture for 5

minutes, then blotted dry on filter paper and transferred for 48 hours to a co-culture *medium* in the dark.

Table 4.1. Composition of MGL and TY *media* used for *Agrobacterium* growth.

Compound	MGL *medium* pH 7	TY *medium* pH 5.8
Tryptone	5 g l-1	5 g l-1
Yeast extract	2.5 g l-1	3 g l-1
NaCl	-	-
5 g l-1	-	-
Glutamic acid	1.02 g l-1	-
K2HPO4	0.25 g l-1	-
MgSO4.7H2O	0.1 g l-1	0.5 g l-1 (2 mM)
Biotin	1 mg l-1	-
Autoclave		
Acetosyringone	-	200 µM
Kanamycin/Spectinomycin	50 mg l-1	-
Rifampicin	50 mg l-1	-

4.2.4. Organogenesis and regeneration

The regeneration of plantlets from the transformed cotyledon explants was performed according to the method described by Qiu *et al.* (2007), with modifications. Briefly, after 48 h of co-culture with *A. tumefaciens*, explants were grown on a *medium* to induce the formation of *callus* and shoots (induction *medium*). Explants were moved to a fresh *medium* every 21 days, or when explants size or shoot formation required it. Shoots were then transferred to an elongation *medium* and finally moved to a rooting *medium* (Table

4.2). Kanamycin selection was maintained at all stages of regeneration. Plants were grown *in vitro* to a size of 5-8 cm and then were moved to soil in the greenhouse, where they were gradually acclimated to environmental growing and humidity conditions.

Table 4.2. *Media* composition for co-culture, induction of *callus* and shoot formation, elongation and rooting. IAA = indoleacetic acid, IBA=indolebutyrric acid, ZR=zeatin riboside.

Compound	**Co-culture *medium***	**Induction *medium***	**Elongation *medium***	**Rooting *medium***
MS vitamins	5 g l^{-1}	5	5	2.5
Sucrose	30 g l^{-1}	30	30	10
Plant agar	6 g l^{-1}	6	6	6
IAA	0.1 mg l^{-1}	0.1	0.1	-
IBA	-	-	-	0.1
ZR	2 mg l^{-1}	2	2	-
Acetosyringone	200 µM	-	-	-
Carbenicillin	-	500 mg l^{-1}	500	500
Kanamycin	-	100 mg l^{-1}	100	30

4.2.5. DNA extraction and genotyping of the T0 and T_1 generation

DNA was extracted from leaves of plantlets growing *in vitro* according to the CTAB method.

Each DNA sample consisted of different leaves from the same plant. Tissues were collected from a total of 28 *gf* and 14 *t* individuals regenerated from different *calli*. The quality of the extracted genomic DNA was assessed by running it on a 1% (v/v) agarose gel.

Two pairs of primers (*gf1* forward and *gf1* reverse, *t* forward and *t* reverse, Table 4.3) were designed using the Primer3 software (http://primer3.ut.ee/) to amplify a region surrounding the CRISPR target site. The resulting amplicon would be within 700-900 base pairs in size and the expected Cas9 cut site would be located at 250-300 bp downstream from the 5' end of the fragment, as specified by the instructions at https://tide.deskgen.com/. The selected *gf* and *t* primers amplify respectively an 866 and a 746 bp region. Primer specificity was first predicted using the BLAST tool available at https://solgenomics.net/tools/blast/. A second pair of primers was designed to amplify a 700 bp fragment on the *NptII* gene to assess the presence of the transgene in regenerated plantlets (*NptII* forward and *NptII* reverse, Table 3). The PCR reaction mix composition and cycling conditions are listed in Tables 4.4 and 4.5, respectively.

Table 4.3. Primers for gf1 and *t* genotyping and NptII amplification

	Target	**Primer sequence 5' → 3'**	**Tm °C**
	***gf1* forward**	GGGTCTGGGCCAAAACTACT	59
	***gf1* reverse**	ACAGGCACAAGCCTCTTCAC	59.4
	***t* forward**	GCTTTGGGTGATAGCAAACC	59.6
	***t* reverse**	CACCATGGCCAGTATAAACTATG	58.5
	***NptII* forward**	GAGGCTATTCGGCTATGACT	57
	***NptII* reverse**	ATCGGGAGCGGCGATACCGT	63

Table 4.4. GoTaq® PCR reaction mix for genotyping and T-DNA amplification.

GoTaq® PCR reaction set up	Volume
2x GoTaq®Master Mix	5 µl
Template	As required
Primer forward 20 µM	1 µl
Primer reverse 20 µM	1 µl
Water	2 µl

Table 4.5. PCR cycling conditions for GoTaq® polimerase reactions

Step	Temperature	Time	Cycles
Initial denaturation	**95°C**	**5 minutes**	**1**
Denaturation	**95°C**	**30 seconds**	**35**
Annealing	**as required**	**30 seconds/kb**	
Extension	**72°C**	**30 seconds**	
Final extension	**72°C**	**5 minutes**	**1**
Hold	**4°C**	**∞**	**1**

Subsequently, the *gf* and *t* PCR products of respectively 28 and 14 individuals, plus 2 negative individuals and the wild type control, were purified using NucleoSpin® Gel and PCR Clean-up by Macherey-Nagel and directly sequenced through Sanger's method. Each PCR product represented a population of molecules comprising the different alleles present in the plant's genome after editing. The resulting chromatograms were analyzed using the TIDE software (https://tide.deskgen.com/), which performs a decomposition of the chromatogram signals around the expected cut site and calculates allele frequencies (Brinkman *et al.*, 2014).

Seeds were collected from T0 fruits, lines GF2, GF5 and GF28 for *green flesh* and T6 and T13 for *tangerine* were chosen to analyze the T_1 generation, based on the fruit phenotype in alignment with that expected. 20 T1 seeds per line were sowed, including the introgressed *green flesh* and *tangerine* lines, plus San Marzano. Finally, 64 plants were kept in the greenhouse: 8 from each line. Eight plants per accession at the 4–5th true leaf stage were transplanted in twin rows (100 cm between twins, 60 cm between rows and 50 cm between plants within the row), divided in two blocks, in an unheated tunnel located at the University of Tuscia's Experimental Farm at Viterbo, Italy (42°260'N, 12°040'E). Plants were grown following standard cultural practices for indeterminate tomatoes, using tutors and weekly removal of lateral shoots. Daily temperature was maintained below 30 °C by a ventilation system and the plants were irrigated through a drop irrigation system.
A PCR was performed on the *NptII* gene to assess whether or not the transgene had segregated. The genotyping of the *gf1* and *t locus* were performed following the same protocol used for the T0 generation.

4.2.6. Phenotypic evaluation of T1 compared with the correspondent introgressed lines, with respect to SM

The T0 regenerants growing in a growth chambers were inspected for visible phenotypes relating to the *gf* and *t* mutations. In particular, color of fruits, corolla and petals were considered indicators of fruit specific phenotypes. Phenotypic analysis of T1 lines was done as reported in Dono et al., 2020. Aspects of vegetative (chlorophyll, flavonoid and anthocianin content), of reproductive (flowering, number of flowers per inflorescence and pollen viability) and of fruit quality (brix°, shelf life and wrinkling) were collected.

4.2.7. Fruit sampling and volatile detection and quantification

Fruit sampling and volatile detection and quantification was done as previously reported in Chapter 3, paragraphs *2.2* and *2.3*. For each line, eight fruits were harvested at full ripening. Two biological replicates for genotype, each represented by four fully ripe berries, were harvested over a period of three days.

4.3. Results

4.3.1. Target selection and vector assembly

A number of spontaneous mutations are known for both *SGR* and *CRTISO* loci (Barry and Pandey, 2009; Isaacson et al., 2002) and all have the same effects on plant and fruit phenotype. It was chosen to design two guides RNA specifically targeting a structurally important region of both the *SGR* and *CRTISO* coding sequences of the proteins, with the aim to obtain loss of function variants. Regarding *gf,* the gRNA was selected to target the third exon, with the Cas9 cut site at position 1749, in correspondence with the conserved central core of the protein structure. For *t*, a gRNA was selected to target the second exon, with the Cas9 cut site at position 972 pb (Figure 4.1a-4.1.b).

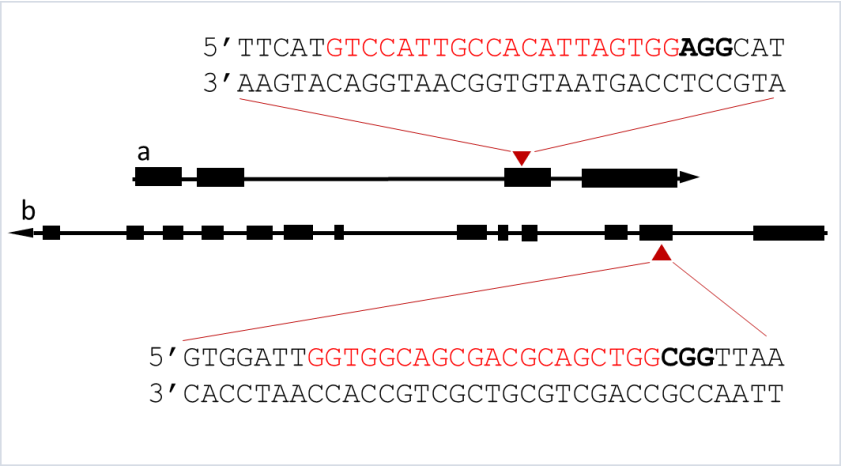


Figure 4.1. Gene structure of (a) *gf* and (b) *t loci.* Exons are identified with black rectangles and gRNA targets with red triangles. gRNAs sequences are reported in red, PAM sequence in bold black.

4.3.2. green *flesh* and *tangerine* mutants were efficiently obtained in San Marzano

Twenty-eight *green flesh* and 14 *tangerine* T_0 plants regenerated on selective culture media were selected to be analyzed after confirming that they had integrated the transgene, through PCR amplification of the *NptII* gene (not shown). Genotyping was performed by PCR amplification of the target locus and the resulting chromatograms were analyzed using TIDE (https://tide.nki.nl/). Overall, the average gene editing efficiency was 67.8 % for *green flesh* and 71.4 % for *tangerine.* The most frequent *green flesh* allelic assets were biallelic (GF 1, GF 3, GF 4, GF 16, GF 19, GF 20, GF 21) and homozygous edited events (GF 2, GF 5, GF 7, GF 9, GF 10, GF 11, GF 13), each representing 36,4% of the transformed plants. 26.3 % of edited plants (GF 6, GF 12, GF 17, GF 27, GF28) were heterozygous at the target *locus*, meaning that they retained a wild type allele (Figure 4.2a). Edited alleles ranged in size from an insertion or deletion of 1 nucleotide to a -5 bp deletion (Figure 4.3a). The most frequent mutation was a single nucleotide insertion, found in 16 individuals; T was the inserted nucleotide in nine edited plants, A in five and G in two (Table 4.6). No insertions larger than one nucleotide were found. Lines GF2, GF5, GF28 were chosen for studying the T1 generation.

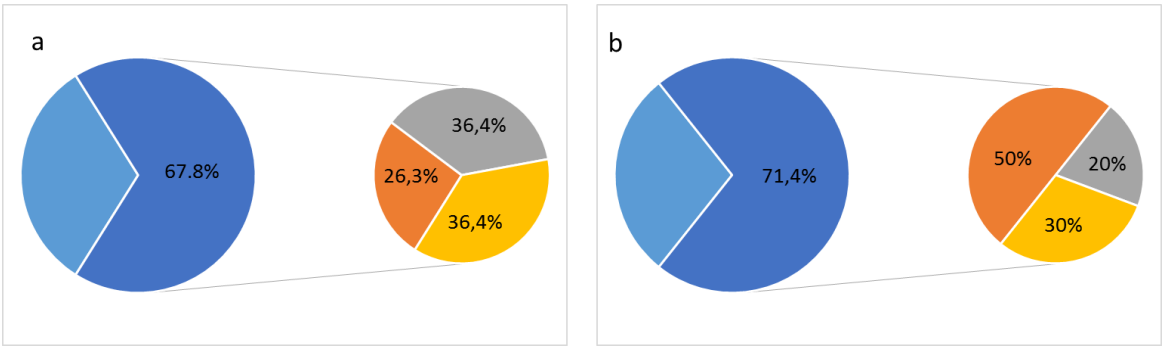


Figure 4.2. Percentage of transformed, biallelic, homozygous, heterozygous and chimeric events for *gf* (a) and *t* (b).

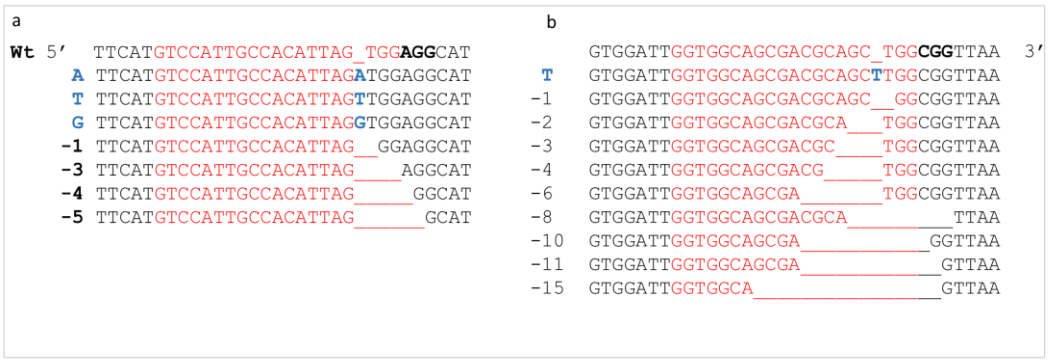


Figure 4.3. Type of Crisper-Cas9 edited alleles for *gf* (a) and *t* (b).

Table 4.6. Genotyping of 28 T0 regenerant transformed with the *green flesh* editing construct by PCR amplification of the target locus, direct sequencing and TIDE chromatogram decomposition. The R2 value indicates the percentage of variance explained by the model. Overall efficiency is calculated as the difference between R2 and the amount of wild type allele. Values in the remaining columns indicate the percentage of each particular allele.

SAMPLE	Overall efficiency	R^2	WT	TYPE OF INSERTION					
				INSERTION		**DELETION**			
				+1	**Inserted base**	**-1**	**-3**	**-4**	**-5**
GF 1	97.7	0.98		47.7	A				49.7
GF 2	96.3	0.98		86.1	G				
GF3	95.9	0.97		46.6	T	48.6			
GF4	91.4	0.91		39.6	T	41.1			
GF5	97.5	0.98		93.9	A				
GF6	13.8	0.98	83.7	12.5	G				
GF7	99.3	0.99						97.1	
GF8	5.7	0.99	93.0						
GF9	90.5	0.98		85.0	T				
GF10	95.9	0.99		91.1	T				
GF11	94.9	0.98		90.2	A				
GF12	87.0	0.97	10.2	81.7	A				
GF13	97.5	0.99		89.4	T				
GF14	5.7	0.99	93.2						
GF15	6.3	0.99	92.5						
GF16	83.0	0.83		41.1	T	22.3			
GF17	21.0	0.94	72.7	7.2	A				
GF18	17.5	0.94	76.9						
GF19	90.4	0.9		39.0	T		38.7		
GF20	94.6	0.95		57.6	T	19.2			
GF21	94.3	0.95		45.3	T	45.9			
GF22	17.4	0.97	79.7						
GF23	12.0	0.96	84.5						
GF24	23.7	0.96	72.0						
GF25	15.6	0.95	79.1						
GF26	9.8	0.97	87.1						
GF27	26.9	0.97	69.8	14.7					12.2
GF28	31.1	0.93	62.4	13.0					8.3

The most frequent *tangerine* allelic asset was chimeric (T4, T7, T8, T13, T14), followed by heterozygous (T2, T6, T11) and homozygous (T3, T9) events (Figure 4.2b). Edited alleles ranged in size from an insertion or deletion of one nucleotide to a -15 bp deletion (Figure 4.3b). The most frequent mutation was a single nucleotide insertion of T, as well as -1 and -10 deletions, each found in four plants. Deletion -10 was found with the highest percentage between the alleles, other large deletions were found at lower frequencies (Table 4.7). Lines T6 and T13 were chosen for studying the T1 generation.

Table 4.7. Genotyping of 14 T0 regenerants transformed with the *tangerine* editing construct by PCR amplification of the target locus, direct sequencing and TIDE chromatogram decomposition. The R2 value indicates the percentage of variance explained by the model. Overall efficiency is calculated as the difference between R2 and the amount of wild type allele. Values in the remaining columns indicate the percentage of each particular allele.

SAMPLE	Overall efficiency	R^2	WT	TYPE OF INSERTION										
				INSERTION		**DELETION**								
				+1	**Inserted base**	**-1**	**-2**	**-3**	**-4**	**-6**	**-8**	**-10**	**-11**	**-15**
T1	1.9	0.99	99.0											
T2	99.2	0.99										98.4		
T3	99.3	0.99	60.2									98.9		
T4	98.7	0.99		22.2	T	15.1	9.9	51.6						
T5	2.7	1.0	96.8											
T6	99.4	0.99										97.6		
T7	96.0	0.96							13.7	11.4		58.9		
T8	96.7	0.97		32.6	T			32.6		31.5				
T9	65.4	0.96	30.7	24.6	T						23.7			10.7
T10	13.6	0.99	85.1											
T11	99.5	1.0				98.4								
T12	2.5	0.98	95.8											
T13	86.4	0.94		21.0	T	10.0		8.1	6.0		28.6			
T14	96.2	0.96		25.8		26.7	25.8	12.5					5.4	

4.3.3. Segregation of the transgene and of edited alleles in T1 progenies

Three *green flesh* (GF2, GF5, GF28) and two *tangerine* (T6, T13) T0 plants were selected for phenotypic characterization and eight plants of each line were grown. Their progeny was analyzed for transgene segregation and a total of respectively 4 *gf* and 2 *t* edited Cas9-free T1 plants were identified (data not shown). Segregation of the transgene occurred with different patterns between the edited lines: one GF2 and three GF5 segregated the transgene, but none of the GF28 resulted to be transgene-free. Regarding *tangerine*, only two T13 progeny plants were *NptII*-free. Analysis of edited alleles in T1 plants showed that *green flesh* did not confirm the mutations of the T0 generation, with GF2 which changes from +1 insertion to -5 deletion and GF28 with a very big supposed deletion; further analysis will be carried on in the T2 generation. *tangerine* mutations were stable and inherited in a Mendelian fashion, with T6 lines that fixed the -10 deletion at the homozygous state and T13 progenies carrying -8 and -4 deletions (Figure 4.4).

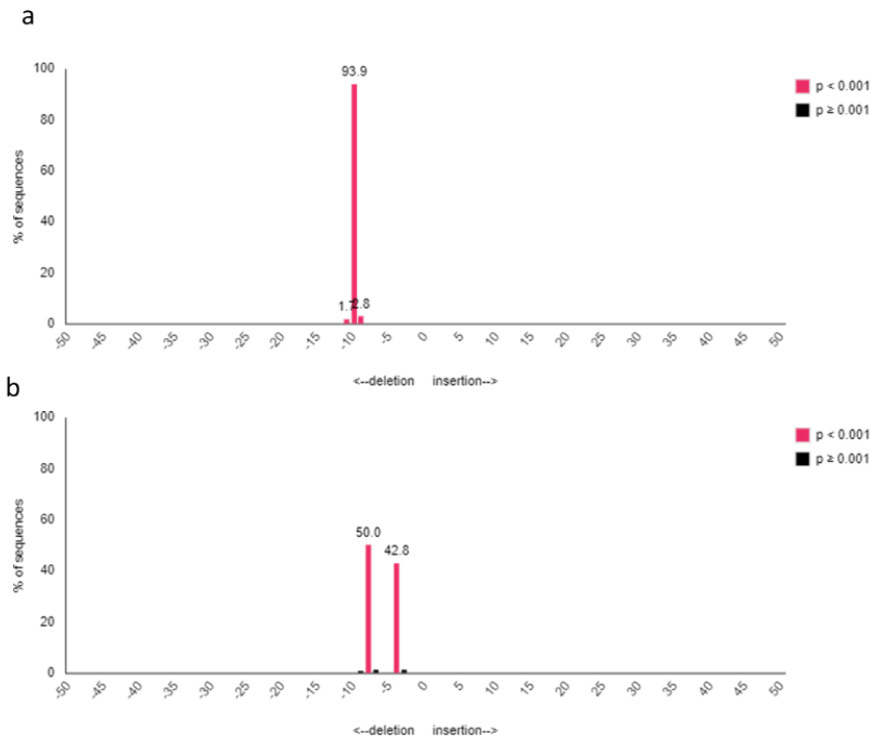


Figure 4.4. Tide graphics corresponding to (a) a -10 deletion allele and (b) two alleles, -8 and -4 deletions, inherited respectively in a T6 and in a T13 edited T1 *t* plants.

4.3.4. Phenotypic traits of T1 plants edited for *green flesh* and *tangerine*

Essential vegetative, flower and fruit traits of the *gf* and *t* mutation were properly recapitulated by the edited mutations fixed in T1 (Figure 4.5).

Figure 4.5. Details of leaves, flowers and fruit from SM, edited *t* and edited *gf* (Figure 4.6 h), confirmed by the higher resistance to wrinkling in GF5 and GF28 (data not shown).

WT plants showed the following data for vegetative aspects: a leaf CHL content of 33.04 µg/cm^2, a FLAV value of 0.63 µg/cm^2 and the ANTH average was 0.29 µg/cm^2. Regarding the reproductive aspects, WT plants produced the first flower about 38 DAT and had NF equal to 7. Lastly two parameters for fruit quality were collected: the WT Brix° was 6.3. Analyzing the behaviour of both introgressed and edited *green flesh* lines, we found that GF28 plants showed CHL levels lower than the WT and *gf* was higher for the ANTH content (Figure 4.6 a-c). For reproductive aspects all the edited *green flesh* plants showed a delay in FLOW compared to the WT and only *gf* lines had a higher NF (Figure 4.6 d-e). PV did not

show any variations. Interestingly, GF5 reported higher Brix° values compared to all the other values. Furthermore, all the edited *green flesh* plants showed higher SL than the WT

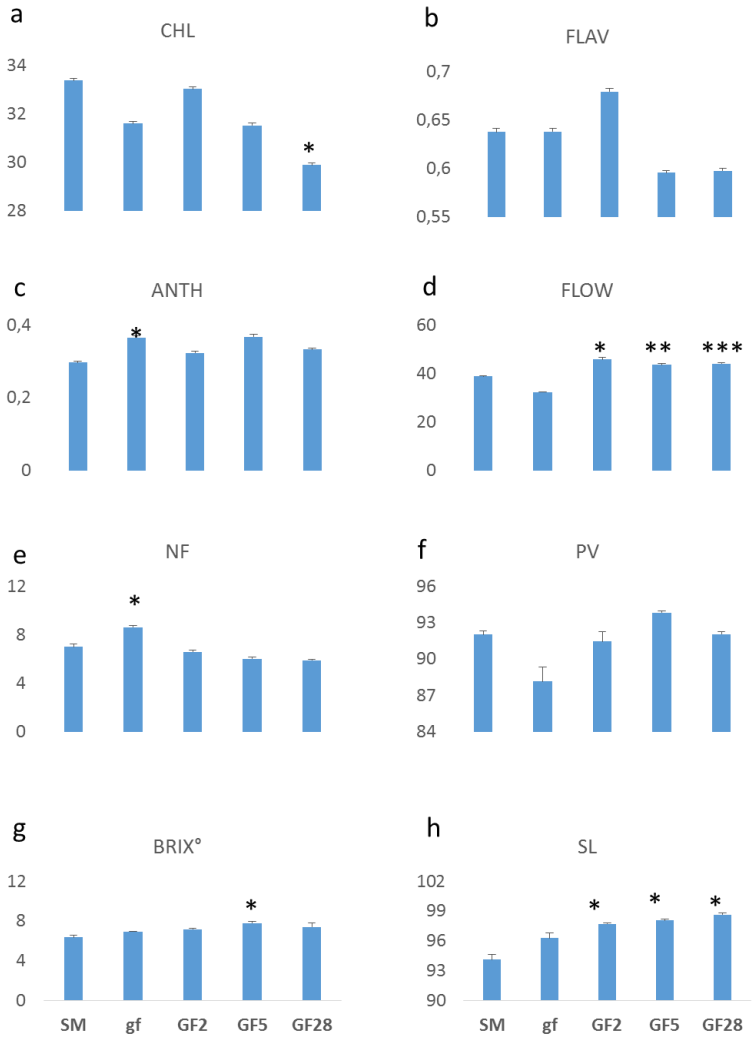


Figure 4.6. Variations in (a) chlorophylls, (b) flavonoids, (c) anthocyanins, (d) flowering, (e) numbers of flowers per inflorescence, (f) pollen viability, (g) soluble solids, (h) shelf life in *gf* introgressed line and GF2, GF5 and GF28 CRISPR lines in San Marzano (SM) background compared with the recurrent parent. Means significantly lower and higher than SM indicated with * for $0.01 \leq P \leq 0.05$, ** for $0.001 \leq P \leq 0.01$ and *** for $P \leq 0.001$ after Student's *t* test respectively.

Focusing on *tangerine*, T6 and T13 had lower CHL content, as expected. *t* together with T13 and T6 coupled with T13 had respectively higher FLAV and ANTH levels. For reproductive aspects, t and T16 reported respectively early and delayed FLOW and no differences for NF. Finally, T6 showed a higher Brix° value than the WT, in contrast with that lower one reported for *t* (Figure 4.7).

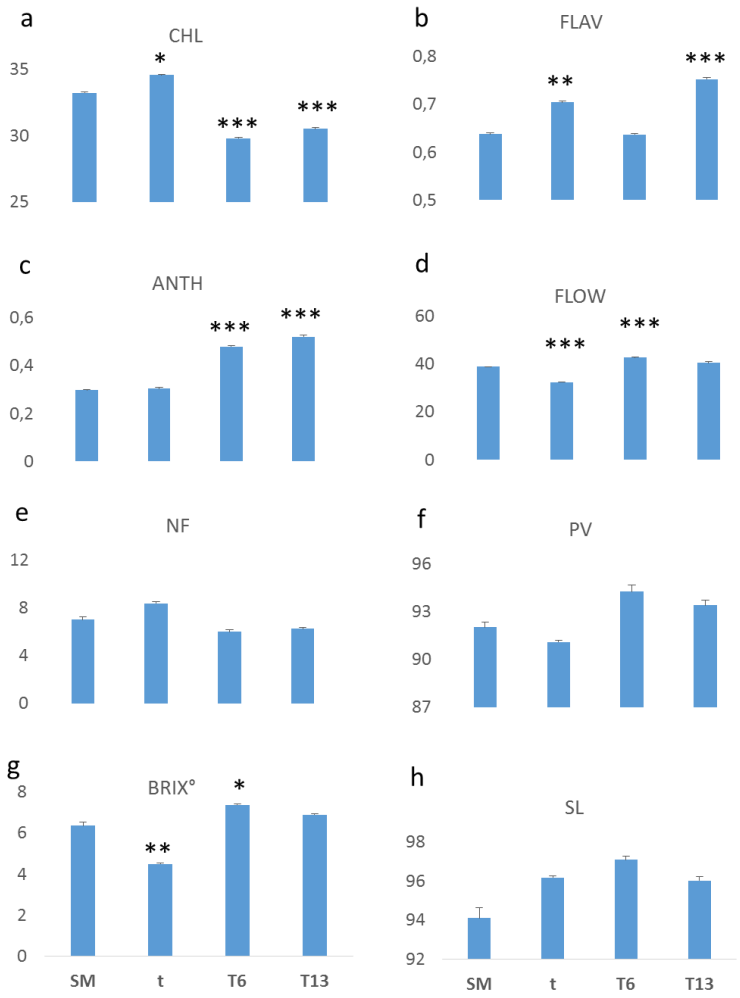


Figure 4.7. Variations in (a) chlorophylls, (b) flavonoids, (c) anthocyanins, (d) flowering, (e) number of flowers per inflorescence, (f) pollen viability, (g) soluble solids, (h) shelf life in *t* introgressed line and T6 and T13 CRISPR lines in San Marzano (SM) background compared with the recurrent parent. Means significantly lower and higher than SM indicated with * for $0.01 \leq P \leq 0.05$, ** for $0.001 \leq P \leq 0.01$ and *** for $P \leq 0.001$ after Student's *t* test respectively.

4.3.5. Volatile compound profile of edited and mutated lines compared with SM

In order to detect the differences in flavour between the introgressed and edited *green flesh* and *tangerine* mutated lines in comparison with the wild-type SM, a HS-SPME/GC-MS analysis of the volatile fraction of the tomato pericarps was carried out, as described in chapter 3. The analytical strategy chosen for the analysis allowed the unequivocal relative quantification of 69 volatile compounds (VOCs) by comparison of both mass spectra and retention index with those of authentic standards. Thirty-three of them showed statistically significant differences in at least one of the lines under study (Table 4.8).

PCA was performed based on the volatile profile of the genotypes involved. *gf* volatile profile did not show any different trend with respect to the WT. For *t*, the first two principal components, together explaining about 42% of the total variance, revealed that each of the genotypes has a particular volatile profile. The introgressed *t* line showed the most distinct profile, which was characterized by higher production of several benzenoids, terpenoids and a few apocarotenoids. On the other hand, both the introgressed *t* line and T6 and T13 edited lines showed a depleted production of many fatty acid derivatives as compared to the wild type San Marzano variety (Figure 4.8).

Table 4.8. Means of volatile compounds for the fold change of individual values over the San Marzano value; means in bold indicate values significantly higher and lower than SM for Student's t-test.

Identified numbers	Metabolic pathway	Compounds	gf	GF2	GF28	GF5	t	T13	T6
1	B	Benzaldehyde	1.690	1.106	2.079	0.690	**0.380**	0.912	3.014
2	B	Benzylalcohol	**1.855**	0.619	2.480	1.182	**0.925**	3.211	1.668
3	B	Salicylaldehyde	1.053	0.544	0.847	0.635	**1.408**	1.580	0.780
4	B	Methyl salicylate	1.732	3.371	**1.274**	1.793	1.309	1.203	1.018
5	B	Ethyl salicylate	3.998	**13.175**	0.980	0.395	3.567	19.361	4.484
6	B	Eugenol	2.066	0.624	0.816	0.976	**9.382**	1.063	1.547
7	BCAA	2-methylbutanol	**5.717**	**3.640**	**4.448**	**3.290**	2.200	**4.347**	0.751
8	C	beta-cyclocitral	0.913	0.370	**0.543**	0.541	1.066	0.836	1.012
9	C	Geranylacetone	0.778	**0.212**	0.843	0.786	**6.431**	**0.520**	2.063
10	C	beta-ionone	0.446	**0.297**	**0.193**	**0.468**	1.086	0.340	0.899
11	E	2-methylpropyl acetate	**20.981**	11.789	10.128	29.595		2.289	20.607
12	E	2-methylbutyl acetate	**1858.154**	790.192	1206.500			**724.692**	**1005.974**
13	L	1-penten-3-one	0.659	0.109	1.187	0.482	**0.937**	0.831	0.480
14	L	Pentanal	2.107	1.016	1.559	1.977	**0.933**	**1.037**	**0.882**
15	L	2-ethylfuran	0.458	0.743	0.862	**0.379**	1.115	0.999	0.955
16	L	1-pentanol	5.576	2.927	7.762	5.008	**1.666**	**1.936**	**2.634**
17	L	(Z)-2-penten-1-ol	1.215	0.775	0.712	1.235	0.909	**0.682**	0.389
18	L	(Z)-3-hexenal	1.275	0.083	1.500	**0.086**	0.669	**1.406**	0.655
19	L	Hexanal	0.950	**0.425**	0.746	**0.454**	0.750	0.295	0.096
20	L	Hexanoic acid	1.649	1.781	0.962	**0.752**	1.577	**0.727**	**1.155**
21	L	(E)-2-heptenal	2.710	**1.699**	1.547	0.692	0.174	**1.213**	0.121
22	L	2-pentylfuran	0.246	**0.553**	**0.045**	0.365	0.412	**0.229**	0.359
23	L	(E)-2-octenal	1.396	**1.017**	1.537	0.574	**0.618**	1.027	1.220
24	L	(E,Z)-2,4-decadienal	2.300	**1.223**	0.777	1.923	**0.360**	0.441	1.412
25	L	(E,E)-2,4-decadienal	1.139	**0.639**	**0.519**	0.976	**0.257**	**0.421**	**0.220**
26	Phe	Phenylacetaldehyde	0.450	1.047	**0.174**	1.127	0.253	0.624	1.613
27	Phe	Benzylnitrile	0.155	1.377	**0.855**	2.049	1.170	1.144	0.391
28	Phe	1-nitro-2-phenylethane	**0.149**	**0.089**	**0.089**	**0.250**	0.512	**0.341**	**0.172**
29	S	Methanethiol	0.600	**0.574**	0.432	0.734	**0.377**	0.518	**0.782**
30	T	(Z)-linalool oxide	1.171	1.110	2.262	1.708	**1.707**	1.350	1.355
31	T	(E)-linalool oxide	1.301	0.148	0.852	1.016	**0.843**	0.835	1.100
32	T	Ocimenol	0.775	0.361	1.214	**1.042**	1.622	1.100	1.009

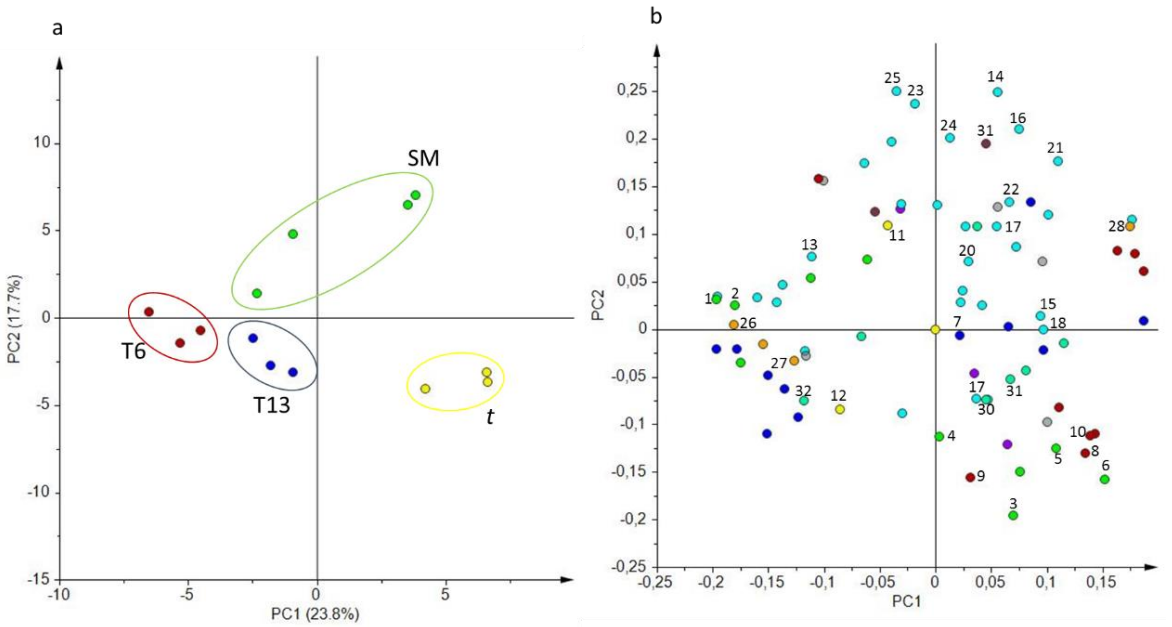


Figure 4.8. PCA of log_2 values of 85 volatile compounds measured by HS-SPME/GC-MS. (**a**) PC1 X PC2 score plot of the *tangerine* mutated lines compared with the T6 and T13 edited lines, plus San Marzano (SM). (**b**) PC1 X PC2 loading plot. Color symbols in the loading plot (**b**) correspond to benzenoids in green, branched chain amino-acids in blue, apocarotenoids in red, esters in yellow, fatty acid derivatives in light blue, phenylalanine derivatives in orange, sulphur compounds in brown, terpenoids in water green.

4.4. Discussion

4.4.1. Response of San Marzano to transformation and regeneration protocols

An efficient and reproducible regeneration protocol, yielding a high percentage of regenerants has been standardized for tomato (Sharma et al., 2014). Additionally, it is known that tomato cotyledons are among the tissues with the highest regenerative potential (Moghaieb et al., 1999). The growth medium, as well as environmental conditions of light and temperature also played an important role; its winning balance of phytohormones and antibiotics easily induced callus formation and the ensuing organogenesis process (Gerszberg et al., 2019). The union of these elements allowed us to regenerate efficiently and to obtain a performant percentage of trasformants.

4.4.2. Editing efficiency and pattern

We obtained 67.8% and 71.4% of edited plants for *green flesh* and *tangerine*, respectively. Our data reported an editing efficiency higher than 90%, while the editing efficiency within each plant ranged from 8.3 to 93.9 for *gf* and from 5.4 to 98.9 for *t*, as explained by the R^2 model. This is consistent with previous findings; Brooks et al. (2014) reported an average efficiency frequency in tomato of 75% and Pan et al (2016) of 83.56%. The latter authors also confirmed that the T0 mutations can be stably transmitted through the germ line, in line with our T1 *tangerine* generation data, which fixed the alleles obtained in the T0. This did not happen for *green flesh* and further analysis will be carried on in the T2 generation to clarify this aspect.

Another debated aspect concerns the types of mutations introduced in plants by gene editing, indeed the breakpoint introduced by Cas9 is usually at 3 nucleotides upstream of the PAM (however it can be found even downstream). In plants, insertion of one nucleotide or deletion of 1-10 nucleotides are the most frequent mutations, according with our results; larger deletions of 20-150 nucleotides are rarer but do occur (Bortesi et al., 2016). Indeed in our *gf* edited plants, a single base insertion was the most common type of mutation, with T being the predominant inserted base. A similar result was reported by Gianoglio (2018), founder of the two CRISPR-Cas9 constructs we used, in Money Maker tomato variety, highlighting that the achievement of the result is “genotype-independent”. For *tangerine* the predominant editing event was a -10 deletion. Cas9 is able to continue acting during plant growth and development, driven by the strong constitutive promoter 35S; no changes can appear in homozygous and biallelic mutants which have already lost the target site, but heterozygous may still be modified in the course of the plant growth. A consistent number of chimeras and polyploids have occurred in the *tangerine* edited lines, supposedly as major drawback for transformation and regeneration protocols, in response to *in vitro* culture, as reported for many different species (Van den Bulk et al., 1990). Alleles with the least editing efficiency were lost in the T1 generation, fixing the lines. Finally the segregation of the transgene occured with different patterns between the edited lines; the hypothesis already advanced for this by Gianoglio (2018) is that the number of copies of the transgene which had been integrated in the plant genome is different, making it easier or not to eliminate it. Our results add to the countless events already present in the literature that describes the CRISPR-Cas9 technique, here specifically delivered through

GoldenBraid modular assembly, as an efficient modification tool for the transformation of plants.

4.4.3. Effects of both the introgressed and edited *gf* and *t* alleles on the plant phenotype

Both introgressed and T1 edited *green flesh* and *tangerine* plants were grown and phenotypic data were collected. All *green flesh* plants reported lower leaf CHL content than WT, and GF28 had a statistically significant inferior mean (at an early growth stage). This is consistent with our previous results (Dono et al., 2020) and with the knowledge that this class of type C stay-green mutations are unable to degrade chlorophyll in the senescence phase of the fruit ripening (Hortensteiner et al., 2009), but do not increase chlorophyll synthesis. No changes occurred in the FLAV and ANTH leaf content in the edited lines as expected, except for the higher ANTH level in *gf*; since this aspect is not mantained in the edited plants, we can conclude that it is mainly due to the effect of the genetic background. A delay in flowering occured in the three edited lines, a phenotype not reported before in *gf* mutants. We already observed the higher NF in *gf* (Chap. 1), which is likely an effect of the genetic background since both other double mutants involving the same gene and the edited lines had a NF comparable to SM. The content of soluble solids was measured on fruits; our introgressed line never reported higher Brix° values that the WT, but we interestingly obtained this result in GF5. Finally we fouond longer SL for the three edited lines, confirming that this is strictly associated with the edited alleles. The introgressed line has never given this result, neither along nor even coupled with other mutations (Dono et al., 2020), meaning that its ability to acquire a longer SL was hidden by the genetic background. In support of our results, chlorophyll content in tomato fruits was shown to be related to reduced oxidative stress and consequently longer shelf-life. In tomato a high chloroplast content in green fruits leads to ripe tomatoes with more active chromoplasts, which in turn accumulate more antioxidants and metabolites which positively affect organoleptic and nutritional quality (Cocaliadis *et al.*, 2014). On this basis, we believe that the higher Brix° value in GF5 could be allele-specific and related with a still active photosynthetic apparatus, being able to produce more sugars than usual.

Shifting the attention on *tangerine* genotypes, the decrease in leaf CHL content, always quite lower in *tangerine* that in the WT, was in opposition with the higher leaf FLAV and ANTH levels. Indeed, the *tangerine* mutation is known to affect carotenoid biosynthesis also in chloroplasts, leading to a different proportion in xhantophylls and lutein with consequent effects on the color of leaves and flowers (Isaacson et al., 2002) (Figure 4.5). FLAV was higher only for *t* and T13, while ANTH was higher in both T6 and T13 edited alleles, suggesting that this feature can be considered "allele-dependent". Moreover, FLAV results confirmed our previous ones for *t*, giving additional support to further investigate how the variation in one class of pigments can influence the levels of other classes. The NF did not show any differences, but *t* had lower FLOW and Brix°, highlighting that several undesiderable traits can represent an uncomfortable drawback in backcross schemes. The CRISPR-Cas9 technique overcomes them, as it happened for the T6 edited line, which had higher FLOW and Brix°. Certainly, more screening on the subsequent edited progeny will allow us to ensure that the phenotypic traits observed are actually due to the edited alleles.

4.4.4. Characterization of the volatile compound flavour of the introgressed and edited *gf* and *t* alleles

The *green flesh* edited lines did not show any specific trend in the pattern of volatile compounds in the ripe fruit, neither in the introgressed and edited lines, nor in comparison with SM; therefore does not seem to be affected by *gf*, at least in relation to aroma compounds. Further studies will be done on primary metabolites in order to detect changes influencing fruit taste, since increased levels of shacarides have been reported in the literature.

The main source of variation causing the dissimilarity of the introgressed *tangerine* has to be attributed to the Benzenoids Eugenol and Salicylaldehyde, and to the Terpenoids (E)-linalool oxide and (Z)-linalool oxide. Nevertheless, this was not observed in the edited lines, which lead to the conclusion that, despite the four backcrosses, part of the recurrent parent genome has not been recovered yet, mainly due to the genetic distance between SM and the *tangerine* donor parent. Furthermore we observed that the Apocarotenoid class of VOCs did not reveal any differences between the *t* lines, compared to the WT, except for higher levels of Geranylacetone in *t* and T13. The *CRTISO* gene is essential for the correct

formation of cyclized carotenes and xhantophylls, indeed t*angerine* accumulates prolycopene instead of all trans-lycopene (Isaacson et al., 2002). The carotenoid cleavage dioxygenases *LeCCD1A* and *LeCCD1B* have been proved to cleave multiple carotenoids after ζ-carotene, both linear and cyclic, producing a variety of volatiles such as β-ionone, geranylacetone, and pseudo-ionone (Simkin *et al.*, 2004), so it seems that the diverse content of *cis* and *trans* lycopene in *tangerine* had no effect in the apocarotenoid formation, which levels positively correlate with tomato flavour acceptability (Vogel et al., 2010).

Finally, we had a decrease in the majority of fatty acid derivatives in the *tangerine* lines. This controversial result allows us to advance the hypothesis of a crossed-talk between the carotenoid and fatty acids biosynthetic pathways. Further metabolic analysis on the *tangerine* lines is still necessary to evaluate the levels of linoleic and linoleic acids, which are the most important substrates for the production of these volatile compounds.

4.5. Conclusions

In this thesis, CRISPR/Cas9 was used to reproduce two mutations in San Marzano by using vectors based on Golden Braid strategy. MAS and transgenic breeding have become the two major procedures of modern plant breeding schemes. Despite the enormous advantages, there are still the limitations, mainly due to the laborious and time-consuming processes, as well as an unavoidable introgression of closely-linked undesirable traits from donor parents. The advent of gene editing technologies, especially the CRISPR/Cas9 system, offer an extremely efficient and simple customization process.

A deep description of the *gf* and *t* organoleptic had not been thoroughly performed yet, so our deeper researches on their effects on the volatile profile in the San Marzano background has laid the foundations for a reassessment of their sensory properties. Furthermore, these genotypes offered new ways to improve and diversify San Marzano fruit quality by modifying its flavour, and therefore its nutraceutical properties, finally paving the way for new breeding plans aimed at re-evaluating local varieties by the use of new innovative strategies, such as CRISPR-Cas9. Without overflying the countless advantages of using the CRISPR-Cas9 approach to efficiently obtain the desired traits, skipping the long timescales provided by the classical backcross scheme, the final aim of this work was to reveal the real effects of two mutations that were reproduced in the San Marzano variety thanks to the power of this system. This analysis lauches the possibility to easily alter the biochemical properties of traditional tomato varieties, without loosing their positive features and to go on searching for interesting new variants with a specific final destination on the market.

Bibliography

Bortesi, Luisa, et al. "Patterns of CRISPR/Cas9 activity in plants, animals and microbes." *Plant biotechnology journal* **2016**, 14.12, 2203-2216.

Brinkman, E. K., Chen, T., Amendola, M., & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic acids research*, **2014**, 42(22), e168-e168.

Brooks, C., Nekrasov, V., Lippman, Z. B., & Van Eck, J. Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. *Plant physiology*, **2014**, 166(3), 1292-1297.

Cocaliadis, M. F., Fernández-Muñoz, R., Pons, C., Orzaez, D., & Granell, A. Increasing tomato fruit quality by enhancing fruit chloroplast function. A double-edged sword? *Journal of experimental botany*, **2014**, 65(16), 4589-4598.

Cooperstone, J. L., Ralston, R. A., Riedl, K. M., Haufe, T. C., Schweiggert, R. M., King, S. A., ... & Schwartz, S. J. Enhanced bioavailability of lycopene when consumed as cis-isomers from tangerine compared to red tomato juice, a randomized, cross-over clinical trial. *Molecular nutrition & food research*, **2015**, *59*(4), 658-669.

Dono, G., Picarella, M. E., Pons, C., Santangelo, E., Monforte, A., Granell, A., Mazzucato, A. Characterization of a repertoire of tomato fruit genetic variants in the San marzano genetic background. *Scientia Horticulturae*, **2020**, 261, 108927.

Gerszberg, Aneta, and Izabela Grzegorczyk-Karolak. Influence of Selected Antibiotics on the Tomato Regeneration in In Vitro Cultures. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, **2019**, 47.3, 558-564.

Hörtensteiner, S. Stay-green regulates chlorophyll and chlorophyll-binding protein degradation during senescence, *Trends in Plant Science*, **2009**, 14(3), pp. 155–162. doi: 10.1016/j.tplants.2009.01.002.

Isaacson, T., Ronen, G., Zamir, D., & Hirschberg, J. Cloning of tangerine from tomato reveals a carotenoid isomerase essential for the production of β-carotene and xanthophylls in plants. *The Plant Cell*, **2002**, 14(2), 333-342.

Kerr, E.A. Green flesh, gf. *Rpt. Tomato Genet. Coop.* 1956, 6: 17.

Moghaieb, Reda EA, Hirofumi Saneoka, and Kounosuke Fujita. "Plant regeneration from hypocotyl and cotyledon explant of tomato (*Lycopersicon esculentum Mill*.). *Soil Science and Plant Nutrition*, **1999**, 45.3, 639-646.

Pan, C., Ye, L., Qin, L., Liu, X., He, Y., Wang, J., Lu, G. CRISPR/Cas9-mediated efficient and heritable targeted mutagenesis in tomato plants in the first and later generations. *Scientific reports*, **2016**, *6*, 24765.

Qiu, D., Diretto, G., Tavarza, R., Giuliano, G. Improved protocol for Agrobacterium mediated transformation of tomato and production of transgenic plants containing carotenoid biosynthetic gene CsZCD. *Scientia horticulturae*, **2007**, 112(2), 172-175.

Sharma, P., Srivastava, D. K. In vitro plant regeneration from cotyledon and hypocotyls tissues of tomato (Solanum lycopersicum L. cv. Solan Vajr). *Vegetos*, **2014**, 27(3), 151-160.

Simkin, A.J., Schwartz, S.H., Auldridge, M., Taylor, M.G., Klee, H.J. The tomato carotenoid cleavage dioxygenase 1 genes contribute to the formation of the flavor volatiles beta-ionone, pseudoionone, and geranylacetone. *The Plant Journal*, **2004**, 40, 882–892.

Van den Bulk, R. W., et al. Somaclonal variation in tomato: effect of explant source and a comparison with chemical mutagenesis. *Theoretical and applied genetics*, **1990**, 80.6, 817-825.

Vogel, J.T., Tan, B.C., McCarty, D.R., Klee, H.J. The carotenoid cleavage dioxygenase 1 enzyme has broad substrate specificity, cleaving multiple carotenoids at two different bond positions. *Journal of Biological Chemistry*, **2008**, 283, 11364–11373.

Funding

This work was supported by:

- the Latium Region, FILAS, project “MIGLIORA”
- the Italian Ministry of Agriculture (MiPAAF), AGROENER project (D.D. n. 26329, 1 april 2016)
- the Italian Ministry for Education, University and Research (MIUR), project SAFE-MED (initiative “Department of excellence” - Law 232/216)
- the European Commission through-H2020 SFS-7a-2014 TRADITOM (634561)
- the Eurocaroten (CA15136) (COST- European Cooperation in Scienze & Technology) supported by the EU Framework Programme Horizon 2020
- the Consorzio Interuniversitario per le Biotecnologie (CIB)
- the Spanish Ministry project BIO2016-78601-R

Acknowledgements

Non posso dichiarare concluso questo percorso di dottorato senza aver prima ringraziato tutte le persone che mi hanno permesso di raggiungere questo soddisfacente traguardo.

Prima di tutto il mio supervisor, il Prof. Andrea Mazzucato, che mi ha supportato in questi tre anni riponendo fiducia in me, aiutandomi a superare gli ostacoli incontrati e a raggiungere obiettivi anche inaspettati con calma, dedizione e razionalità, senza i quali non si può ambire a diventare un buon ricercatore.

Il Laboratorio delle biotecnologie delle colture ortofrutticole è stato il mio punto di riferimento scientifico e umano; in particolare voglio ringraziare Maurizio, per la pazienza e la gentilezza con la quale mi ha trasmesso il suo sapere e Alessandro, collega insostituibile, ora anche un caro amico. Grazie a loro l'ambiente di lavoro è sempre stato rilassato, sostenibile, basato sulla collaborazione e sull'aiuto reciproci e mai sulla competizione. Ho apprezzato particolarmente questo aspetto positivo e fondamentale della mia formazione, e me lo porterò dietro in tutte le mie future esperienze lavorative.

Voglio ringraziare tutti i ragazzi che ho conosciuto nel laboratorio, in particolare Chiara Di Iorio, Fabrizio De Angelis, Mauro Parenti, Eleonora Fabene; ognuno di loro ha contribuito a lasciare un ricordo positivo di questi tre anni.

Un grazie va a Gianfranco Diretto e al suo gruppo in Enea, che mi hanno coinvolto con grande professionalità e simpatia nelle loro attività, permettendomi di migliorare e ampliare la mia formazione.

Non posso dimenticare di ringraziare il Professor Antonio Granell, dell'IBMCP dell'Universitat Politecnica di Valencia, che ho conosciuto durante il mio Erasmus a Valencia e che mi ha dato la possibilità di continuare a svolgere dei periodi di formazione sotto la sua tutela. Nel suo laboratorio ho conosciuto Josè Luis Rambla, ricercatore brillante, che mi ha trasmesso le sue conoscenze ma soprattutto il suo amore per la conoscenza. Ringrazio anche tutti i ragazzi e i tecnici di questo laboratorio che mi hanno accolta e fatta sentire a mio agio anche lontano da casa.

Ringrazio i miei genitori per il loro sostegno continuo e forte, senza i quali non sarei arrivata fin qui.

Grazie a Domenico, che non è mai venuto meno, accompagnandomi anche attraverso i momenti di maggiore stress con pazienza e sensibilità.

Grazie a mia nonna, un mix di altruismo e comicità, unica.

Grazie al resto della mia famiglia e ai miei amici, a quelli vicini e a quelli lontani, che mi hanno sempre dimostrato affetto, sostegno e teso una mano nei momenti di difficoltà e indecisione.

Grazie alla mia maestra di danza Roberta e a tutta la scuola di danza, un porto sicuro, un luogo che mi ha sempre restituito la calma e la tranquillità, fondamentali per portare a termine questo percorso.

Questo lavoro è dedicato a tutti voi,
senza i quali non sarebbe stato possibile
e a chi lo troverà utile
per futuri approfondimenti scientifici.