University of Naples Federico II

Department of Agricultural Sciences

Division of Plant Genetics

Ph.D. Agricultural and Agri-food Sciences

Exploiting genetic and genomic resources to increase the ascorbic acid content in the tomato fruit

Antonietta Aliberti

Tutor:

Prof.ssa Amalia Barone

INDEX

Abstract

The wild tomato species such as *Solanum pennellii* are an important source of genes that were lost during tomato selection and cultivation. The S. *pennellii* Introgression lines (ILs), which carry defined homozygous segments of the wild genome in the cultivated genetic background of cv.M82, are an important genetic resource to map quantitative traits loci (QTLs), such as those controlling plant yield and fruit quality, and to exploit the genetic diversity present in the wild species. Two introgression lines (IL12-4 and IL7-3) harbouring QTLs for ascorbic acid content were previously identified in the laboratory of Structural and Functional Genomics at the Department of Agricultural Sciences of University of Naples Federico II. These two ILs showed increased content of antioxidant compounds in the fruit compared to the cultivated S. *lycopersicum* cv. M82. Afterwards, sub-lines with reduced sizes of the introgressed region were obtained from IL7-3 and IL12-4.

The first aim of the present thesis was to perform the phenotypic selection and characterization of the *S. pennellii* sub-lines in different environmental conditions. In the fruit of the different sub-lines, the level of soluble solids content in terms of °Brix, firmness and ascorbic acid was highly variable in three different environmental conditions. The sub-lines R182 and B27, deriving from IL7-3 and IL12-4, respectively, were selected for their better performances in terms of fruit quality since they exhibited a significantly higher firmness, °Brix and ascorbic acid content compared to M82. Moreover, the two sub-lines also showed a production comparable to that of the control line M82 in all the environment conditions tested.

The second aim of the present thesis was to identify candidate genes involved in determing the high level of ascorbic acid in the fruit and mapping in the introgressed regions of the sub-lines R182 and B27. Since today the only reference genome used to investigate gene positions and functions in tomato still remains that of the first completely sequenced genome, i.e. that of S. *lycopersicum* cv. Heinz, the first step to reach this second aim was to reassembly the S. *pennellii* IL7-3 and IL12-4 genomes. In order to reassemble the genome of M82 and of the ILs, several resources were used. These consisted of Next Generation Sequence resources available today, such as the sequence data of the S. *pennellii* and different RNA-seq data related to the IL populations.

By using an integrated bioinformatic approach a new reference genome and annotation for IL12-4 and the IL7-3 was built. Moreover, to confirm the reliability of the IL7-3 assembly and define the size of S. *pennellii* introgression region in the sub-line R182, a group of species-specific molecular markers were designed based on polymorphisms found comparing the genomes of the cultivated and the wild species. Finally, to identify candidate genes mapping in the wild regions better defined in the present thesis, a transcriptomic approach with RNA-Seq was carried out. Transcriptome analyses allowed identifying three candidate genes in the two sub-lines: the Solyc12g098480 encoding for the UDP-glucoronic-acid-4-epimerase in the wild introgressed region of B27 and the Solyc07go49310 and Solyc07g049290 corresponding to two Major facilitator superfamily (MFS) proteins in in the wild introgressed region of R182.

The third aim of the present thesis was to carry out the functional validation of candidate genes potentially involved in the control of ascorbic acid content in the red ripe fruit. The first gene selected was the UDP-glucoronic-acid-4-epimerase gene mapping in the introgressed region of the sub-line B27. Its role in increasing ascorbic acid in this sub-line fruit was investigated through its over-expression in tomato fruit by stable and transient techniques. In addition, the genotyping and phenotyping evaluation of CRISP/Cas9 knock-out mutants for a non-canonical uORF carried by two GDP-L-galactosephosphorylases (GGP1, Solyc06g073320 and GGP2, Solyc02g091510) allowed demonstrating that these two genes are involved in the regulation of the ascorbate biosynthesis in tomato.

In conclusion, the results obtained in the present thesis, allowed selecting two elite sub-lines that in the near future could be used as breeding material to improve tomato fruit for nutritional traits. In addition, the results achieved allowed increasing knowledge about genes involved in the control of ascorbic acid content in tomato fruit. Indeed, three candidate genes were identified exploiting the genomics resources available for tomato, and other two genes have been functionally validated. The transfer of these genes by conventional and innovative strategies will aid in the future the creation of new improved varieties.

Chapter 1. Introduction

The cultivated tomato (*Solanum lycopersicum*) belongs to the nightshades Solanaceae, one of the most important and large plant family in the world, being present on all continents except Antarctica. Indeed, it was supposed that the Solanaceae family consists of about 98 genera and 2,700 species (Yadav et al., 2016), and is divided in three subfamilies, Solanoideae (to which belongs the genus Solanum), Cestroideae, and Solanineae (Knapp et al., 2004) with a great diversity of habitat, morphology and ecology. The greatest species diversity has been found in South America and Central America, where most members of the Solanaceae are erect or climbing, annual or perennial herbs, shrubs are not uncommon and there are also a few trees (Yadav et al., 2016). The Solanaceae family contains several of the most important species with a high economic impact and used every day in the human nutrition, such as tomato, potato (*S. tuberosum* L.) and aubergine or eggplant (*S. melongena* L.) (Knapp and Peralta, 2016) and other common species as tobacco and petunia, used for ornamental purposes. The cultivated tomato is widely grown around the world and supplies a major agricultural industry. It is the second most important vegetable crop after potato with a global production of about 164 million tons (t) of fresh fruit harvested on a 4.7 million hectares (ha) surface (FAOSTAT, 2016, http://faostat.fao.org). Most of tomatoes are consumed fresh as raw vegetable or added to other food items, however some varieties are harvested to be processed as paste, whole peeled tomatoes, diced products, and various forms of juice, sauces, and soups. Although tomato optimal growing temperature is around 25ºC during the day and around 20ºC during the night, it is grown in almost every corner of the world, in open fields or greenhouses from the tropics to the Arctic Circle. In the 2016, the most important producing countries were China, India and United States, Turkey, Egypt, Italy and Iran (Figure 1.1), followed by minor producing countries, such as Spain, Brazil, Mexico, Russia, Uzbekistan, Nigeria, Ukraine, Portugal, Tunisia, Algeria, Morocco, Camerun and Greece (FAOSTAT, 2016).

Country	Tonnes
China	56,308,914
India	18,399,000
USA	13,038,410
Turkey	12,600,000
Egypt	7,943,285
Italy	6,437,572
Iran	6,372,633

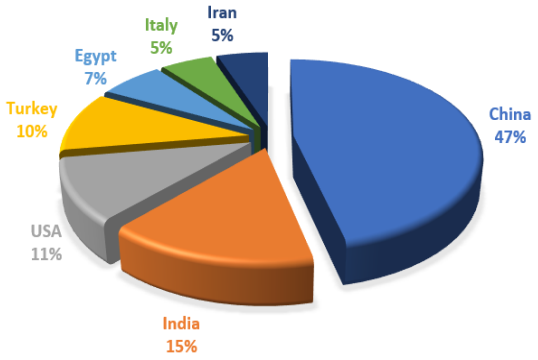


Figure 1.1- World tomato production. Major tomato producing countries in the world during the year 2016 (http://faostat.fao.org).

As for processing and fresh tomato cultivation in Italy, the major producing regions are Puglia, Emilia-Romagna and Campania (USDA Foreign Agricultural Service, 2016). Nowadays, tomato is involved as a key food in the Mediterranean diet, indeed it is an important source of substances such as vitamins, minerals, and antioxidants, already known for their beneficial effects on human health (Raiola et al., 2014).

1.1 The tomato fruit quality

Quality is a key trait in plant breeding, especially for fruit and vegetables (Diouf et al., 2018). Commercial quality mainly relies on external attractiveness (colour, form, size), firmness and shelf life; organoleptic quality depends on physical (texture or firmness) and biochemical traits (mainly the contents in sugars, acids and volatile compounds) determining the overall flavour (Bertin and Genard, 2018). In the last years, the interest in quality traits, allowed to detect quantitative trait loci (QTL), which represent a very useful tool for gene cloning, MAS breeding, and trait improvement (Li et al., 2017). Tomato, among all Solanaceae species, is considering as a model system for dissection of genetic determinants of QTLs. They were detected in tomato for many important traits such as yield (Eshed and Zamir, 1995), disease resistance (Mutschler et al., 1996), tolerance to abiotic stress (Foolad et al., 1998) and plant architecture (De Vicente and Tanksley, 1993, Paran et al., 1997).

In the last years researchers have focused their attention especially on the fruit compositional quality for human health such as antioxidant compounds. Indeed, tomato fruits are rich in phytochemicals such as carotenoids (mainly lycopene and β-carotene), phenolic compounds (mainly flavonoids, such as naringenin), vitamins C and E (Valderas-Martinez et al., 2016) and their consumption has been associated with a reduced risk of inflammatory processes, cancer, and chronic non-communicable diseases (CNCD) including cardiovascular diseases (CVD), such as coronary heart disease, hypertension, diabetes, and obesity (Canene-Adams et al., 2005). Antioxidant metabolites are a group of vitamins, carotenoids, phenolic compounds, and phenolic acid, with health-enhancing effects on human body (Raiola et al., 2014).

Carotenoids are the pigments synthesized during fruit ripening and responsible for the final red colour of the fruit. In tomato two main carotenoids were found: the lycopene that represents around 90% of the carotenoids and the β-carotene, which is approximately 10% of the total carotenoids (Frusciante et al., 2007). Lycopene has antioxidant ability since it acts as free radical scavenger from reactive oxygen species (ROS). Free radicals are highly reactive, short-lived molecules that can react with damaging essential structural proteins, enzymes, and DNA. Such damage has the potential to cause cancer, atherosclerosis, cardiovascular, and other diseases. Lycopene has the potential to reduce such undesirable molecular events because the high-energy, highly reactive free electron on DNA is transformed to a much less reactive more stable free electron after it is delocalized along the conjugated 13 double bonds of the lycopene molecule (Friedman, 2013).

β-carotene is an organic compound classified as a terpenoid and it is a precursor of vitamin A. As lycopene, it is directly involved in reducing the formation of free radicals. Some researchers demonstrated that β-carotene prevents photooxidative damage and sunburn (Raiola et al., 2014). Indeed, β-carotene is particularly used for skin treatments.

1.1.1 Ascorbic acid

Ascorbic acid (AsA), also known L-ascorbic acid or Vitamin C, is one of the most important antioxidant compounds in plants, which are the major source of this vitamin in the human diet since humans are unable to produce it. Indeed, to provide antioxidant protection, the recommended dietary allowance (RDA) for vitamin C for non-smoking is approximately 90 mg/day for adult men and 75 mg/day for adult women (Food and Nutrition Board, Institute of Medicine, 2000), but it varies between different countries (Fang et al., 2017). Tomato is one of the most important AsA source for human health especially for people feeding with the Mediterranean diet.

Usually the AsA content in tomato fruit ranges from 10 to 88 mg/100 g of fresh weight (FW) even if some commercial cultivars show a lower value (from 10 to 40 mg/100 g of FW) (Ruggieri et al., 2016). This is probably consequence of the tomato domestication/breeding process that led to select for agronomical traits negatively associated with AsA content (Locato et al., 2013). In plants, AsA serves as a major redox buffer, as a cofactor for many enzymes, and as a regulator of cell division and growth, as well as in signal transduction (Gallie, 2013).

The biosynthesis of AsA in higher plants takes place in mitochondria *via* several proposed routes (Akram et al., 2017) (Figure 1.2). The primary pathway is the D-mannose/L-galactose pathway (Wheeler et al., 1998). In this pathway, D-glucose is converted to D-glucose 6-phosphate by the enzyme hexokinase. The D-glucose 6-phosphate formed is then converted to GDP-D-mannose *via* a four steps reversible process catalysed by the enzymes phosphogluco-isomerase, mannose 6-phosphate isomerase, phosphomannose mutase and GDP-D-mannose phosphorylase, respectively (Akram et al., 2017). The second major phase involves the conversion of GDP-D-mannose into GDP-L-galactose following three reactions catalysed successively by GDP-D-mannose 3′,5′-epimerase, GDP-L-galactose phosphorylase and L-galactose 1-P phosphatase. The GDP-L-galactose is subsequently converted to L-galactose by the action of enzyme L-galactose dehydrogenase and finally into L-galactono-1, 4-lactone (final precursor of AsA). Lastly, ascorbic acid is formed from L-galactono-1,4-lactone in an enzymatic reaction catalysed by L-galactono-1,4-lactone dehydrogenase (Akram et al., 2017).

The second pathway involves the cell wall pectins degradation, which results in the formation of methyl-galacturonate that is converted into L-galactonate *via* two reactions catalysed by methyl esterase and D-galacturonate reductase. Later on, the enzyme aldono lactonase catalyses the conversion of L-galactonate into L-galactono-1, 4-lactone and is finally used in ascorbate synthesis (Smirnoff et al., 2001).

The third pathway involves the conversion of GDP-D-mannose to GDP-L-gulose and subsequent formation of L-gulono-1, 4-lactone via L-gulose (Wolucka and Van Montagu, 2003). This is similar to the primary pathway which starts from glucose however this pathway branches off from GDP-D-mannose. From here, GDP-D-mannose is converted to L-gulose in three reactions catalysed by GDP-D-mannose-3′, 5′-epimerase, GDP-L-gulose-1-P-phosphatase and L-gulose-1-P-phosphatase, respectively. At this point, the enzyme L-gulono-1, 4-lactone dehydrogenase catalyses the conversion of L-gulose into L-gulono-1, 4-lactone which is then finally converted to AsA (Smirnoff et al., 2001).

In a fourth pathway, the synthesis of ascorbate from myo-inositol is reported. Briefly, myo-inositol is converted to L-gulono-1, 4-lactone *via* three reactions catalysed by a myo-inositol oxygenase, a glucuronate reductase and an aldono lactonase (Valpuesta and Botella, 2004). The gulono-1, 4-lactone is finally used in ascorbate synthesis (Smirnoff et al., 2001). For convenience, the four major pathways and important precursor molecules have been elucidated in the form of a schematic diagram (Akram et al., 2017).

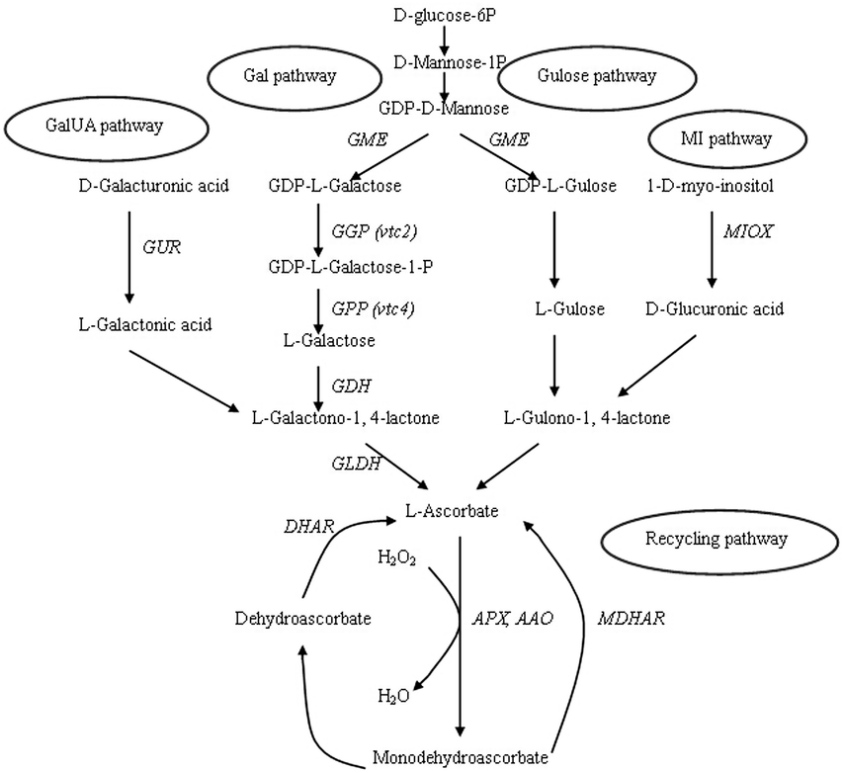


Figure 1.2- AsA synthetic and recycling pathways in plants. The four pathways included GalUA (D-galacturonic acid) pathway, Gal (L-galactose) pathway, Gulose(L-gulose) pathway and MI (Myo-inositol) pathway (Lu et al., 2016).

1.2 Tomato introgression lines

In addition to the cultivated species *S. lycopersicum* there are some related wild species, including *S. pimpinellifolium* (Jusl.) Mill. (current tomato), *S. cheesmanii* Riley, *S. chmielewskii* Rick, Kes., Fob. and Holle, *S. chilense* Dun., *S. neorickii* Rick, Kes., Fob. and Holle, *S. peruvianum* (L.) Mill., *S. habrochaites* Humb. and Bonpl., and *S. pennellii* (Corr.) D'Arcy. Among the wild species, the researchers interest is focused on S. *pennellii*, species of the South America that is considered an important donor of germplasm for the cultivated tomato S. *lycopersicum* (Bolger et al., 2014) due to its greater genetic variability

respect to S. *lycopersicum*, which lost lots of important traits during domestication and evolution. The cross among *S. pennellii* × *S. lycopersicum* cv M82 allowed to identify 76 ILs (Figure 1.3), which present different and defined wild chromosomal segments in the genetic background of the cultivated variety M82 (Eshed & Zamir, 1995, Zamir, 2001).

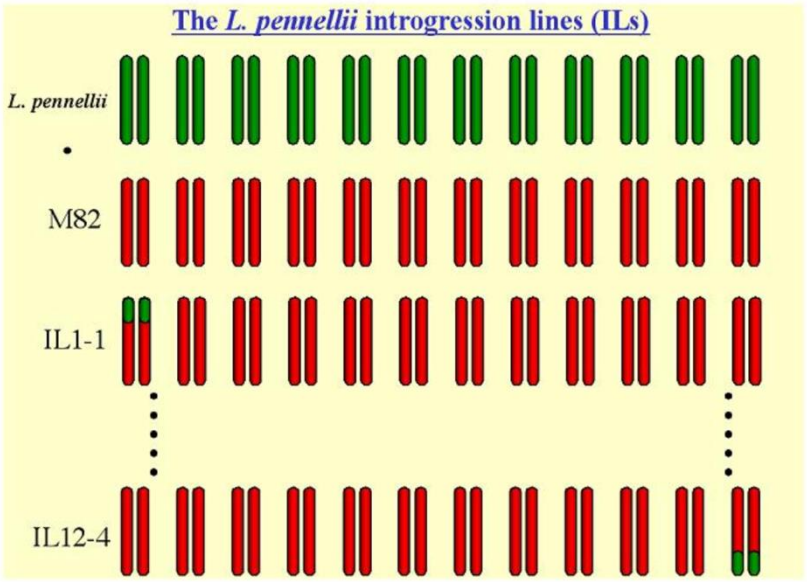


Figure 1.3- Schematic representation of *S. pennellii* ILs (http://zamir.sgn.cornell.edu).

Overall, these lines offer a complete coverage of the parental donor genome divided into 107 bins (with an average length of 12 cM each) (Smit et al., 2012) (Figure 1.4), which define the position of the donor parental introgression in the genome of the recurrent parent. These lines are an excellent resource of genetic variation so far used worldwide by several researchers for the identification of more than 2,700 agronomically useful QTLs, such as those controlling plant biomass yield, drought tolerance, morphology, chemical composition and other qualitative characteristics (Kamenetzky et al., 2010).

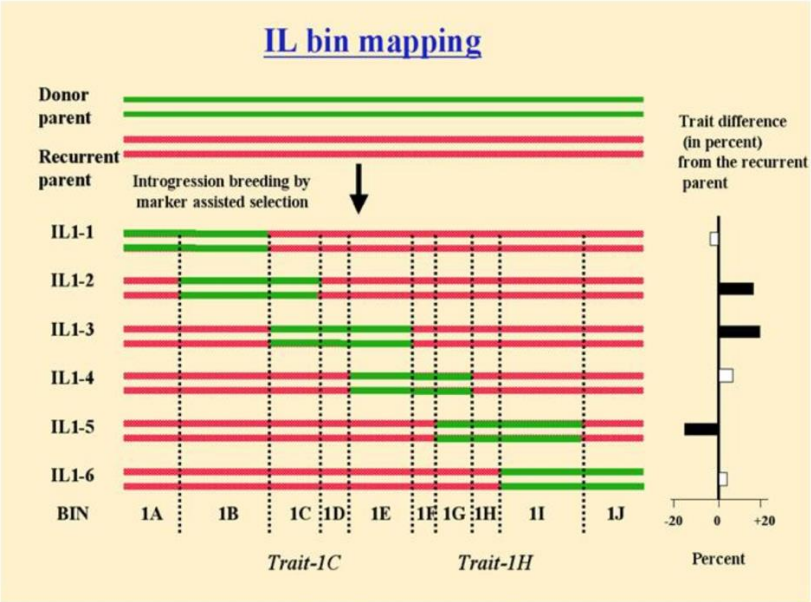


Figure 1.4- ILs bin mapping provides a rapid method for assigning a map position to DNA sequences (http://tgc.ifas.ufl.edu/vol49/VOL49/html/Vol4_26.htm), (http://zamir.sgn.cornell.edu).

In the last years, at the Department of Agricultural Sciences at the University of Naples Federico II, the attention was focused on two ILs, selected for the high content of AsA in the mature red fruit (Di Matteo et al., 2010, Sacco et al., 2013, Rigano et al., 2014). Indeed, the IL7-3 (Figure 1.5) carrying a S. *pennellii* introgression region of 35 cM, showed a QTL with possible candidate genes for the increase of antioxidant compounds in tomato fruit, as well as the IL12-4 (Figure 1.6) with an introgression fragment of 46 cM. Moreover, during breeding programs, these two ILs were crossed with the cultivated genotype S. *lycopersicum* cv. M82 in order to reduce the introgression size: different sub-lines were obtained and have been previously investigated to better identify possible candidate genes for the increment of AsA content (Figure 1.7, Figure 1.8) (Ruggieri et al., 2015, Calafiore et al., 2016). In the future these genes, using the new genomic techniques as RNA-seq and genome editing, could be identified and validated, to be then transferred in cultivated varieties to improve the content of quality traits as AsA content in fruits.

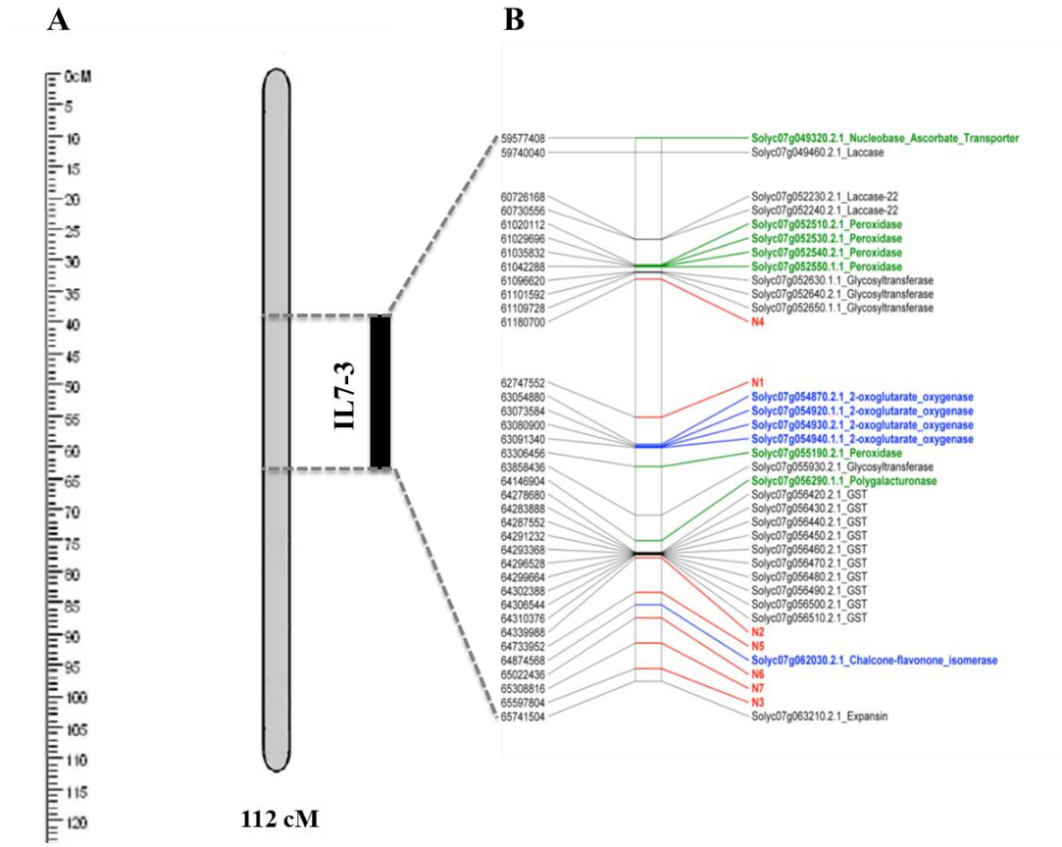


Figure 1.5- Chromosome 7 of S. *pennellii* IL7-3. A) Position of the introgression line IL7-3 respect to the chromosome 7. B) Mapping of possible candidate genes and of polymorphic molecular markers (N) in region 3 of chromosome 7. On the left, the position in bp of each gene/marker on the chromosome is also reported.

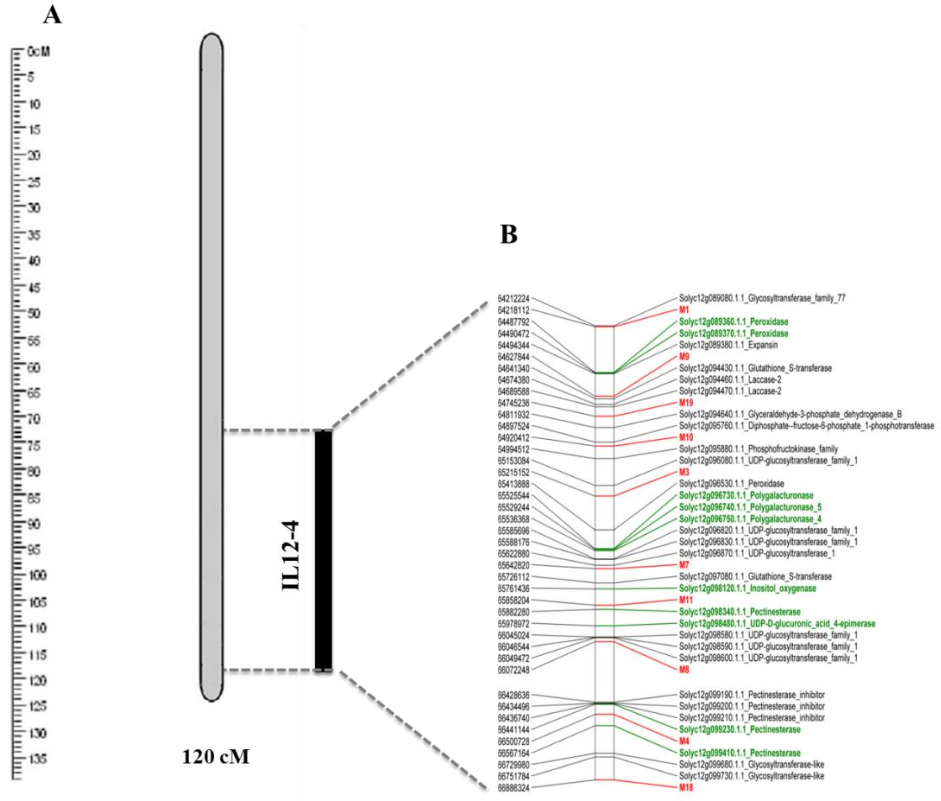


Figure 1.6- Chromosome 12 of S. *pennellii* IL12-4 A) Position of the introgression line IL12-4 respect to the chromosome 12. B) Mapping of possible candidate genes and of polymorphic molecular markers (M) in region 4 of chromosome 12. On the left, the position in bp of each gene/marker on the chromosome is also reported (Ruggieri et al., 2015).

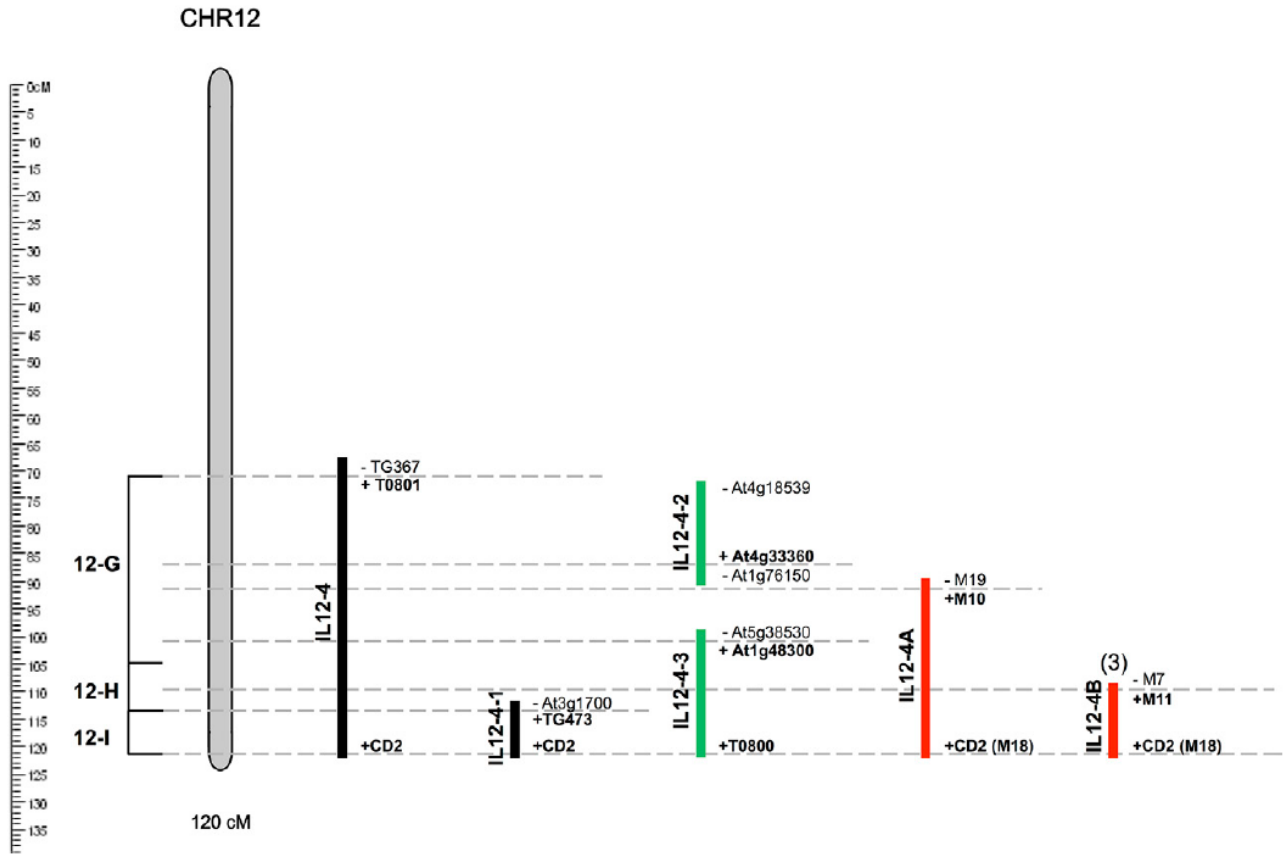


Figure 1.7- Position of the IL12-4 sub-lines selected by Ruggieri et al. (2015).

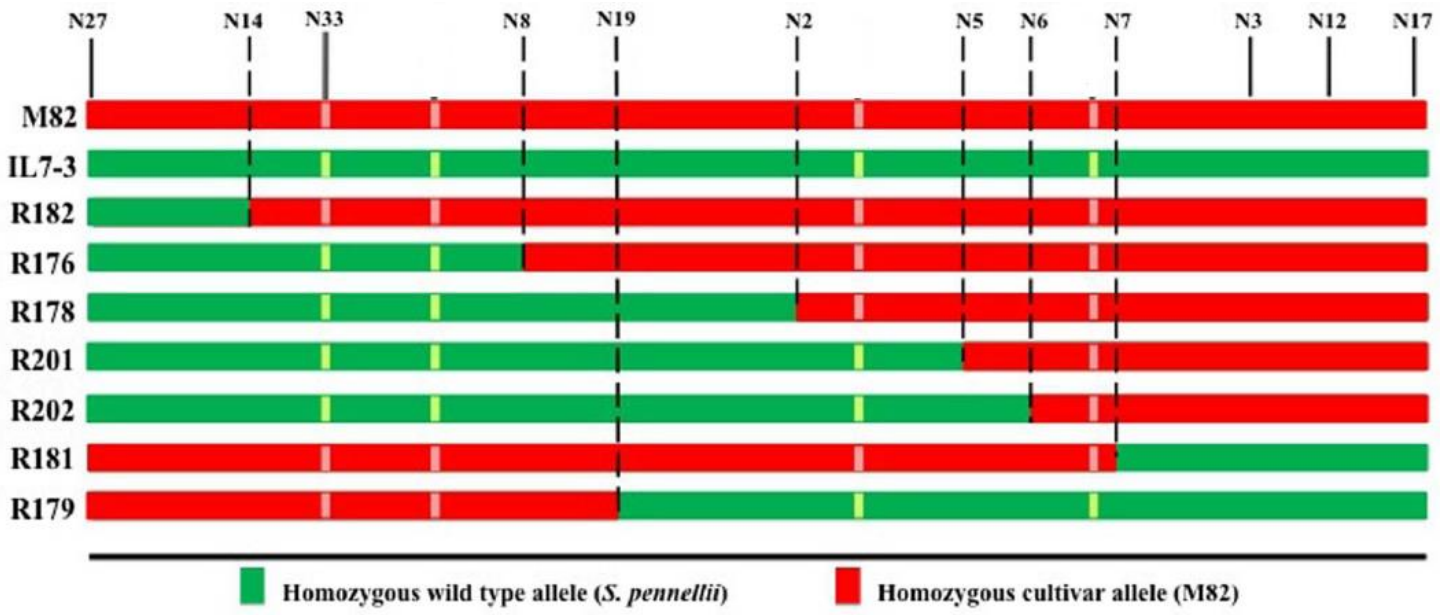


Figure 1.8- Size and genomic identity of the seven sub-lines selected from the region 7-3 (Calafiore et al., 2016).

1.2 RNA sequencing technique

The term Next-generation sequencing (NGS) is used to describe several different modern sequencing technologies, including:

- Illumina sequencing;
- Roche 454 sequencing;
- Ion torrent: Proton / PGM sequencing;
- SOLiD sequencing.

These technologies allow to sequence DNA and RNA much more quickly and cheaply than the previously sequencing technique, revolutionizing the study of genomics and molecular biology (European Informatics institute, https://www.ebi.ac.uk). Today the Illumina technology is most frequently used for RNA sequencing.

RNA sequencing (RNA-seq) now is the most common method to analyse gene expression and to uncover novel RNA species (Hrdlickova et al., 2017). In general, a population of RNA is converted to a library of cDNA fragments with adaptors attached to one or both ends. Each molecule, with or without amplification, is then sequenced in a high-throughput manner to obtain short sequences from one end (single-end sequencing) or both ends (pair-end sequencing) (Wang et al., 2009). After sequencing, reads are aligned to a reference genome or reference transcripts, whereas if the genomic sequence is not available a *de-novo* assembly strategy is performed (Wang et al., 2009). Generally, the RNA-seq is used to evaluate gene expression between two different conditions and to find different expressed genes using bioinformatics tools, as EdgeR packages from Bioconductor (Robinson et al., 2010), that count the number of read and compare them among samples.

This technique has allowed to identify candidate genes controlling different trait, in tomato as in many other species (Fang et al., 2016, Khalil-Ur-Rehman et al., 2017) In particular, RNA-seq analysis can be used in tomato to identify deferentially expressed gene involved in AsA pathways (Ye et al., 2015) or explore the interaction of mycorrhization and fruit development (Zouari et al., 2014).

Therefore, the transcriptomic analysis is key step to identify possible candidate genes, which function will be validated using different technique as stable or transient transformation, VIGS or the innovative genome editing technique.

1.3 Crisp/Cas9 technique

Since the last years, the functional validation of candidate genes was performed using different technique such as agroinfiltration mediated transient expression to overexpress gene involved in the content of AsA (Ye et al., 2015) or stable transformation (Ruf et al., 2001). However, actually, the most important validation technique is the genome editing with the CRISP/Cas9 approach.

The clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) technology is as a new gene-editing technique used to induce deletions or introduce specific changes at precise genomic loci (D'Ambrosio et al., 2018, Li et al., 2018, Tomlison et al., 2018, Ueta et al., 2017, Jinek et al., 2012). This CRISPR/Cas system is based on a Cas9 (Figure 1.9) bacterial endonuclease that forms a complex with a CRISPR RNA (crRNA) and a trans- activating CRISPR RNA (tracrRNA).

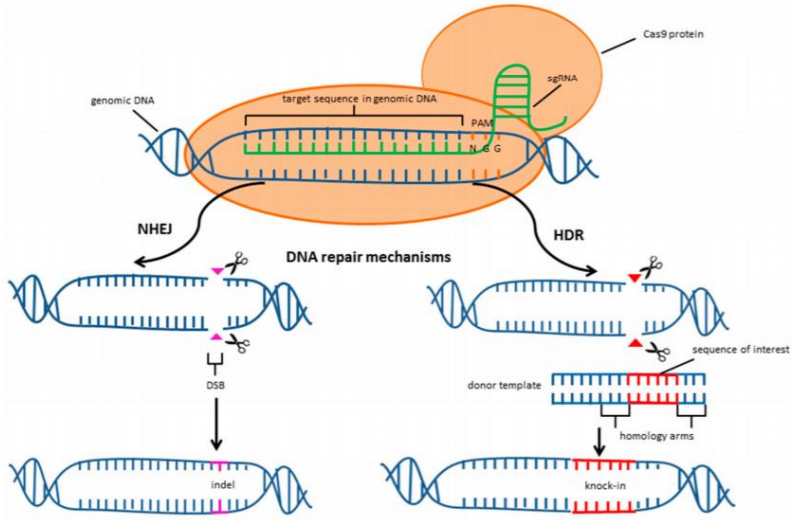


Figure 1.9- CRISPR/Cas9 mediated cleavage of genomic DNA and two major repair pathways (Kruminis-Kaszkiel et al, 2018).

These two RNA structures can be combined into the synthetic guide RNA complex (sgRNA) (Bassett et al., 2013). The sgRNA contains a 20-bp complementary sequence of the target sequence and brings target loci in the genome upstream of a specific protospacer adjacent motif (PAM), such as the canonical "NGG" for the *Streptococcus pyogenes* Cas9 endonuclease (Choo et al., 2017). These endonucleases play the function of a genomic scissors, create a double-strand breaks (DSBs), assuring precise genome editing. The disruption in the DNA sequence triggers various repair mechanisms, such as non-homologous end joining (NHEJ) and homology-directed repair (HDR), leading to the induction of specific knock-outs or knock-ins (Kruminis-Kaszkiel et al, 2018).

1.4 Aims of the thesis

The S. *pennellii* ILs IL7-3 and IL12-4, and their sub-lines already available, constitute a powerful genetic material to reach new results in the selection of superior genotypes and new insights in the identification of candidate genes for the increase of AsA in the tomato ripe fruit. The main goal of the present thesis is to increase the knowledge related to the accumulation of this metabolite in the fruit by exploiting various genetic and genomic resources available for the tomato species.

In particular, specific aims of the present thesis are:

1. to perform the phenotypic selection and characterization of S. *pennellii* IL7-3 and IL12-4 sub-lines, which carry positive QTLs for AsA accumulation in the tomato fruit;

2. to reassembly the S. *pennellii* IL7-3 and IL12-4 genomes and to identify candidate genes controlling high AsA content in the selected sub-lines and mapping in their wild genomic regions;

3. to carry out the functional validation of candidate genes through their over-expression in tomato fruit by stable and transient techniques or by the genotypic and phenotypic characterization of knock-out mutants obtained by the CRISP/Cas9 genome editing technique.

Chapter 2. Selection and characterization of *S. pennellii* sub-lines

2.1 Materials and methods

2.1.1 Plant materials

Plant material used in Italy consisted of two *S. pennellii* in *S. lycopersicum* introgression lines (IL12-4 and IL7-3, accessions LA4102 and LA4066), the cultivated genotype M82 (accession LA3475), the IL12-4 subline B27 and seven IL7-3 sublines. The IL12-4 sub-line B27 was previously selected in our laboratory (Ruggieri et al., 2015). Two IL7-3 sub-lines (coded R201 and R202) were previously obtained in our laboratory (Calafiore et al., 2016), whereas five sub-lines (coded R176-R182) were kindly provided by Dr. Dani Zamir (Hebrew University, Israel). All genotypes were grown in open-field for two years (2016, 2017). Data collected in the year 2016 are related to one field (Acerra, Campania Region), whereas in the 2017 data were collected from three open fields in the South Italy: Acerra, Battipaglia and Giugliano following the traditional area farming practice. In each field a randomized complete block design was applied with three replicates *per* genotype and 10 plants *per* replicate. Fruits were collected at three development stages: Mature Green (MG), Breaker (BR) and Mature Red (MR). Seeds and columella were subsequently removed, and fruits were ground in liquid nitrogen and stored at −80°C until analyses. Leaves were also collected for each genotype, ground in liquid nitrogen by mortar and pestle to a fine powder and stored at −80°C until analyses.

2.1.2 Morpho-agronomic analyses

Three plants *per* genotype were used each week for measuring, from the second to the fifth inflorescence, traits related to fruit set (FS) and number of flowers/inflorescence (NFL), marking for each inflorescence the number of closed and cut flowers. Data recorded as number of fruit (NFR), fruit weight (FW) and yield (YP) were collected from all plants *per* genotype according to each genotype maturity.

The software Tomato Analyzer 3.0 (http://www.oardc.ohiostate.edu/vanderknaap/tomato_analyzer.php) was used for the fruit morphological characterization using six fruits per genotype to measure the fruit perimeter (FP), fruit area (FA), pericarp area (PA), pericarp thickness (PT), fruit shape index I and II (FSI I and II), distal and proximal angle (DA and PA, respectively).

2.1.3 Qualitative analyses

For each genotype and biological replicate 15 fruits were collected at MR stage to evaluate soluble solids content, firmness and colour. The soluble solids content was measured as °Brix in the homogenized juice from ripe fruit by a refractometer (Hanna), the firmness (F) of fruit cuticle was measured on one side of ripe fruit by a penetrometer with an 8 mm shore (PCE-PTR200 penetrometer). The colour of ripe fruit was assessed as percentage of reflectance (L) and absorbance index (a/b), where a is the absorbance a 540 nm and b at 675 nm, using a Konica Minolta CR-400a, and two measures per ripe fruit (six fruits per genotype).

2.1.4 Ascorbic acid determination

A colorimetric method was used for the ascorbic acid determination (Stevens et al., 2006) with modifications reported by Rigano et al. (2014). Briefly, 300 µl of ice-cold 6% TCA was used for 500 mg of frozen powder, afterwards the mixture was vortexed, for 15 min. incubated on ice and centrifuged at 14000 rpm for 20 min at 4°C. In an Eppendorf tube were placed 20 µl of supernatant with 20 µl of 0.4 M phosphate buffer (pH 7.4) and 10 µl of double distilled (dd) water. Then, 80 µl of colour reagent solution were prepared by mixing solution A [31% H3PO4, 4.6% (w/v) TCA and 0.6% (w/v) FeCl3] with solution B [4% 2,2′-dipyridil (w/v)]. The mixture was incubated at 37°C for 40 min and measured at 525 nm by a NanoPhotometerTM (Implen). Three separated biological replicates for each sample and three technical assays for each biological repetition were measured. The concentration was expressed in nmol of AsA according to the standard curve, designed over a range of 0–70 nmol; the values were then converted into mg/100 g of fresh weight (FW).

2.1.5 Carotenoids determination

The method reported by Zouari et al. (2014) was used for the carotenoids extraction. A solution of acetone/hexane (40/60, v/v) was added to one gram of powder and incubated for 15 min at room temperature. The mixture was centrifuged at 4000 rpm for 10 min and the absorbance of the supernatant was measured at 663, 645, 505, and 453 nm. Total carotenoids were determined by the equation reported by Wellburn (1994). Results were expressed as mg per 100 g FW. All biological replicates *per* sample were analysed in triplicate.

2.1.6 Statistical analysis

Data analysis was carried out using SPSS Software version 23. The t-Student test was calculated for qualitative and quantitative traits to verify if genotypes were statistically different from the parental M82. A univariate analysis was performed to detect the genotype per environment interaction.

2.2 Results

2.2.1 Phenotypic characterization during the year 2016

The seven sub-lines derived from IL7-3 and the sub-line B27 derived from IL12-4 were analysed at Acerra in the year 2016 for productivity, measuring five yield-related traits compared to the control genotype M82 and to their parental lines IL7-3 and IL12-4. Moreover, eight different traits related to fruit quality were evaluated.

As for the productivity (Table 2.1), no genotypes showed values of NFL significantly different from M82. Indeed, NFL varied from 5.08 (M82) to 5.94 (R202) with a mean value of 5.48. As for FS, the genotypes R179 and R202 showed a significant reduction of 25% (41.0) respect to M82 (56.0), as well as IL7-3 showed a significant reduction of 37% respect to M82 (35.6 *vs* 56.0). The FW was highly variable among the sub-lines, ranging from 35-42 g (R176, R178, R179, R201 and R202) to 54-55 g (R181 and R182), the latter value being not different from the parental genotypes M82 (51.7 g) and IL12-4 (54.8 g) whereas the other sub-lines showed a significant fruit weight reduction. As for NFR, this value ranged from the minimum of 24.7 recorded for R176 to the maximum of 32.4 recorded for R182, and no statistical differences were detected when comparing the sub-lines with M82, whereas IL7-3 exhibited a significantly lower value compared with the cultivated M82 (15.3 in IL7-3 *vs* 30.1 in M82). Finally, as for the yield *per* plant, in all the sub-lines a yield value similar to M82 was observed, with a mean of 1.26 kg/plant and only the genotype IL7-3 showed a value (0.60 kg/plant) significantly lower than M82 (1.58 kg/plant).

Table 2.1- Productivity of the sub-lines and their parental genotypes at Acerra in the year 2016.

Genotype	Flowers/inflorescence no.	Fruit set %	Fruit weight g	Fruit/plant no.	Yield/plant kg
M82	5.08±0.58	55.98±5.00	51.70±2.89	30.07±7.95	1.58±0.36
IL12-4	5.64±0.42	53.08±0.22	54.83±0.29	31.33±12.51	1.72±0.68
IL7-3	5.25±0.65	35.57±5.87**	39.50±1.80*	15.30±0.44*	0.60±0.03*
B27	5.64±0.46	52.56±0.41	49.97±0.06	27.63±10.80	1.38±0.54
R176	5.44±0.51	40.93±7.05	42.00±4.00*	24.7±10.07	1.05±0.46
R178	5.61±0.59	53.38±1.49	39.00±6.56*	30.9±9.29	1.24±0.56
R179	5.44±0.27	41.74±3.44*	37.77±2.31*	27.7±5.40	1.05±0.17
R181	5.49±0.69	46.00±2.19	54.00±8.89	26.9±7.73	1.43±0.34
R182	5.19±1.47	53.86±13.28	55.00±1.00	32.4±8.76	1.78±0.28
R201	5.52±0.68	51.54±10.02	36.00±1.73**	29.1±15.10	1.07±0.61
R202	5.94±0.63	40.92±3.25*	35.33±3.51*	29.0±11.57	1.01±0.37

Data related to parameters linked to the production, such as the number of flowers per inflorescence and the fruit set, are reported together with those strictly related to the yield/plant, such as fruit weight and the number of fruit per plant. The significance of differences of each genotype *vs* M82 was evaluated by the t-Student's test (*:P<0.05; **:P<0.01).

The analysis of fruit quality traits carried out in Acerra in the year 2016 (Table 2.2) showed for some traits levels significantly different between IL7-3 sub-lines and M82. As for °Brix, a significantly higher level was observed for the sub-lines R178 (5.40), R179 (4.73), R181 (4.97), R182 (4.97), R201 (5.30) and R202 (5.47) respect to M82 (3.87), as well as IL7-3 and IL12-4 showed high significant values. As for fruit firmness R178, R179, R181 and R202 showed a lower value of 5.31 kg/cm^2, 4.81 kg/cm^2, 5.48 kg/cm^2, 5.77 kg/cm^2, respectively, compared with M82 (6.48 kg/cm^2), and only the sub-line R182 showed a very high level of firmness (8.40 kg/cm^2). The detection of fruit colour with Minolta in terms of a/b did not show significant differences among all lines analysed. Differences in the fruit colour were detected when measuring the chroma value, with a reduction of 12% between M82 and the genotypes R178 and R202. Metabolic analysis carried out on fruits (Table 2.3) evidenced that only R182 accumulated the significant highest levels of AsA compared to M82 in the green (18.45 mg/100 gr FW *vs* 11.00 mg/100 gr FW) and mature red (44.81 mg/100 gr FW *vs* 31.13 mg/100 gr FW) stages. In mature red stage significantly high AsA levels were also observed in IL7-3 (39.88 mg/100 gr FW), R202 (40.8 mg/100 gr FW) and B27 (35.35 mg/100gr FW) and only R179 (25.96 mg/100 gr FW) exhibited a significant lower value respect to M82. As for lycopene only R176, R178, IL12-4 and B27 showed significantly lower values respect to M82, whereas for total carotenoids only R176, R178 and B27 exhibited lower values compared with M82. As for β-carotene, only IL12-4 and B27 showed a lower value respect to the cultivated genotype.

Table 2.2- Fruit quality traits of the sub-lines and their parental genotypes at Acerra in the year 2016.

Genotype	Brix (°)	Firmness	Color a/b	Color Chroma
M82	3.87±0.25	6.48±0.15	1.22±0.09	41.99±1.54
IL12-4	4.78±0.03*	6.88±0.11	1.11±0.02	42.16±3.07
IL7-3	5.77±0.55*	6.42±0.66	1.12±0.09	37.10±3.42
B27	4.53±0.06	6.53±0.33	1.15±0.02	41.92±0.15
R176	4.37±0.45	6.31±0.13	1.04±0.04	42.69±0.54
R178	5.40±0.72*	5.31±0.35*	1.09±0.09	37.65±1.12*
R179	4.73±0.42 *	4.81±0.22***	1.18±0.05	41.04±3.29
R181	4.97±1.07	5.48±0.21*	1.23±0.03	43.50±0.90
R182	4.97±0.31**	8.39±0.62*	1.29±0.08	40.87±0.95
R201	5.30±0.26*	6,04±0.68	1.22±0.12	37.75±2.50
R202	5.47±0.84*	5.77±0.14*	1.20±0.11	37.91±0.37**

The significance of differences of each genotype *vs* M82 was evaluated by the t-Student's test (*: P<0.05; ** :P<0.01; ***:P<0.001).

Table 2.3- Metabolic analysis of the sub-lines and their parental genotypes at Acerra in the year 2016 (all data are reported as mg/100 g FW).

Genotype	Lycopene	β-carotene	Total carotenoids	AsA MG	AsA BR	AsA MR
M82	6.76±0.93	2.38±0.79	10.30±0.92	11.00±2.28	36.46±5.87	31.13±1.56
IL12-4	3.36±0.25*	0.89±0.15**	10.54±2.09	10.61±2.79	36.24±6.99	31.35±3.51
IL7-3	6.31±1.04	2.82±0.27	6.10±0.40*	14.90±3.00	43.07±3.87	39.88±1.82*
B27	3.48±0.24*	0.78±0.05**	5.08±1.14*	9.21±2.49	33.90±4.16	35.35±0.17*
R176	3.10±0.15*	1.33±0.02	5.75±0.24*	13.30±5.81	39.63±4.08	33.83±0.44
R178	3.50±0.30*	1.60±0.14	6.36±0.28*	13.40±2.09	33.65±3.96	38.54±7.19
R179	4.43±1.13	1.60±0.29	7.66±1.74	13.45±6.72	37.65±7.12	25.96±2.66*
R181	4.06±0.97	1.68±0.26	6.98±1.41	14.31±1.67	34.83±7.13	32.44±4.39
R182	5.85±.13	2.14±0.53	9.22±1.13	18.45±3.08*	38.61±11.70	44.81±2.31**
R201	5.17±1.04	2.18±0.26	8.79±2.07	14.37±2.75	43.37±7.96	37.55±6.33
R202	4.15±0.99	1.91±0.12	7.19±2.08	14.66±3.42	41.25±10.37	40.79±3.73*

Data related to parameters linked to the metabolic analyses on fruits, such as Lycopene, β-carotene, total carotenoids and AsA content in the mature green (MG), breaker (BR) and mature red (MR) ripe stages. The significance of differences of each genotype *vs* M82 was evaluated by the t-Student's test (*: $P<0.05$; **:$P<0.01$).

Finally, the morphological analysis of fruit evidenced no significant differences for pericarp area (PA) and thickness (PT) (Table 2.4, Figure 2.1), micro (Micro PA) proximal angles and for the distal angle (DA), whereas fruit area (FA) is significantly reduced in R176 (17.35 cm2) and R201 (16.21 cm^2) compared with M82 (21.32 cm^2), and for fruit perimeter (FP) only R201 is significantly lower compared to M82 (15.5 cm *vs* 17.81 cm). Also, the macro (Macro PA) proximal angles showed significantly lower values from M82 in some sub-lines, particularly R179, R201 and R202, with values of 91.32°, 93.76° and 90.13°, respectively. Only the sub-lines B27, R181 and R182 showed morphological parameters similar to those exhibited by the control genotype M82.

Table 2.4- Morphological analysis of fruit of the sub-lines and their parental genotypes in the year 2016.

Genotype	PA (cm2)	PT (cm)	Micro PA (°)	DA (°)	FA cm2	FP cm	Macro PA (°)
M82	0.42±0.02	0.36±0.02	213.84±14.00	169.56±7.18	21.32±1.67	17.81±0.90	107.50±5.15
IL12-4	0.45±0.03	0.32±0.03	201.79±1.01	170.21±3.21	19.48±0.45	17.16±0.35	93.2±3.97**
IL7-3	0.40±0.03	0.38±0.03	207.22±5.27	171.81±2.08	16.77±1.55*	15.87±0.70*	97.21±3.85
B27	0.42±0.00	0.36±0.01	200.59±15.19	172.51±2.19	20.20±1.66	17.30±0.82	110.67±7.41
R176	0.43±0.03	0.34±0.03	226.55±1.35	170.88±7.18	18.62±0.16	16.94±0.05	93.66±11.71
R178	0.42±0.05	0.36±0.04	312.74±48.79	169.93±2.68	17.35±1.91*	16.19±0.03	104.83±10.05
R179	0.41±0.03	0.36±0.03	210.99±4.84	168.48±3.71	16.80±2.45	15.71±1.26	91.32±1.39*
R181	0.45±0.04	0.33±0.03	212.59±8.10	173.50±6.64	21.08±0.76	17.59±0.36	119.95±6.45
R182	0.44±0.01	0.34±0.01	193.28±25.06	168.48±3.35	18.61±0.53	16.37±0.30	113.47±6.19
R201	0.40±0.02	0.38±0.00	194.57±8.84	175.28±4.42	16.21±0.36*	15.5±0.21*	93.76±1.91*
R202	0.40±0.04	0.37±0.02	194.36±35.77	170.22±0.08	16.78±1.14	16.00±0.81	90.13±0.84*

Data related to parameters linked to morphological analysis: pericarp area (PA), pericarp thickness (PT), fruit area (FA), fruit perimeter (FP), micro proximal angles (Micro PA) and macro proximal angles (Macro PA). The significance of differences of each genotype *vs* M82 was evaluated by the t-Student's test (*: $P<0.05$; **:$P<0.01$).

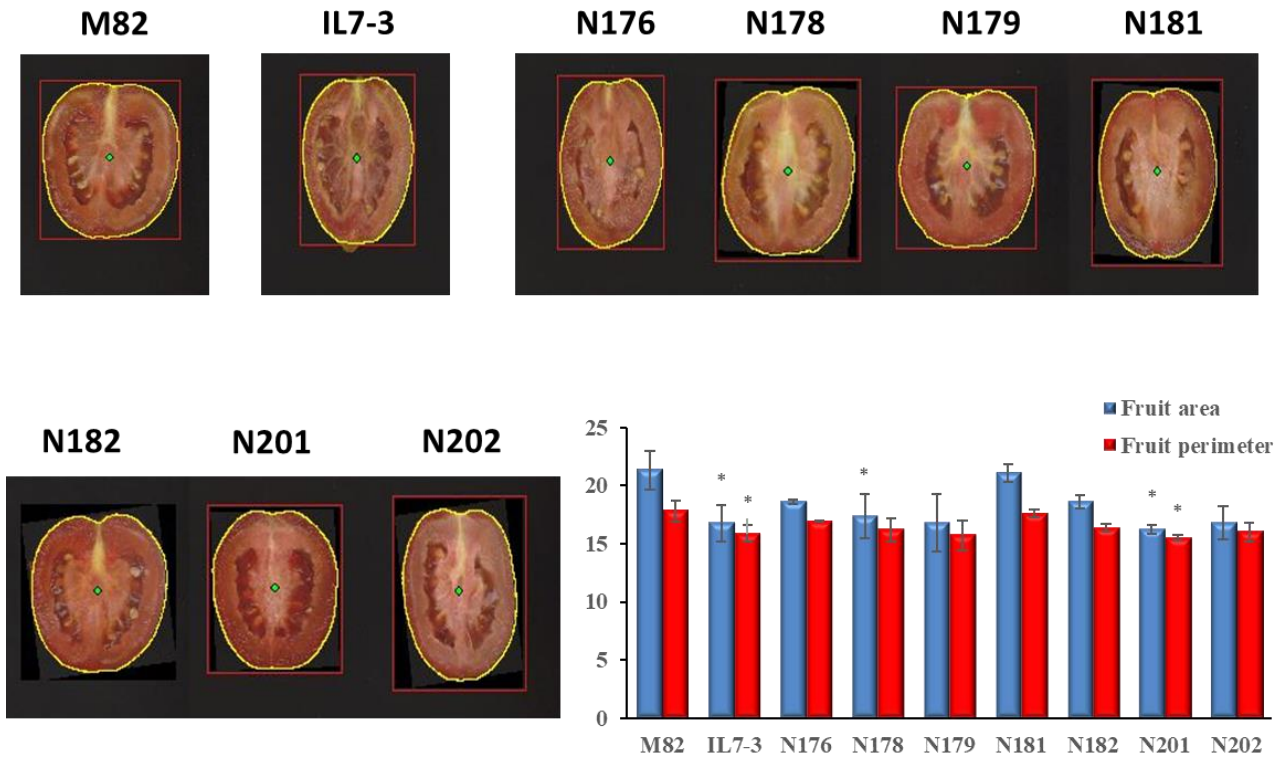


Figure 2.1- Morphological analysis of fruit of the IL7-3 sub-lines and their parental genotypes in the year 2016.

Analyses were carried out using the software Tomato Analyzer. Pictures were obtained by longitudinal cutting of the fruit; the yellow line shows the fruit perimeter and area. Measures of fruit area (cm^2) and perimeter (cm) are reported in the graph for parental genotypes and the sub-lines. The significance of differences of each genotype *vs* M82 was evaluated by the t-Student's test (*: $P<0.05$).

2.2.2 Phenotypic characterization during the year 2017

In the summer 2017 analyses were restricted to the most relevant parameters among those measured in the 2016, with the aim of confirming the data of B27 and the selection of the best IL7-3 sub-lines in three different environments of South Italy (Table 2.5). As for parameters strictly concerning the productivity (FW, NFR and YP), the parental line IL7-3 showed significant lower values than M82 for all traits in all experimental fields, thus confirming the lower productivity exhibited in the previous year. Comparing data of 2017 in the three different environments, no GxE interaction at the two-way ANOVA test was observed for FW ($p=0.446$), as well as for NFR ($p=0.289$) and YP ($p=0.315$). Comparing the FW of the sub-lines with M82, significant differences were mainly observed in the Acerra field, where R178, R179, and R202 showed a significant reduction compared to M82, confirming data obtained in the year 2016. In Giugliano, only R182 exhibited a significantly higher value compared with M82 (50.29 g vs 42.39 g) and in Battipaglia R202 and IL12-4 showed a significant lower value respect to M82. Among the sub-lines, R176 and B27 showed a significant reduction of NFR in Giugliano and only R179 in Battipaglia field, whereas R182 and the other sub-lines exhibited values comparable to M82 in the three environmental conditions tested, except R178 that in Acerra showed a significant higher value. Finally, as for yield, the sub-lines R176, R178, R181 and R182 confirmed no differences compared to M82 in all the three fields whereas R179, R201 and R202 showed significantly lower value in Battipaglia.

Table 2.5- Productivity of the sub-lines and their parental genotypes in three different enviroments in the year 2017.

Genotype	Fruit weight g			Fruits/plant No.			Yield/plant g		
	Acerra	**Giugliano**	**Battipaglia**	**Acerra**	**Giugliano**	**Battipaglia**	**Acerra**	**Giugliano**	**Battipaglia**
M82	51.22±2.41	42.39±2.49	60.14±4.30	33.19±2.69	60.76±8.59	104.48±19.45	1.68±0.14	2.60±0.73	3.36±0.30
IL12-4	55.11±5.00	42.07±4.13	47.01±5.56*	31.96±2.51	56.56±10.69	110.10±59.54	1.75±0.28	2.38±1.49	3.16±0.37
IL7-3	44.89±2.52*	36.78±2.25***	34.33±1.70*	17.89±2.48*	24.00±1.65*	48.58±11.00*	1.10±0.05*	1.30±0.27*	1.17±0.55**
B27	55.22±8.49	48.93±5.14	58.69±0.27	33.40±4.72	38.65±1.55*	123.89±30.49	1.71±0.18	2.15±0.20	3.73±1.56
R176	47.56±3.90	49.02±7.37	51.58±18.63	34.24±3.72	38.67±5.83*	117.0±57.07	1.61±0.17	2.65±0.93	3.15±1.12
R178	40.33±5.67*	36.82±6.32	42.71±12.15	42.56±3.75*	58.42±7.70	102.63±53.18	1.69±0.20	2.63±0.32	2.87±0.97
R179	43.33±2.67**	39.34±5.56	53.95±20.07	23.74±4.99	38.86±7.28	42.13±28.49*	1.01±0.22	2.03±0.33	1.12±0.39*
R181	56.44±6.30	45.57±9.63	59.74±5.72	32.14±1.03	42.93±15.83	106.51±41.69	1.79±0.20	2.33±0.61	3.82±1.60
R182	51.56±5.35	50.29±2.25*	58.03±9.43	36.55±12.55	54.29±10.94	114.30±38.37	1.84±0.63	2.70±1.23	3.51±0.47
R201	39.44±4.60	36.58±2.89	44.67±12.68	39.05±6.93	50.31±18.56	84.09±25.86	1.55±0.41	2.51±1.01	1.99±0.45**
R202	40.26±1.06**	37.29±5.21	46.07±4.93*	32.63±3.23	41.58±14.00*	81.23±12.31	1.31±0.23	2.11±0.74	2.06±0.53*

	F	**Sign**	**TSS%**	**F**	**Sign**	**TSS%**	**F**	**Sign**	**TSS%**
ENVIRONMENT	11.6	0	0.7	66.23	0	15.21	22.22	0	4.1
GENOTYPE	5.82	0	1.77	2.67	0.009	3.07	4.65	0	4.29
GxE	1.03	0.446	0.62	1.19	0.289	2.74	1.16	0.315	2.14

The Two-way ANOVA test results are reported for each trait as F value, to estimate the environment, the genotype and their interaction effect, considering as level of significance a<0.05. The percentage of total sum of squares (TSS%) for E, G and GxE is calculated dividing the TSS of each variable by the grand total TSS. The significance of differences of each genotype *vs* M82 was evaluated by the t-Student's test (*:P<0.05; **:P<0.01; ***:P<0.001).

Qualitative analyses carried out in 2017 were restricted to °Brix, firmness and ascorbic acid content in the ripe fruit, since these traits were the only showing relevant differences during the year 2016 (Table 2.6). As for °Brix, data collected in the year 2017 confirmed those of the year 2016, indeed significant differences were observed for most sub-lines and IL7-3 respect to M82 in Acerra field, as well as in Giugliano. Particularly, R182 exhibited a significantly high value in Acerra during both the years 2016 and 2017. In Acerra field, also R178, R181 and R202 showed a significantly higher value than M82, though only R178 and R202 confirmed the value exhibited in the year 2016. However, no differences were observed in Battipaglia whereas in Giugliano IL12-4, B27 and R201 were significantly different. Indeed, this trait evidenced a high GxE interaction ($p=0.003$). As for firmness, the two parental genotypes and most sub-lines showed very similar values in all the fields, with a not significant GxE interaction ($p=0.114$). The only relevant result was the significantly higher firmness observed in R182 in all cases (values always higher than 8 kg/cm^2). A significant higher value of AsA compared with M82 was evidenced in IL7-3, IL12-4 and B27, whereas most sub-lines showed a high variability, evidencing the presence of GxE interaction ($p=0.006$). The sub-line R182 was the only showing a higher level of AsA respect to M82 in all the three experimental fields confirming the data of Acerra 2016. Furthermore, in the year 2017 the AsA content was also evaluated in leaves in Acerra and Giugliano (Table 2.7) to understand if R182 showed a high value of this metabolite also in leaves. The AsA level in the leaves was not significantly different between the parental lines M82 IL12-4 and IL7-3. Comparing the AsA content of all sub-lines with M82, in Acerra only B27 and R182 showed a significant increase of AsA in the leaves (19.24 mg/100 g FW and 19.42 mg/100 g FW, respectively, vs 13.40 mg/100 g FW), indeed the other sub-lines exhibited values from 11 mg/100 g FW to 16 mg/100 g FW. Also in Giugliano, B27 and R182 showed a significantly higher value respect to M82 (12.01 mg/100 g FW and 14.90 mg/100 g FW, respectively, vs 9.93 mg/100 g FW) whereas R176, R179 and R202 exhibited a significant reduction of AsA respect to M82 with values of 7.04 mg/100 g FW, 6.01 mg/100 g FW and 6.32 mg/100 g FW, respectively. Finally, a not significant GxE interaction ($p=0.10$) was observed for AsA in the leaves.

Table 2.6- Fruit quality of the sub-lines and their parental genotypes in three different enviroments in the year 2017.

Genotype	Brix (°)			Firmness			AsA mg/100 g FW		
	Acerra	**Giugliano**	**Battipaglia**	**Acerra**	**Giugliano**	**Battipaglia**	**Acerra**	**Giugliano**	**Battipaglia**
M82	4.60±0.20	4.07±0.57	4.97±0.47	7.07±0.52	6.49±0.15	7.34±0.39	44.35±6.55	27.38±2.04	26.10±3.24
IL12-4	4.90±0.62	5.37±0.32*	4.87±0.67	missingvalue	missingvalue	missingvalue	56.46±1.02*	37.24±2.05**	41.97±1.13**
IL7-3	6.47±0.32**	6.00±0.26**	5.30±0.36	6.52±0.88	6.42±0.66	6.38±0.82	64.21±5.87*	38.17±3.08**	40.87±3.70**
B27	4.47±0.51	5.53±0.64*	4.43±0.50	missingvalue	missingvalue	missingvalue	55.31±1.39*	43.66±2.52**	38.95±1.36*
R176	4.90±0.26	4.40±0.46	4.57±0.15	7.73±0.90	6.27±0.15	6.95±0.75	39.90±6.53	16.51±5.36*	22.26±2.80
R178	5.93±0.47*	5.40±0.72	4.73±0.35	5.98±0.75	5.47±0.57*	6.64±1.02	51.75±9.90	24.63±0.40	26.96±4.98
R179	5.13±0.85	4.73±0.42	5.30±0.53	6.70±1.43	4.81±0.33**	7.38±0.39	43.91±8.97	31.13±9.67	15.10±5.88*
R181	5.10±0.17*	4.97±1.07	4.87±0.51	8.46±0.81	7.16±1.50	6.46±0.85	49.67±2.96	35.81±8.45	30.34±3.97
R182	5.93±0.55*	4.96±0.31	3.93±0.38	8.32±0.57*	8.29±0.54**	8.77±0.53*	57.86±3.77*	34.82±2.21*	34.34±1.24*
R201	4.97±0.15	5.30±0.26*	5.40±0.87	6.53±0.39	6.12±0.56	5.94±0.82*	46.87±3.17	35.93±3.31*	25.56±4.32
R202	5.63±0.38*	5.47±0.84	4.70±0.35	5.86±0.53*	6.06±0.21*	6.98±0.76	55.26±6.47	33.08±4.40	25.53±1.18

	F	**Sign**	**TSS%**	**F**	**Sign**	**TSS%**	**F**	**Sign**	**TSS%**
ENVIRONMENT	6.386	0.003	0.132	4.943	0	0.154	190.9906137	0	5.595
GENOTYPE	4.806	0	0.499	8.5	0.011	1.058	18.7580631	0	2.748
GxE	2.528	0.003	0.525	1.555	0.114	0.387	2.329985841	0.006	0.682

The Two-way ANOVA test results are reported for each trait as F value, to estimate the environment, the genotype and their interaction effect, considering a level of significance a<0.05. The percentage of total sum of squares (TSS%) for E, G and GxE is calculated dividing the TSS of each variable by the grand total TSS. The significance of differences of each genotype *vs* M82 was evaluated by the t-Student's test (*: P<0.05; ** :P<0.01).

Table 2.7- AsA content in leaves of the sub-lines and their parental genotypes in Acerra and Giugliano in the year 2017.

Genotype	AsA leaves mg/100g FW	
	Acerra	**Giugliano**
M82	13.40±0.85	9.93±0.23
IL12-4	16.45±2.12	10.35±0.14
IL7-3	9.18±1.28	9.48±2.17
B27	19.24±1.91**	12.01±0.52*
R176	13.31±1.87	7.04±0.55*
R178	11.95±0.72	6.99±1.17
R179	11.71±1.34	6.01±0.19*
R181	16.91±0.95	11.64±1.04
R182	19.42±1.75**	14.90±0.81*
R201	14.12±1.20	8.89±0.67
R202	13.40±0.85	6.32±0.12**

	F	**Sign**	**TSS%**
ENVIRONMENT	128.21	0	4.02
GENOTYPE	14.62	0	4.58
GxE	2.02	0.06	0.63

The Two-way ANOVA test results are reported for each trait as F value, to estimate the environment, the genotype and their interaction effect, considering a level of significance a<0.05. The percentage of total sum of squares (TSS%) for E, G and GxE is calculated dividing the TSS of the each variable by the grand total TSS. The significance of differences of each genotype *vs* M82 was evaluated by the t-Student's test (*: P<0.05; ** :P<0.01).

For a general selection of the IL7-3 sub-lines based on all evaluated traits during the year 2016, an Evaluation Index (EI) was assayed, which takes into consideration the scores obtained by each line compared to the control genotype M82. The EI estimated for productivity and fruit quality traits evidenced that the best sub-line in terms of quality (QS) and yield (YS) score was R182, which exceeded the control genotype M82 for QS and had the same value for QY, as shown in the scatter diagram of Figure 2.2 A. All the other sub-lines exhibited values of QS and YS worse than M82, and intermediate values between the two parental genotypes M82 and IL7-3. The EI was also estimated on data collected in the year 2017 for the three productive (FW, NFR, YP) and the three fruit quality traits (°Brix, FI, AsA) evaluated in three different environmental conditions.

The scatter diagram (Figure 2.2 B) evidences that the sub-line R182 exhibited the highest QS and YS, and that sub-lines R178 and R181 were comparable to M82. Therefore, following the second year of phenotypic analyses carried out in three different environmental conditions, the sub-line R182 was confirmed as the elite one.

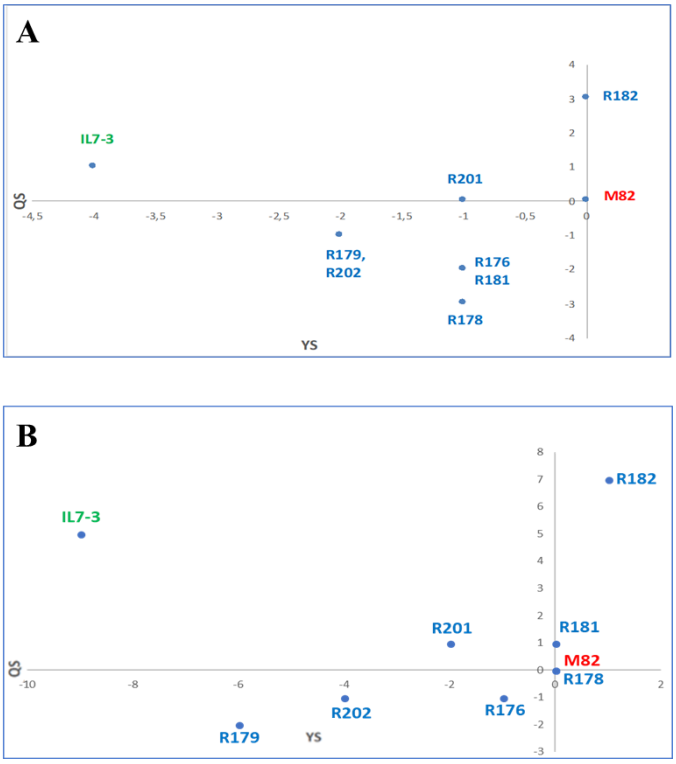


Figure 2.2-Scatter diagram of the Evaluation Indexes (EIs) measured for the sub-lines and IL7-3 in term of quality (QS) and yield (YS), in comparison to the control genotype M82, whose EI was fixed to zero.
A) Scatter plot of EIs measured in the year 2016; B) Scatter plot of EIs measured in the year 2017.

Chapter 3. Identification of candidate genes controlling high AsA content in the selected sub-lines

3.1 Materials and methods

3.1.1 Raw data materials

DNA-seq materials used for the assembly of the M82 S. *lycopersicum* genome (Table 3.1) were retrieved from publicly available M82 raw data (Bolger et al., 2014). RNA-seq and RAD-seq materials used for the assembly of IL12-4 and IL7-3 genomes were publicly available raw data (Chitwood et al., 2013).

Table 3.1- RNA-seq, DNA-seq and RAD-seq material used for the assembly.

		DNA-seq		
	BioProject	**Center name**	**Platform**	**SRA Study**
M82	PRJEB6302	RWTH AACHEN	ILLUMINA	ERP005818
		RNA-seq		
ILs	PRJNA222538	DEP. OF PLANT BIOLOGY, UNIVERSITY OF CALIFORNIA, DAVIS	ILLUMINA	SRP031491
		RAD-seq		
ILs	PRJNA222545	DEP. OF PLANT BIOLOGY, UNIVERSITY OF CALIFORNIA, DAVIS	ILLUMINA	SRP031490

Information related the BioProject number, the name of the center where analyses were carried out, the platform used to prepare the RNA and DNA libraries and the Sequence Read Archive (SRA) study of RNA-seq and RAD-seq experiments are reported.

3.1.2 *S. lycopersicum* cv M82 genome assembly

The SRA files were converted to FASTQ file using fastq-dump of SRA Toolkit (http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=toolkit_doc&f=fastq-dump). FASTQ files were checked for quality using FASTQC [Andrews S. (2015 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/]. The adapter

sequences, poly-N stretches and low-quality reads (Phred score <20) were removed using the BBDuck module of the BBMap software package (version 34_90, http://sourceforge.net/projects/bbmap).

The reconstruction analysis has been performed using Heinz SL3.0 (ftp://ftp.solgenomics.net/genomes/Solanum_lycopersicum/assembly/build_3.00/S_lycopersicum_chromosomes.3.00.fa) as reference genome. The genome reconstruction analysis was carried out using the Reconstruction tool (v.1.0). The reconstruction pipeline is based on an iterative read mapping and a *de-novo* assembly (Scala, Grottoli et al., 2017).

3.1.3 *S. lycopersicum* cv M82 annotation

The *S. lycopersicum* cv M82 annotation has been performed using Heinz ITAG 3.2 (ftp://ftp.solgenomics.net/genomes/Solanum_lycopersicum/annotation/ITAG3.2_release/ITAG3.2_cDNA.fasta) and NCBI annotated genomes release 102 (ftp://ftp.ncbi.nlm.nih.gov/genomes/Solanum_lycopersicum/RNA) as reference annotations. The genome annotation was carried out effecting two liftover using Maker programs (http://www.yandell-lab.org/software/maker.html). The quantitative measures for the assessment of the assembled genome and transcriptome completeness were carried out using BUSCO v3 program (https://busco.ezlab.org/).

3.1.4 IL12-4 and IL7-3 assembly

The IL7-3 and IL12-4 break points and introgression size were defined using an IL check pipeline (Figure 3.1) kindly provided by Dr. Valentino Ruggieri (CRAG Centre for Research in Agricultural Genomics Campus UAB - Edifici CRAG | 08193 Cerdanyola | BARCELONA). The SRA files (RNA-seq and RAD-seq) were converted to FASTQ files using FASTQ-DUMP. FASTQ files were checked for quality using FASTQC. Trimming and merging were done using the Trimmomatic software. Then, RNA-seq data were aligned against the *S. lycopersicum* reference genome sequence (SL3.2 release) and *S. pennellii* genome

(annotation release ID:100) using STAR aligner (version 2.4.2a), and the variant calling of RNA-seq and RAD-seq data were performed using the SUPER W pipeline (Sanseverino et al., 2015). Then, variants were filtered and the ILs definition and assembly were carried out.

3.1.5 Molecular marker analysis

SCAR and CAPS markers were defined in order to estimate the size of the wild region present in one IL sub-line (R182). Genomic DNA was extracted from young leaves using the PureLinkTM Genomic DNAKit. Primers for the PCR amplification were designed based on polymorphisms detected in the IL7-3 introgression region between the *S. lycopersicum* (SL3.0 assembly and iTAG3.2 annotation) and the *S. pennellii* (v2 Assembly) genomes, by investigating the Genome Browser available in the Sol Genomics Network database (http://www.tomatogenome.net/VariantBrowser). PCR amplification was carried out in 50 µl reaction volume containing 50 ng DNA, 1X reaction buffer, 0.2mM each dNTP, 1.0 mM primer and 1.25 U GoTaq polymerase (Promega). For designing CAPS markers, restriction enzymes suitable to detect polymorphic SNPs between the amplified fragments of the two species were found using the tool CAPS Designer available at the Sol Genomics Network (https://solgenomics.net/). Amplified and restricted fragments were visualized on agarose gel at different concentrations depending on their expected size.

3.1.6 RNA-seq analysis

RNA extraction was performed on B27, M82 and R182 mature red fruits and M82 and R182 breaker fruits cultivated in Acerra field using the Trizol reagent (Thermo Fischer Scientific) and following the manufacturer's guidelines. RNA sequencing experiment was performed using the "TrueSeq mRNA" protocol for preparing libraries and a pair-end strategy was chosen for sequencing. The RNA-seq analysis was carried out using the new Sequentia Biotech data analysis software “AIR” (https://transcriptomics.sequentiabiotech.com/) and the *S. lycopersicum* reference genome sequence (SL3.20 release). The differential expression analysis was performed using edgeR.

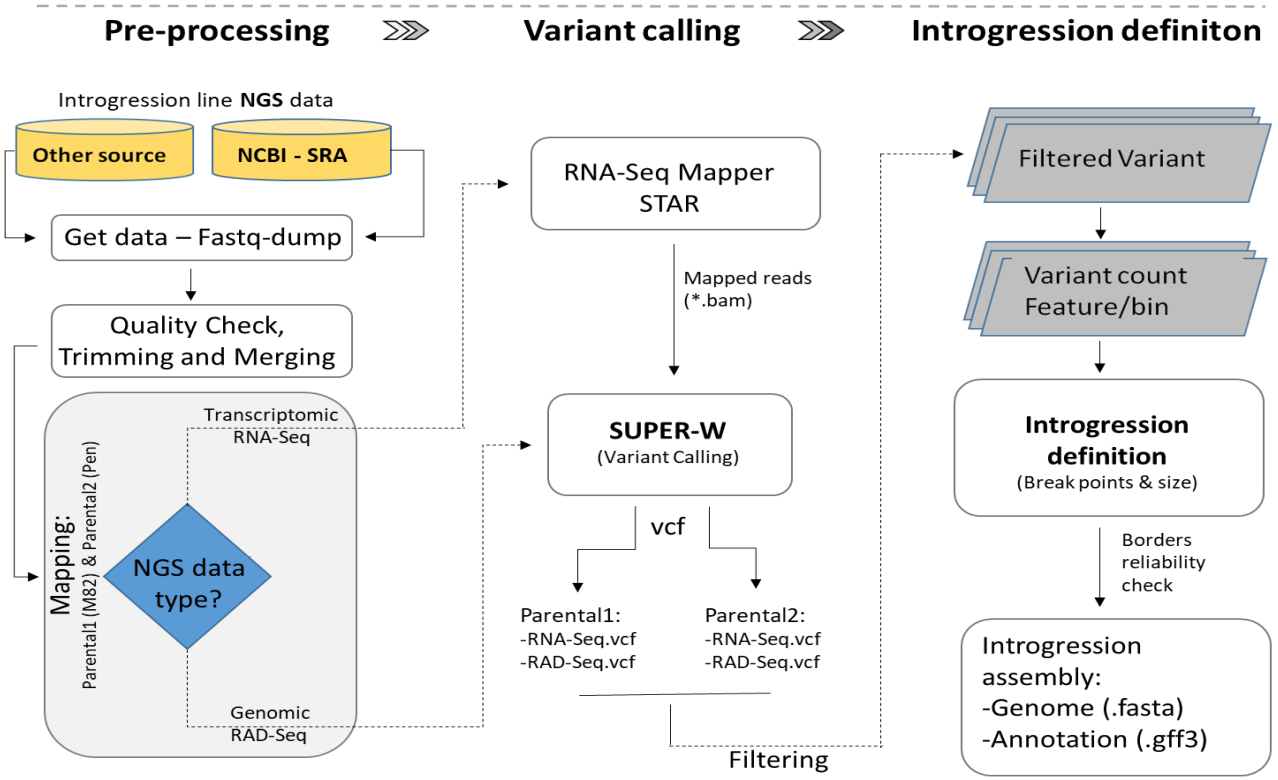


Figure 3.1-IL check pipeline kindly provided by Dr. Valentino Ruggieri, showing all the steps related to pre-processing data, variant calling and introgression definition.

3.2 Results

3.2.1 *S. lycopersicum* cv M82 genome reconstruction

The *S. lycopersicum* M82 genome assembly was obtained using a Reconstruction pipeline (Figure 3.2). As for the first step (Table 3.2), six iteractions were performed. For each iteraction all the variants were filtered and added to the reference genome to get an intermediate genome. In the Iteration 1, 1,188,034 variants were found with deletion (del) and insertion (ins) size of 96,403 bp and 111,813 bp respectively, a mean coverage of 39.58 and a genome size of 828.092 Mb and a genome size variation respect to Heinz of 0.002%. As for the Iteration 2, 65,929 variants were detected with a M82 intermediate genome size of 828.091 Mb and a size variation between M82 and Heinz of -0.0001%, whereas the Iteration 3 showed 12,405 variants with a genome size of 828.090 Mb and a genome size variation of M82 respect Heinz of 0.0017%. As for Iteration 4, 4,385 variations were found and a genome size variation between M82 and Heinz of -0.000018%, whereas the Iterations 5 and 6 exhibited only 1,943 and 643 variations respectively and a genome size variation of -0.20% for the interaction 6. As for the Step 1, 1,273,341 variants were identified with a del and ins size of 1,799,145 bp and 185,818 bp, respectively, and a total genome size variation of -0.20% among M82 and Heinz (826.444 Mb M82 vs 828.077 Mb Heinz). However, most of the variations were in the intergenic regions and only 10% of these were identified in coding regions. As for the Step 2, no mapped reads were assembled in contigs in order to add them to the new genome, indeed the Reconstruction pipeline kept 212 contigs with minimum length of 400 bp, thought only 1 contig with length of 450 bp was added to chromosome 10 of M82 genome in the Step 3. As for the Step 4 the new M82 genome with a size of 826,444 Mb was assembled.

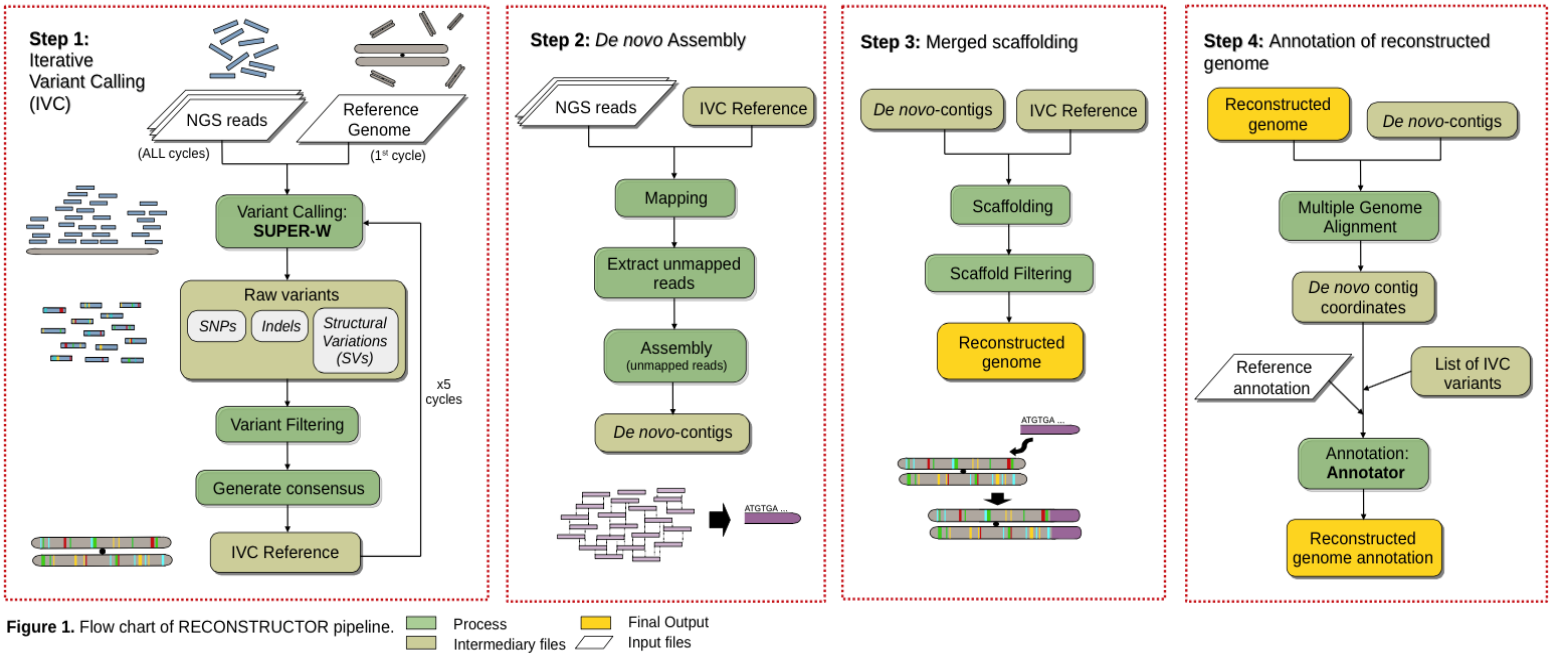


Figure 1. Flow chart of RECONSTRUCTOR pipeline.

Figure 3.2- Flow chart of the Reconstruction pipeline kindly provided by Dr. Walter Sanseverino. Step 1) calling variant in the common regions between reference and resequencing experiment; Step 2) de novo assembly of private regions of resequencing experiment; Step 3) merging data of common and private regions; Step 4) new genome annotation.

Table 3.2- Step 1 of the Reconstructor pipeline.

Step	Nº variant	Del size (bp)	Ins size (bp)	Mean_coverage	Genome size	Variation
Iteration 1	1188034	96403	111813	39.58	828092366	0.001861%
Iteration 2	65929	8799	7637	39.73	828091204	-0.000140%
Iteration 3	12405	1483	1011	39.76	828090732	0.001664%
Iteration 4	4385	413	268	39.77	828090587	-0.000018%
Iteration 5	1943	164	100	39.77	828090523	-0.000008%
Iteration 6	645	1691883	64989	43.89	826444336	-0.198793%
Tot	1,273,341	1,799,145	185,818	40.41666667	-1,613,327	-0.197158%

For each iteration the variants were added to the reference genome to get an intermediate genome. The variation percentage represents the variation of the genome size for each iteration.

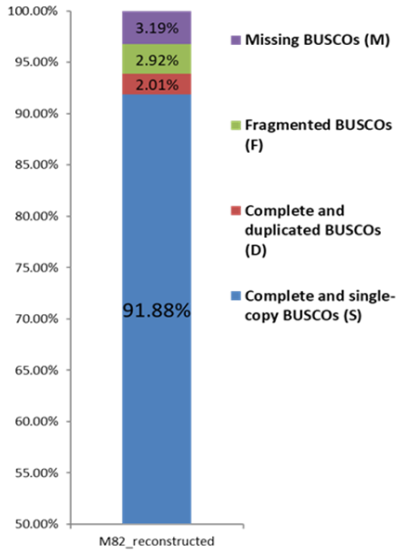
The M82 genome annotation was carried out using the tomato annotations SGN SL 3.1 and NCBI v.102. As for the Tomato Annotation SGN SL 3.1, 37,768 transcripts were considered, whereas for the Tomato Annotation NCBI v.102 35,921 transcripts (Table 3.3).

As the first liftover (SGN SL3.1), 35,364 transcripts passed the quality filter and were correctly assigned to each chromosome, whereas only 404 transcripts were excluded as they did not pass the quality filter. As for the second liftover (NCBI v.102), 32,921 transcripts passed the quality filter and were correctly assigned to each chromosome and 3,459 were excluded. As for the transcripts passed for each annotation, almost 99.8% were in common and among these, SGN SL 3.1 transcripts were preferred, whereas only 2,675 NCBI transcripts, not in common between the annotations, were added to the M82 genome annotation for a total of 38,443 transcripts. Furthermore, the quantitative measures for the assessment of the assembled genome and transcriptome completeness using Busco v3 (Figure 3.3) showed 1,323 (91.88%) recovered genes found in single copy and only 29 (2.09%) recovered genes found in more than one copy. As for fragmented and missing genes, only 42 (2.92%) were considered fragmented genes as partially recovered and 46 (3.19%) considered missing genes as not recovered.

Table 3.3- Results of SGN SL3.1 and NCBI v.102 liftovers using the Maker software.

	SGN SL3.2	**NCBI v.102**
Transcripts	35768	35921
Passed	35364	32462
Excluded	404	3459
Used	35364	2675
M82 Ttanscripts	38443	

A



B

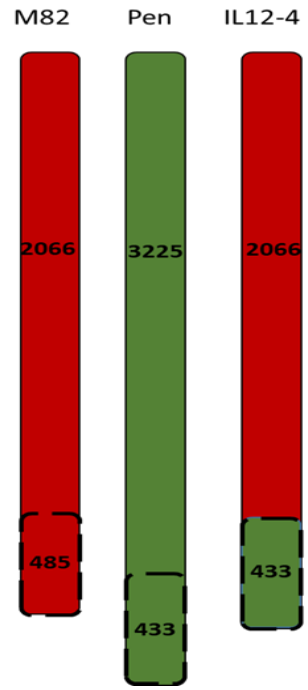
BUSCO	M82_annot
N of proteins	37010
Complete BUSCOs (C)	1352
Complete and single-copy BUSCOs (S)	1323
Complete and duplicated BUSCOs (D)	29
Fragmented BUSCOs (F)	42
Missing BUSCOs (M)	46
Total BUSCO groups searched	1440

Figure 3.3- Quantitative measures for the assessment of the assembled genome and transcriptome completeness using Busco v3. A) percentage of: complete and single copy genes (S); complete and duplicated copy genes (D); fragmented genes (F) and missing genes (M). B) results of BUSCO v3 analysis considering 37, 010 proteins.

3.2.2 IL12-4 assembly

Break points and introgression size of IL12-4 were defined by mapping the reads coming from RNA-Seq data on both the parental genomes (*S. pennellii* and the reconstructed M82). By doing this, we identified an exhaustive set of markers that allowed precisely defining both the break points on the parent M82 and the extension of the introgression of the donor parent *S. pennellii*. Finally, the introgression was re-assembled. The line IL12-4 (Figure 3.4) was built by replacing on chr12 of M82 a region of 3.088 Mbp and harbouring 485 genes (447 in Heinz) by a region of 3.398 Mbp derived from chr12 of *S. pennellii* and harbouring 433 genes. The introgression size was identified looking at the variation pattern of chr12 of IL12-4 respect to chr12 of M82, whereas break points were found considering the variation pattern of chr12 of IL12-4 respect to chr12 of *S. pennellii*. In Figure 3.5, comparing the variation of chr12 of IL12-4 with the same chromosome of both the parental genotypes, it was possible to identify exactly the introgression fragment. Furthermore, in order to verify the borders reliability of the introgression fragment, the break points were compared with the genetic expression score. The comparison (Figure 3.6) showed that Solyc12g088840 is the upper border of the introgression and Solyc12g100357, which is the last gene of chr12, was the lowest border, whereas the comparison among the *S. pennellii* material (Supplemental figure S3.1) evidenced that Sopen12g030710 was the first and the Sopen12g035030 the last gene of the introgression fragment.

A



B

	Penn. Genes	M82 genes	Size Penn. Region (Mbp)	Size M82 Region (Mbp)
IL12-4	433	485	3.398	3.088

Figure 3.4- IL 12-4 genome assembly.

A) Chromosome 12 of IL12-4 assembly; B) number of M82 genes removed, *S. pennellii* genes added and the fragments size removed and added.

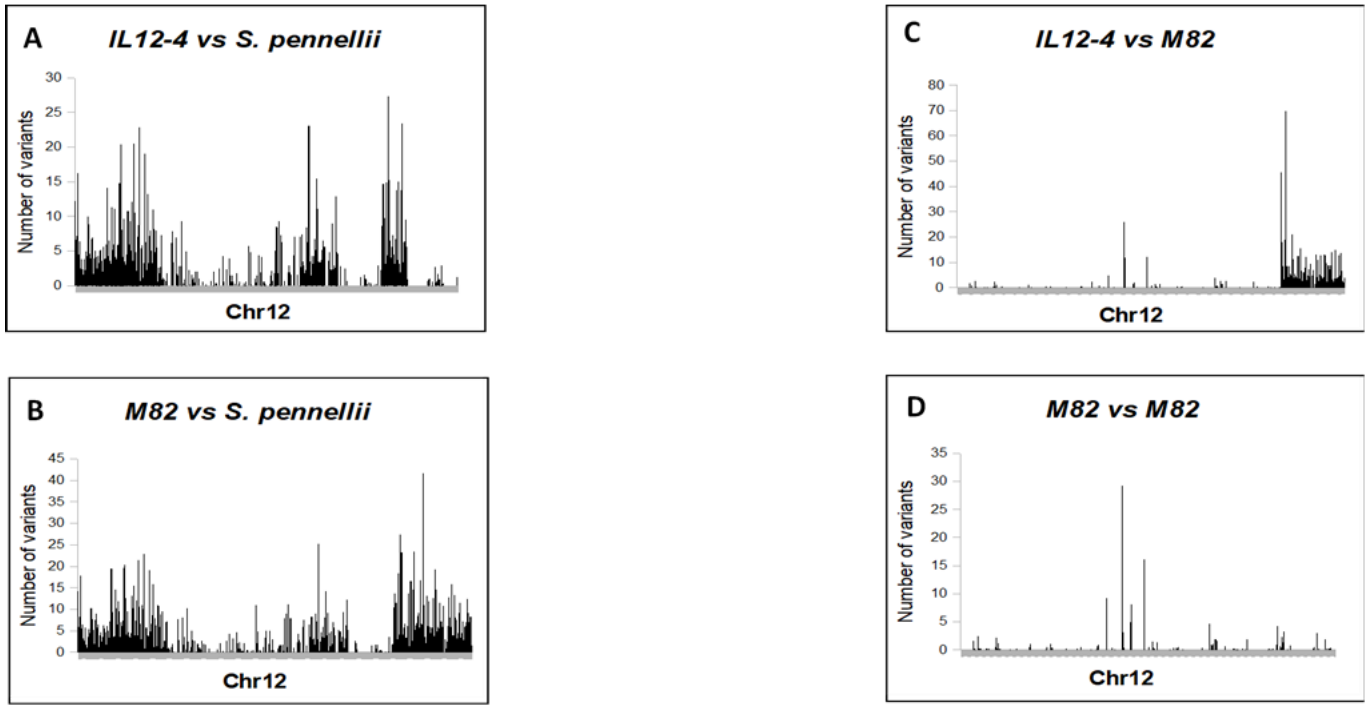


Figure 3.5- Comparison of variants among the chromosome 12 of IL12-4 and of the parental genotypes.

A) and B) Identification of the introgression break points. The introgression fragment showed no variations compared with *S. pennellii*

C) and D) Identification of the introgression size. The introgression fragment exhibited variations compared with *S. lycopersicum* cv. M82.

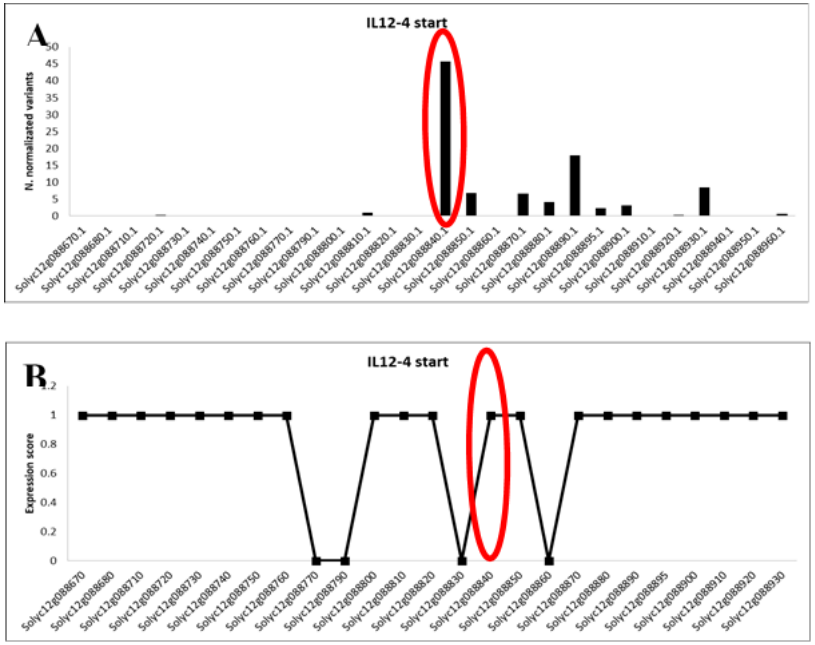
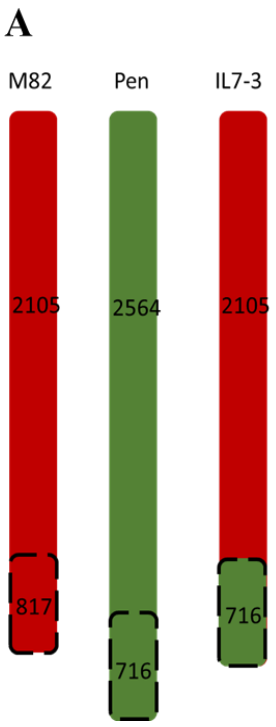


Figure 3.6- Borders reliability check of the introgression fragment to identify the exact upper introgression border.

The red cycle shows the first gene removed from M82. A) Number of normalized variations for each gene found using the IL check pipeline; B) Expression score of each gene. The value 1 indicates that the genes were expressed whereas the value 0 genes that they were not expressed, considering a threshold expression value of 0.05.

3.2.3 IL7-3 assembly

Break points and introgression size of IL7-3 were defined by mapping the reads coming from RNA-Seq data on both the parental genomes (*S. pennellii* and the reconstructed M82). By doing this, we identified an exhaustive set of markers that allowed precisely defining both the break points on the parent M82 and the extension of the introgression of the donor parent *S. pennellii*. Finally, the introgression was re-assembled. The line IL7-3 (Figure 3.7) was built by replacing on chr07 of M82 a region of 6.660 Mbp and harbouring 817 genes (745 in Heinz) by a region of 7.142Mbp derived from chr07 of *S. pennellii* and harbouring 716 genes. The introgression size was identified looking at the variation pattern of chr07 of IL7-3 respect to chr07 of M82, whereas break points were found considering the variation pattern of chr07 of IL7-3 respect to chr07 of *S. pennellii*. In Figure 3.8 comparing the variation of chr07 of the introgression line with the same chromosome of both the parental genotypes, it was possible to identify exactly the introgression fragment. Furthermore, in order to do the borders reliability check of the introgression fragment, the break points were compared with the expression score. The comparison (Figures 3.9 and 3.10) showed that Solyc07g048010 was the upper border of the introgression and Solyc07g063390 was the lower border, whereas the comparison among the *S. pennellii* material (Supplemental figure S3.2, S3.3) exhibited that the Sopen07g024420 was the first and the Sopen07g031590 the last gene of the introgression fragment.



B

	Penn. Genes	M82 genes	Size Penn. Region (Mbp)	Size M82 Region (Mbp)
IL7-3	716	817	7.142	6.66

Figure 3.7- Il 7-3 genome assembly.

A) Chromosome 7 of IL7-3 assembly; B) number of M82 genes removed, *S. pennellii* genes added and the fragments size removed and added.

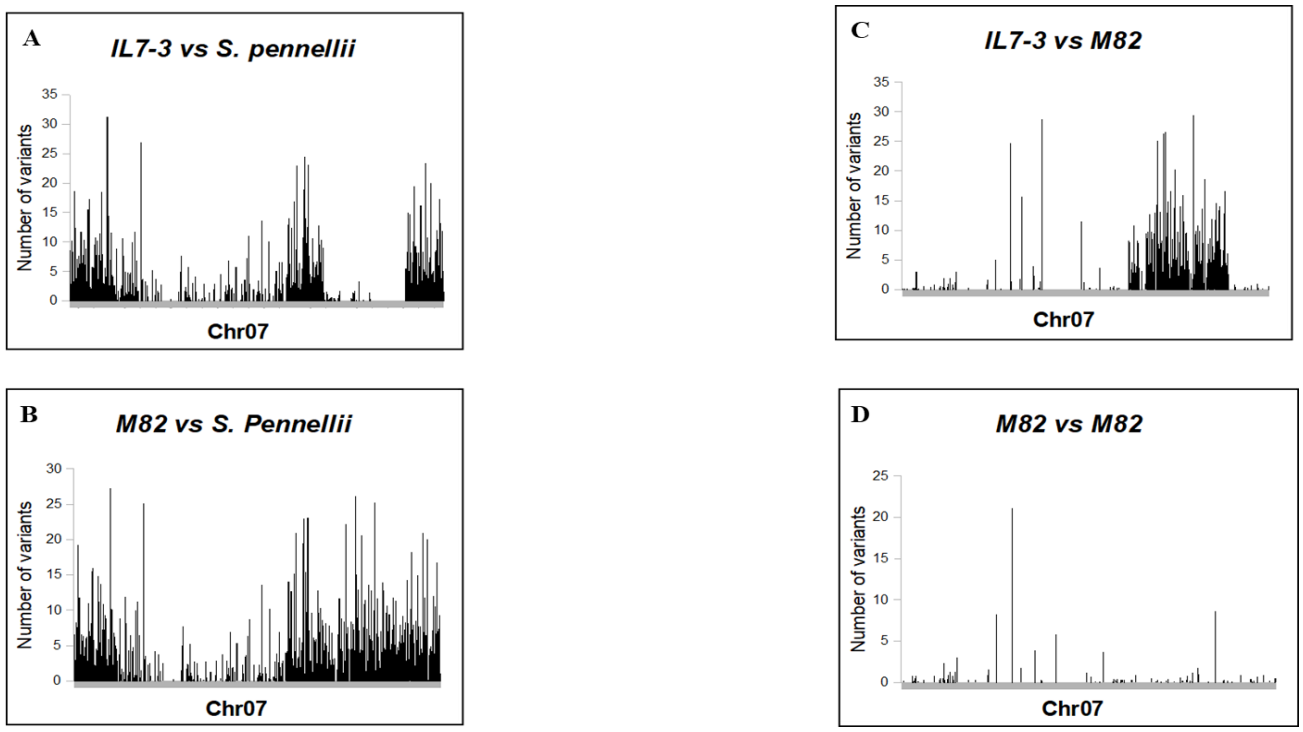


Figure 3.8- Comparison of variants among the chromosome 7 of IL7-3 and the parental genotypes.

A) and B) Identification of the introgression break points. The introgression fragment showed no variations compared with *S. pennellii* C) and D) Identification of the introgression size. The introgression fragment exhibited variations compared with *S. lycopersicum* cv. M82.

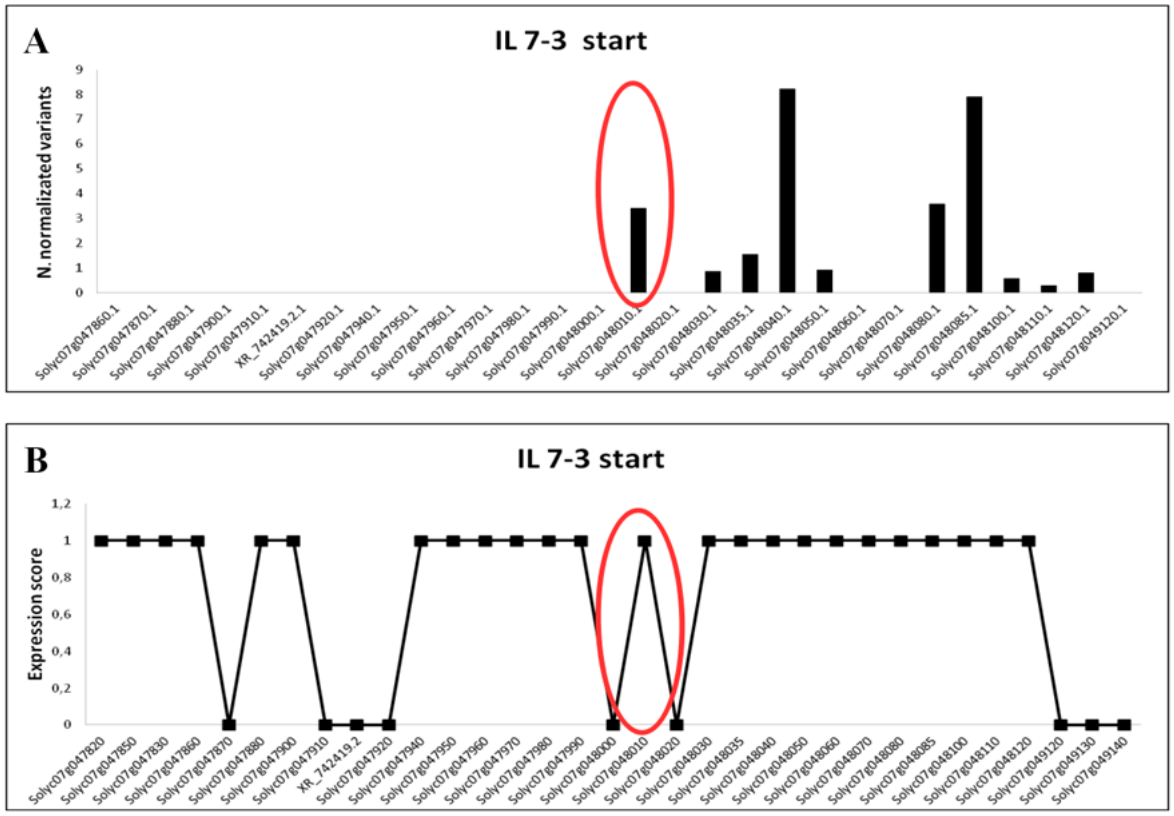


Figure 3.9- Borders reliability check of the introgression fragment to identify the exact upper introgression border.

The red cycle shows the first gene removed from M82. A) Number of normalized variations for each gene B) Expression score for each gene. The value 1 indicates that the genes were expressed whereas the value 0 genes that they were not expressed, considering a threshold expression value of 0.05.

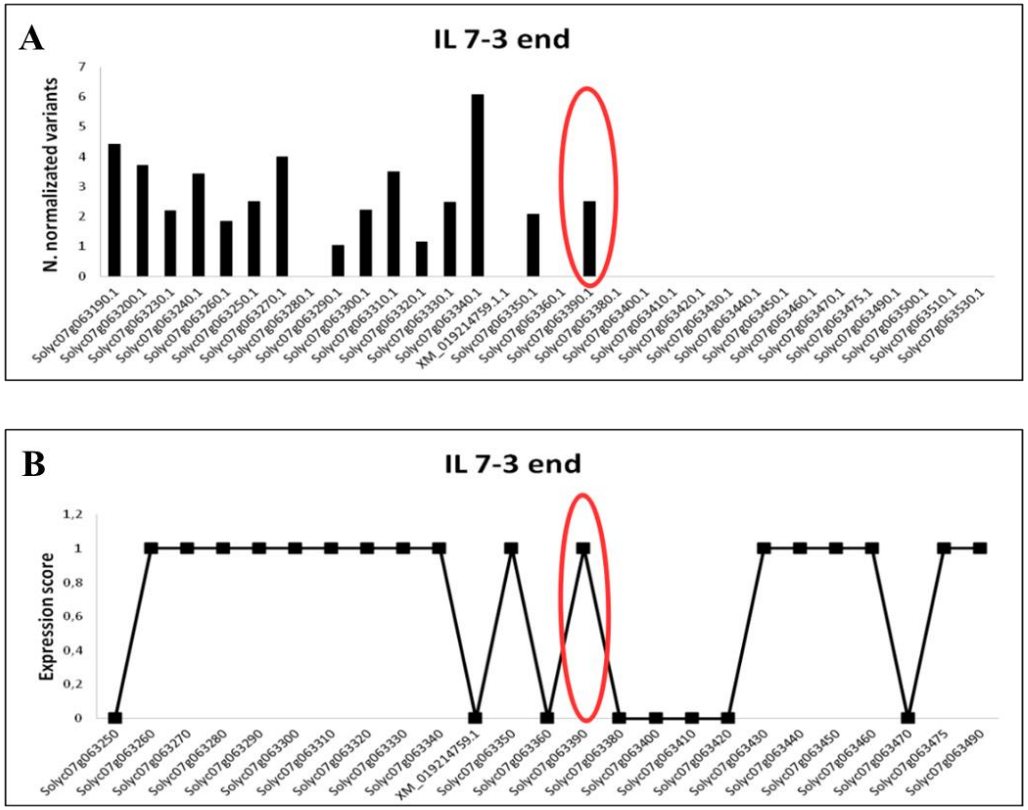


Figure 3.10- Borders reliability check of the introgression fragment to identify the exact lower introgression border.

The red cycle shows the last gene removed from M82. A) Number of normalized variations for each gene B) Expression score for each gene. The value 1 indicates that the genes were expressed whereas the value 0 genes that they were not expressed, considering a threshold expression value of 0.05.

3.2.4 R182 introgression size definition

To confirm the reliability of the IL7-3 assembly and define the size of *S. pennellii* introgression region in the elite sub-line R182, whose first gene overlaps the start break point of IL7-3, 15 SCAR and three CAPS (N34, N35 and N18) were screened (Table 3.4). All markers were constructed based on polymorphisms found comparing the genomes of the cultivated and the wild species and were tested on the parental genotypes M82 and IL7-3 together with R182. As for the CAPS marker N34, the restriction enzyme used was *Rsa*I with M82 and IL7-3 digested fragment size of 415 bp and 217+198 bp respectively, whereas N35 was based on the restriction enzyme *RsaI*, which showed M82 and IL7-3 digested fragment sizes of 476+54 bp and 530 bp, respectively. As for the last CAPS marker N18, the restriction enzyme used was *ScaI* with M82 and IL7-3 digested fragment sizes of 223+217 bp and 440 bp, respectively. The analysis carried out using molecular marker allowed identifying a more precise starting point of the IL7-3 and R182 introgression fragments (Table 3.4), as the first introgression wild gene verified was the Solyc047990 instead of the Solyc07g048010 reported by elaborating the RNA-seq data. The cultivated region covered by the designed molecular markers spans from 59,320,656 bp to 59,705,558 bp of chromosome 7 (version ITAG3.0 of the Sol Genomics tomato annotated genome), whereas the corresponding *S. pennellii* region spans from 69,831,629 bp to 70,130,606 bp. This analysis allowed determining whether the wild or the cultivated allele is present for each investigated gene in the introgressed region of the sub-line R182, thus precisely defining its borders. As reported in the table, the region spanning from marker MK2 to marker MK21 carries wild alleles in the sub-line R182, since R182 showed the same amplified or digested DNA fragments observed in the parental line IL7-3. The last wild gene of the introgressed region is Sopen07g024640, for which two different CAPS markers were designed (MK21 and MK22), targeting different regions of the gene. Since the genotype R182 showed the wild allele for marker MK21 and the cultivated one for marker MK22, we hypothesized that the recombination event leading to the definition of the introgression region in this line occurred in within this gene.

Table 3.4- Molecular marker designed to identify the size of the R182 introgression region and results obtained for each gene.

Marker code	Marker type	Primer sequence	Amplified fragment size M82	Amplified fragment in IL7-3	S. *lycopersicum* region targeted	Chr. 7 start position SL3.0--	S. *pennellii* region targeted	Chr. 7 start position S.*pennellii*	M82	IL7-3	R182
MK1	SCAR	F- 5'-3' AAGAGAGCGAAGGTAGTAAC R- 5'-3' TCAGTCAAGGCTACTTCCC	134	-	Solyc07g047980	59320014	Sopen07g024410	6981629	-	-	-
MK2	SCAR	F- 5'-3' TATCGACATGCCTTATGTTG R- 5'-3' TCAGTGATGAAATTGAGTGTG	135	-	Solyc07g047990	59325229	Sopen07g024420	69836399	-	+	+
MK3	SCAR	F- 5'-3' TCATAGATGATTCACACTGAG R- 5'-3' ACAAGTTCACAGGATATTCAC	-	246	Solyc07g047990	59325229	Sopen07g024420	63836805	-	+	+
MK4	SCAR	F- 5'-3' TTAAAAGAGTCGTGCGGAGC R- 5'-3' GAAGTGAACAGCCAATGTGG	323	-	Solyc07g048000	59330722	Sopen07g024430	69840798	-	+	+
MK5	SCAR	F- 5'-3' GGCGCCAAATAAAAAATGG R- 5'-3' CTCTTCTCCCAACGATTTG	-	1146	Solyc07g048010	59335411	Sopen07g024440	69849442	-	+	+
N34	CAPS	F- 5'-3' CTCAATCATTCTTGGCCAACTA R- 5'-3' CTTTCCCCGACAAAGTGCTC	415	415	Solyc07g048010	59338020	Sopen07g024440	69848612	-	+	+
MK6	SCAR	F- 5'-3' AAGCCGTGTAGCTCGGAC R- 5'-3'CACTGAAGTGTACATAGAGAGG	623	-	Solyc07g048030	59347798	Sopen07g024450	69859998	-	+	+
MK8	SCAR	F- 5'-3' CCATTTCTCCTCTCAGTAG R- 5'-3' AGCAACTTGTTCCGGTTAG	122	-	Solyc07g048080	59394940	Sopen07g024500	69913622	-	+	+
MK9	SCAR	F- 5'-3' GCTATACCGCAATCATAAATC R- 5'-3' GAGGCCACTGCATTAGGTT	-	1150	Solyc07g048100	59477649	Sopen07g024530	69961500	-	+	+
MK11	SCAR	F- 5'-3' CTGATTGCATCGGAATTACCC R- 5'-3' CGCGAAAGGAACAAAAATACG	348	-	Solyc07g049140	59494927	Sopen07g024560	69998524	-	+	+
MK12	SCAR	F- 5'-3' GTTGATATGTCTGGAATTTCG R- 5'-3 GTGGAATAATTATTTCACTCATG	429	446	Solyc07g049140	59494979	Sopen07g024560	69998568	-	+	+
MK13	SCAR	F- 5'-3' TAGGTATCAAGTCGCTGAAC R- 5'-3' ATCTGTAAATTTTGTGCTCATC	816	-	Solyc07g049170	59554063	Sopen07g025160	70804301	-	+	+
MK14	SCAR	F- 5'-3' TCTCCTTCCAATTCCACTTC R- 5'-3' CAAGGCGGAACCACCTAAG	1539	-	Solyc07g049200	59621820	Sopen07g025210	70873975	-	+	+
N35	CAPS	F- 5'-3' CTTGAAAGCATTGCCAGAGAA R- 5'-3' GCACACTCTATGGATTCAAGGG	530	530	Solyc07g049280	59667971	Sopen07g024610	70081676	-	+	+
MK20	SCAR	F- 5'-3' ACTTGTTGTTGCTTACATATC R- 5'-3' AACTTCAGTGATGCGACTAC	247	-	Solyc07g049290	59670170	Sopen07g024620	70082413	-	+	+
MK21	SCAR	F- 5'-3'CTTCAATGCTTCTCAACGAC R- 5'-3' CAGAAACCATACCTGCATG	-	615	Solyc07g049310	59686924	Sopen07g024640	70111820	-	+	+
MK22	SCAR	F- 5'-3' CTAGAACTCGTTCAGAAATTTG R- 5'-3' GCATAGAGTAGTGCTATTTTC	270	-	Solyc07g049310	59695121	Sopen07g024640	70120070	-	+	-
N18	CAPS	F- 5'-3' GCCATTTAACATTGGGACTCG R- 5'-3' AGCTTACATCTGATCCGCCC	440	440	Solyc07g049320	59705558	Sopen07g024650	70130606	-	+	-

3.2.5 RNA-seq B27vsM82

During the years 2016 and 2017, in order to define the transcriptome profiling and differentially expressed genes (DEG), the sub-line B27 and the parental genotype M82 were analysed using an RNA-seq approach. For each genotype two replicates of red ripe fruits cultivated in the Acerra fields were considered. The quality check (QC) performed on the raw sequencing data to remove low quality portions and Illumina adapters (Supplemental Figure S3.4, S3.5) showed reads reduction after the QC (Table 3.5). As for the year 2016, M82 reads exhibited a total reduction of approximately 0.14%, indeed M82_A, M82_B showed a reduction of 0.15% reads and 0.13%, whereas B27 showed an average reduction of 0.13% reads with B27_B and B27_C. As for 2017, M82_A and M82_B exhibited a total reduction of 15.5% with values of 14, 6% (and 16,30%, whereas B27_B and B27_C showed a total reduction of 15,25%, with values of 14% and 16,40% respectively. The high-quality reads aligned against the *S. lycopersicum* reference genome (SL3.20 release) with STAR aligner, showed that an average of reads of approximately 96% for the year 2016 and 94% for the year 2017 were uniquely mapped on the genome and an average mismatch rate *per* base of 0.6% and 0.13% was observed, respectively

Table 3.5- Number of reads of B27 and M82 before and after the quality check (QC).

	2016		2017	
Sample	**Reads before QC**	**Reads after QC**	**Reads before QC**	**Reads after QC**
M82_A	49194352	49118966	38678218	33732410
M82_B	48288299	48266214	30200906	25967758
B27_B	53960561	53888025	40955694	35896764
B27_C	52880116	52812189	42909710	36862446

Data related to the number of reads of M82 and B27 replicates for the year 20176 and 2017 are reported.

The R HTSFilter package was applied for the statistical analysis aimed at removing the not expressed genes and the ones showing a high variability and used a filtering procedure for replicated transcriptome sequencing data based on a Jaccard similarity index. This showed that for the year 2016 (Figure 3.11) the replicates had the highest similarity with a TMM normalized read counts (s) of 36.71, whereas for the year 2017 the TMM normalized read counts (s) was 77.65 (Supplemental Figure S3.6). These values were used as threshold and loci with TMM normalized read counts (S) < of 36.71 and 77.65 respectively were removed. As for the first RNA-seq (2016), 17,647 were filtered and 17,078 were removed (34,725 total genes), whereas for the second RNA-seq (2017) 13,185 were filtered and 21,576 were removed (34,761 total genes). The overall quality of the experiment was evaluated, based on the similarity between replicates, by a PCA analysis using the normalized expression values of the genes filtered with HTSFilter. The genes passing the HTSFilter were used for differential expression analysis (Table 3.6). The comparison among genotypes B27 and M82 for the first RNA-seq, showed 1,387 DEGs, of which 785 up-regulated and 602 down-regulated, whereas the second RNA-seq evidenced 63 DEGs, of which 37 up-regulated and 26 down-regulated. Genes were considered significantly differentially expressed if the false discovery rate (FDR) of the statistical test was less than 0.05 and the logarithm of the fold-change was greater than 2. As for the two RNA-seq experiments (Figure 3.12), in the B27 introgression region 20 up-regulated and 10 down-regulated and 8 up-regulated and 8 down-regulated genes were identified in the years 2016 and 2017, respectively. Particularly, only 8 up-regulated and 4 down-regulated genes were in common between the two analyses. As for the year 2016, among the differential expressed genes the Solyc12g098480 encoding for the UDP-glucoronic-acid-4-epimerase and only the Solyc12g098380 for the year 2016 and 2017 were identified as genes involved in the pathway of the AsA biosynthesis. The Solyc12g098380 was previously studied as key gene for the increment of AsA content, whereas the Solyc12g098480 was selected as candidate gene.

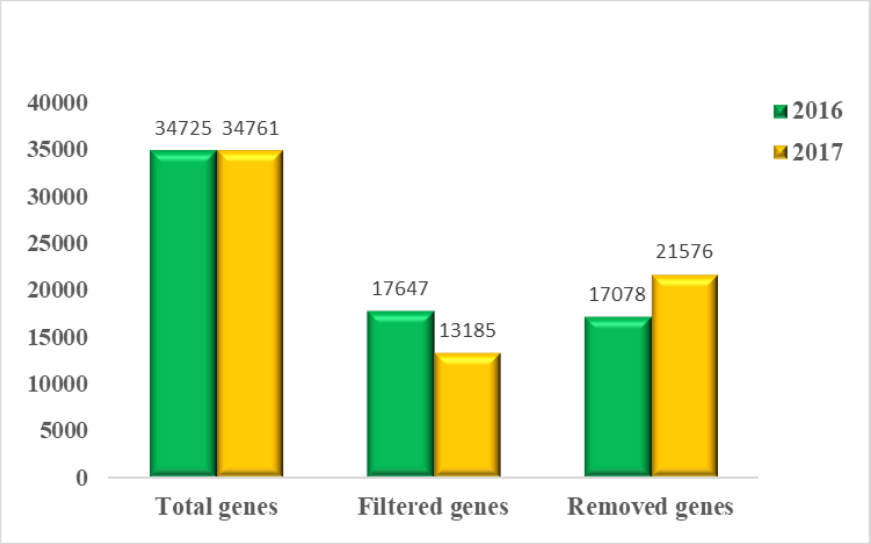


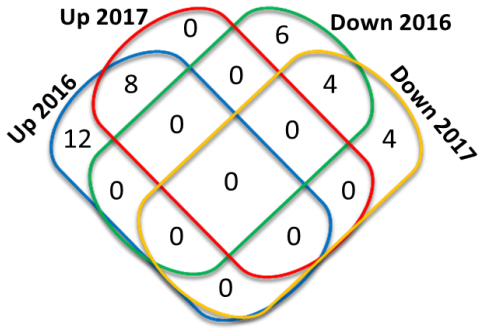
Figure 3.11- Bar plot of total, filtered and removed genes following the HTSFilter analysis of two RNA-seq experiments carried out in the years 2016 and 2017.

Table 3.6- Number of DEGs for the year 2016 and 2017.

	2016	**2017**
Tot. DE genes	1387	63
UP genes	785	37
Down Genes	602	26

Data related to the total, up-regulated and down-regulated DEGs of M82 and B27 for the year 2016 and 2017.

A



B

UP regulated genes		Down regulated genes	
2016	2017	2016	2017
Gene ID	Gene ID	Gene ID	Gene ID
Solyc12g098340	*Solyc12g098340*	Solyc12g098510	Solyc12g098580
Solyc12g098480	*Solyc12g098660*	*Solyc12g098620*	*Solyc12g098620*
Solyc12g098610	*Solyc12g099000*	Solyc12g098710	Solyc12g098640
Solyc12g098660	*Solyc12g099190*	Solyc12g098810	Solyc12g098700
Solyc12g098740	*Solyc12g099515*	*Solyc12g099200*	*Solyc12g099200*
Solyc12g098840	*Solyc12g099930*	Solyc12g099430	Solyc12g099250
Solyc12g098900	*Solyc12g100010*	*Solyc12g099600*	*Solyc12g099600*
Solyc12g098910	*Solyc12g100180*	Solyc12g100250	*Solyc12g100260*
Solyc12g099000		*Solyc12g100260*	
Solyc12g099180		Solyc12g100270	
Solyc12g099190			
Solyc12g099450			
Solyc12g099515			
Solyc12g099780			
Solyc12g099870			
Solyc12g099930			
Solyc12g100010			
Solyc12g100030			
Solyc12g100050			
Solyc12g100180			

Figure 3.12- Up-regulated and down regulated genes evidenced from the RNA-seq experiments carried out in the years 2016 and 2017 A) Venn diagram of DE genes in the two years. B) ID of common DEGs between the two years are evidenced in italics.

3.2.6 RNA-seq R182*vs*M82

During the year 2017 two RNA-seq analyses were carried out between the sub-line R182 and the parental genotype M82, comparing fruits collected in the experimental field of Acerra at breaking and mature red stages. For each sample, two replicates were considered. The quality check (QC), performed on the raw sequencing data to remove low quality portions and Illumina adapters, showed reads reduction after the QC (Table 3.7). As for the breaker RNA-seq, M82 reads exhibited a total reduction of 15%, indeed M82_A, M82_B showed reads reduction of 14.43% and 15.5%, whereas R182 showed a total reduction of approximately reduction of the 14.5% with R182_A 14.73%, and R182_B 14.24%. As for the mature red RNA-seq, M82 and R182 exhibited a total reduction of 15.5% and 15% respectively. As for M82 replicates, M82_A showed reads reduction of 14.6%, whereas M82_B 16.30%. As for R182 replicates, R182_A and R182_B exhibited a reduction of 16.20% and 14%. The high-quality reads aligned against the *S. lycopersicum* reference genome (SL3.20 release) with STAR aligner showed an average of approximately 94% for the first RNA-seq and 95% for the second RNA-seq were uniquely mapped on the genome and an average mismatch rate *per* base of 0.14% and 0.13% was observed, respectively.

Table 3.7- Number of reads of R182 and M82 before and after the quality check (QC).

	BR		MR	
Sample	**Reads before QC**	**Reads after QC**	**Reads before QC**	**Reads after QC**
M82_A	31807464	27795714	38678218	33732410
M82_B	33222414	28773386	30200906	25967758
R182_A	37086080	32324302	35189096	30285052
R128_B	42663592	37343756	30979878	27188354

Data related to the number of reads of M82 and R182 replicates for the Breaker RNA-seq and Mature Red RNA-seq.

The R HTSFilter package was applied for the statistical analysis aimed at removing the not expressed genes and the ones showing a high variability and used a filtering procedure for replicated transcriptome sequencing data based on a Jaccard similarity index. This showed that for the Breaker RNA-seq the replicates had the highest similarity with a TMM normalized read counts (s) of 5.11, whereas for the Mature Red RNA-seq the TMM normalized read counts (s) was 13.63 (Supplemental Figure S3.7). These values were used as threshold and loci with TMM normalized read counts < of 5.11 and 16.63 respectively were removed. As for the first RNA-seq (BR) (Figure 3.13), 19,592 were filtered and 15,169 were removed (34,761 total genes), whereas for the second RNA-seq (MR) 16,649 were filtered and 18,112 were removed (34,761 total genes). The overall quality of the experiment was evaluated, based on the similarity between replicates, by a PCA analysis using the normalized expression values of the genes filtered with HTSFilter. The genes passing the HTSFilter were used for differential expression analysis (Table 3.8).

The comparison among the genotypes R182 and M82 for breaker RNA-seq, showed 50 DEGs of which 32 up-regulated and 18 down-regulated, whereas the MR RNA-seq evidenced 98 DE genes of which 39 up-regulated and 59 DE down-regulated. Genes are considered significantly differentially expressed if the false discovery rate (FDR) of the statistical test was less than 0.05 and no filter was applied to the logarithm of the fold-change. As for the two RNA-seq experiments (Figure 3.14), in the R182 introgression region 2 up-regulated and 3 down-regulated and 3 up-regulated and 3 down-regulated genes were identified in the BR and MR RNA-seq, respectively.

Particularly, only 7 up-regulated and 4 down-regulated genes were in common between the two analyses. As for the first analysis, among the differential expressed genes the Solyc07go49310 and the Solyc07g049290 corresponding to two Major facilitator superfamily (MFS) proteins. These two genes were selected as candidate genes for the increment of Ascorbic Acid content in the sub-line R182.

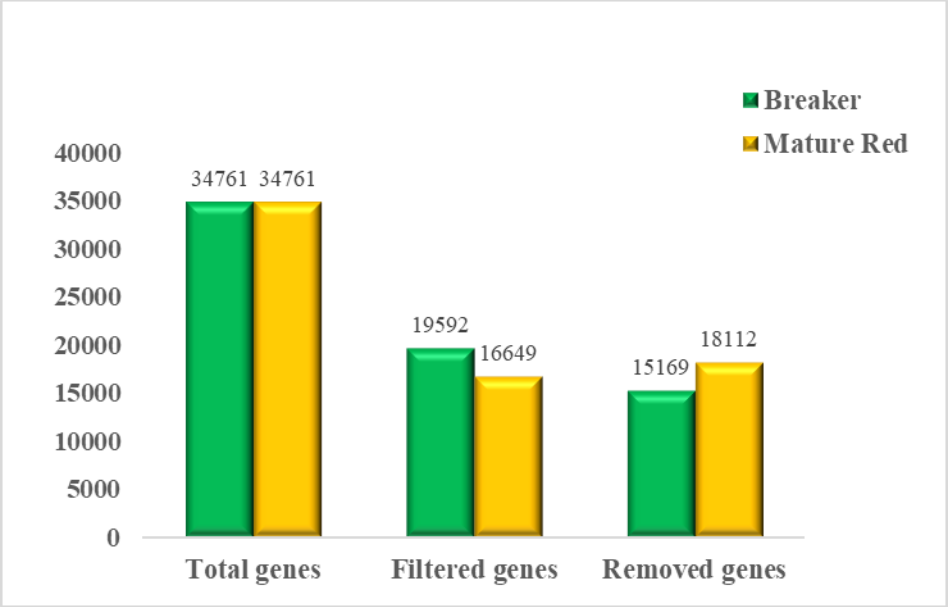


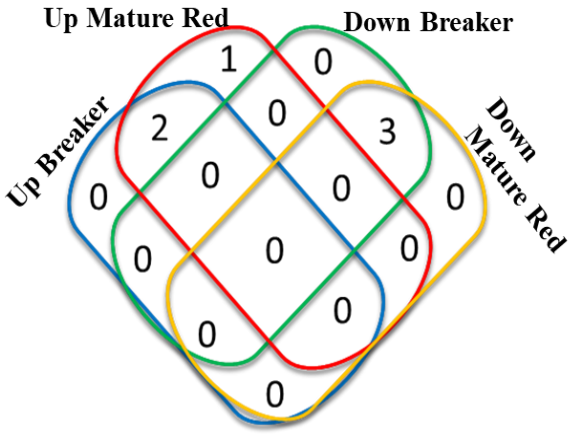
Figure 3.13- Bar plot of total, filtered and removed genes following the HTSFilter analysis of two RNA-seq experiments carried out in the years 2017.

Table 3.8- Number of DEGs for the year 2017.

	Breaker	Mature Red
Tot. DE genes	50	98
UP genes	32	39
Down Genes	18	59

Data related to the total, up-regulated and down-regulated DEGs of M82 and R182 for the year 2017.

A



B

UP regulated genes		Down regulated genes	
Breaker	**Mature Red**	**Breaker**	**Mature Red**
Gene ID	**Gene ID**	**Gene ID**	**Gene ID**
Solyc07g048040	*Solyc07g048040*	*Solyc07g048100*	*Solyc07g048100*
Solyc07g049310	Solyc07g049290	*Solyc07g049140*	*Solyc07g049140*
	Solyc07g049310	*Solyc07g049200*	*Solyc07g049200*

Figure 3.14- Up-regulated and down regulated genes for the two RNA-seq analyses for the year 2017. A) Venn diagram of DE genes in the two analyses. B) ID of common DEGs between the two years are evidenced in italics.

Chapter 4. Strategies to study the function of genes controlling AsA content in tomato fruit

4.1 Materials and methods

4.1.1 Plant materials

Various tomato genotypes were used with the purpose of studying the function of genes controlling AsA content in tomato fruit:

- the cultivated genotype M82
- the cultivated genotype Money Maker
- the introgression sub-line B27
- two genotypes called GGP1 and GGP2 provided by Cathie Martin.

All genotypes were grown in a greenhouse at the John Innes Centre. B27 and M82 fruits were collected at Mature Green (MG) development stage, whereas GGP fruits at Mature Red (MR) development stage, seeds and columella were subsequently removed, and fruits were ground in liquid nitrogen and stored at −80°C until the analyses. Leaves were also collected from each genotype, ground in liquid nitrogen by mortar and pestle to a fine powder and stored at −80°C until the analyses.

4.1.2 RNA extraction and cDNA synthesis

RNA was extracted from B27 MG fruits using the TRI REAGENT (Sigma-Aldrich) and following the manufacturer's guidelines with small modifications. First, 1mL of TRI REAGENT was used for 100 mg of frozen powder, afterwards the homogenate was vortexed, stored for 5 min on ice and centrifugated at 14000 rpm for 10 min at 4°C. The supernatant was transferred in a new Eppendorf tube and stored for 5 min on ice. Afterward, 200 µL chloroform were added in the Eppendorf, for 5 min stored on ice and centrifugated at 14000 rpm for 15 min at 4°C. The aqueous phase was transferred and 500 µL of isopropanol were added. Samples were stored for 10 min on ice and centrifugated at 14000 rpm for 10 min at 4°C.

The supernatant was removed, and RNA pellet was washed by vortexing with 1 mL of 75% ethanol and centrifugated at 14000 rpm for 5 min at 4°C. This step was repeated for five times. The ethanol wash was removed and briefly the RNA pellet was air-dried for 10 min. RNA was dissolved in 40 µL of H_2O free of DNA. Samples were stored at -80°C. One µg of total RNA was reverse transcribed to cDNA using oligo-dT (20) primers and SuperScript III reverse transcriptase (Invitrogen) in 20 µL of final reaction according to the manufacturer's instructions

4.1.3 DNA isolation

DNA was isolated from M82 and GGP fruits using the DNeasy plant mini kit (Qiagen) according to the manufacturer's protocol. Plant tissue was ground using a mortar and pestle and about 100 mg of powder was added to 400 µL of buffer AP1 (10 mM Tris-HCl pH8.0, 1 mM EDTA pH 8.0, 0.1% SDS, 0.1M NaCl, 1X PVP, 10mM DTT, RNase) and incubated at 65°C for 10 minutes. 130 µL of buffer P3 was then added and the mixture was incubated on ice for 5 minutes. The mixture was then centrifuged in the shredder column for 2 minutes at 20000 x g. The supernatant was transferred to a new tube and 1.5 volumes of AW1 was added. The mixture was centrifuged through the DNA binding membrane for 1 minute at 6000 x g and washed twice with 500 µL of buffer AW2. The DNA was eluted from the column through the addition of 100 µL buffer AE, (10 mM Tris-HCl, 0.5 mM EDTA, pH 9.0) and centrifuging at 6000 x g for 1 minute.

4.1.4 Primer design

Primers were designed using the online software Primer3 plus, (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) and Oligo Evaluator (Sigma Aldrich) (http://www.oligoevaluator.com) according to the following output parameters: the optimal length of oligo primers was from 22 to 30 bp (excluding primers designed for Gateway™ cloning). The melting temperatures (Tm) of primers were restricted to 58 – 65°C with a GC content from 35 – 65% with complementary sequences avoided. A list of primers used can be found in Supplemental Table S4.1

4.1.5 Polymerase Chain Reaction

Amplification condition was performed in 200 µL microcentrifuge tubes. 50 µL reaction volume consisting of 10 µL 5x Phusion HF Buffer, 2.5 µL forward primer (10 uM), 2.5 µL reverse primer (10 uM), 1 µL DNTPs (10 mM), 0.5 µL Thermo Scientific™ Phusion™ Flash High-Fidelity DNA Polymerase (2 U/µL), 50 ng\µL of template and nuclease free water to a final volume of 50 µL. The mixture was amplified in a DNA Thermal Cycler (G-Storm). Thermal cycle was programmed for 2 min at 95°C as initial denaturation, followed by 30-40 cycles of 30 sec at 95°C for denaturation, 30 sec from 60°C to 65°C as annealing (temperature depended on the primer), from 30 sec to 2 min at 72°C for extension (time depended on the amplified fragment size), and final extension at 72°C for 10 min. PCR products were analysed on 2% (w/v) agarose gel and amplicons obtained were gel-purified with QIAquick Gel Extraction Kit (Qiagen).

4.1.6 *Escherichia coli* DH5α Calcium Chloride competent cells

A single *E. coli* colony was inoculated into 10 mL LB broth in 50 ml falcon tube and grown by shaking overnight at 37°C. Then, 3 mL were inoculated in 400 mL LB broth and shaken at 37°C until the OD_{550} was approximately 0.5. After that the culture was transferred to 200 mL flasks and centrifuged at 2500 rpm for 10 min. Each pellet was resuspended in one third original volume (66 ml) of TBD [$RbCl_2$ 100mM, $MnCl_2.4H_2O$ 45 mM, $CaCl_2(+2H_2O)$ 10mM, KAcetate 35 mM and Glycerol 15%]. Resuspended bacteria were incubated for 90 min on ice and spinned at 2000 rpm for 5 min. The supernatant was poured off and the pellet resuspended in 1/25 of the original volume, in freezing buffer [$RbCl_2$ 10mM, $CaCl_2(+2H_2O)$ 75mM, MOPS 10 mM and Glycerol 15% pH 6.8]. After resuspending the pellet, the competent cells were aliquoted into cryostat tubes, immediately frozen in liquid nitrogen and stored at -80°C.

4.1.7 Overexpression Gateway cloning

PCR products of Solyc12g098480 and Sopen12g033260 with attBI and attB2 sites at each end were cloned into the pDONR 207 entry vector using BP clonaseII (Invitrogen). pDONR 207 (150 ng) was mixed with at least 50 ng DNA fragment and 2 µL of BP clonaseTM enzyme in a total volume of 10 µL and incubated overnight at 25°C. Proteinase K (1 µL) was added and incubated for 10 minutes at 37°C, to stop the reaction. The whole mixture was diluted to a final volume of 80 µL with sterile water and added to competent DH5α *E. coli* cells for transformation. *E. coli* were grown on plates, then in liquid culture, using 10 µg/ml gentamycin as selection antibiotic and the plasmid DNA was extracted using the QIAprep Miniprep Kit (Qiagen) according to manufacturer's instructions. Inserts were confirmed using sequencing (Mix2Seq Eurofins Genomic service).

LR clonaseII (Invitrogen) was used to insert the DNA of interest into destination vectors, 100 ng of pDONR 207 vector and 100 ng of destination vector pJAM2288, obtained in Cathie Martin's laboratory, were mixed with 2 µL of LR clonaseTM in a total volume of 10 µL. Reactions were stored at 25°C overnight and then stopped by the addition of 1 µL of proteinase K solution and incubated at 37°C for 10 minutes. The mixture was added to competent DH5α for heat shock transformation. *E. coli* were grown on plates, then in liquid culture, using 50 µg/ml kanamycin as selection antibiotic and the plasmid DNA was extracted using the QIAprep Miniprep Kit (Qiagen) according to manufacturer's. Inserts were confirmed using sequencing.

4.1.8 Agrobacterium transformation

Agrobacterium tumefaciens transformation was performed using the freeze-thaw method. Agrobacterium strain AGL1 was grown in 5 ml LB broth containing Rifampicin (50µg/mL) overnight at 28°C. 5 mL of overnight culture were added to 100 mL LB media and shaked at 28°C until culture was grown to an $O.D._{600}$. Then the culture was chilled for 10 min on ice, spinned down at 4000 rpm for 10 min at 4°C in pre-chilled 50 ml Falcon tubes. The supernatant was discarded, the cells were resuspended in 1mL of ice-cold LB media and dispensed 300 µL aliquots into pre-chilled 2 mL microtubes. One µg of plasmid DNA in a 20 µL volume was added to the cells, gently mixed by flicking the tube.

Then the cells were frozen in liquid nitrogen for 15 sec and immediately thawed by incubation at 37°C for 5 min. 700 µL of LB medium were added to the tube and shaken for 2-3 h at 37°C. The tube was centrifuged at 4000 rpm for 5 min in a microfuge tube. 600 µL of supernatant were discarded, the cells were resuspended in the remaining 400 µL of supernatant and spread on LB plate containing Kanamycin (50 µg/ml). Following overnight growth at 28°C with shaking at 280 rpm DNA was extracted to verify the presence of the gene of interest.

4.1.9 Stable and transient overexpression

Stable tomato overexpression of Solyc12g098480 and Sopen12g033260 was carried out using M82 seeds. Tomato seeds were treated with 70% EtOH to loosen gelatinous seed coat. After 2 min EtOH vas removed and seeds were rinsed once with sterile water. 10% NaClO was added, for 3 h shaken, and seeds were washed 4 times with water. 100 seeds were put into tubs containing germination medium (Murashige and Skoog basal salt mixture (MS) + vitamins+ agarose (pH=5.8), left at 4°C for 3 days and after that stored in the culture room. Seedlings were grown for 7-10 days. After 10 days, cotyledons were cut with a rolling action of a rounded scalpel blade to minimize damage to the tissue. In a Petri dish the tip of the cotyledon was cut off and then two more transverse cuts were made to give two explants of about 0.5 cm long. The explants were transferred to a new Petri dish of water to prevent any damage during further cutting. *Agrobacterium* culture was spinned down and the pellet was resuspended in Murashige and Skoog basal salt mixture (MS) + vitamins medium 3% sucrose (pH=5.8) to an OD_{600} of 0.4-0.5. Bacterial suspension was put in a petri dish and the explants were immersed for 30 min. Then them were removed, blotted on sterile filter paper and placed about 30-40 on a feeder plate, again taking care not to damage the tissue. The plates were returned to the culture room with 1ow light for 48 h. The pieces from the feeder layers were taken and put onto tomato regeneration plates [MS+vitamins medium 2% sucrose + 0.1% Myo-inositol + Agargel (pH=6.0)] containing Timentin (also known as Ticarcillin) at 320mg/L, Zeatin Riboside at 2mg/L and Kanamycin at 50ug/ml. 12-16 pieces were placed *per* Petri dish. Plates were left unsealed and returned to the culture room for ~2weeks.

Explants were transferred to fresh medium every 2-3 weeks. Once regenerating material is too large for Petri dishes it was put into a deep Petri dish, and then into tubs for 4-6 weeks. The regeneration media [MS+vitamins medium 2% sucrose + 0.1% Myo inositol + Agargel (pH=6.0)] in tubs was contained in addition to the Timentin, 250mg/l Cefotaxime. Shoots were cut from the explants and put into rooting medium [MS+vitamins medium 0.5% sucrose + Gerlite(pH=6.0)] with Timentin at 320mg/L and Kanamycin at 50mg/L. To transfer to soil, as much of the medium as possible was removed by washing the roots gently under running water. Plant carefully in hydrated, autoclaved Jiffy pots (peat pots) and kept enclosed to keep humidity high while in the growth room. Gradually humidity was decreased.

Agroinfiltration for transient overexpression of Sopen12g033260 and Solyc12g098480 in Money Maker fruit was performed as described by Orzaez et al. (2006) with some modifications. 5 ml LB culture with Agrobacterium previously transformed were grown overnight at 28°C, 200 rpm. The next day, 50 µL of the culture were inoculated into 50 ml LB containing Kanamycin (50 µg/ml) and Rifampicin (50 µg/ml), grown overnight at 28°C to a final OD_{600} of 0.5-0.8. The culture was spinned down for 15 min at 3500 rpm. The supernatant was removed completely and cells were resuspended in infiltration medium (10 mM MgCl2, 10 mM MES, pH5.7) to an OD_{600} of 0.5. Then, acetosyringone was added to obtain a final concentration of 200 µM. The tubes were wrapped with aluminium and on a shaker incubated for 2-3 h. Tomato Money Maker fruits at Breaker development stage were infiltrated using a 1-mL syringe with a 0.5-×16-mm needle. Needle was introduced 3 to 4 mm in depth into the fruit tissue through the stylar apex, and the infiltration solution was gently injected into the fruit. Once the entire fruit surface has been infiltrated, some drops of infiltration solution begin to show running off the hydathodes at the tip of the sepals. Agroinfiltrated fruit showing overexpressed epimerase were collected two weeks after breaker stage.

4.1.10 Genotyping of GGP plants

GGP1 and GGP2 genotyping was performed by PCR and sequencing approach considering three fruits at MR development stage. The sequencing of the Solyc02g091510 and the Solyc06g073320, encoding for the GDP-L-galactosephosphorylase1 (GGP1) and GDP-L-galactosephosphorylase2 (GGP2), was carried out using the Mix2Seq kit (Eurofins) according to the manufactures protocol.

4.1.11 AsA determination

AsA determination was performed using Ion-Pair HPLC (Gradient) with UV detection on the Agilent Infinity method (Hu et al., 2012). 1mL of 5% ice-cold meta phosphoric acid was used for 250 mg FW grinded tissue, afterwards the mixture was vortexed, for 30 min. incubated on ice and centrifuged at 14000 rpm for 20 min at 4°C. The supernatant was filtered through a C18 column and a 0.22 µm filter. 10 µL of supernatant were incubated with 10 µL of 10mmol/l TCEP for total ascorbate quantification or with DDW for oxidised ascorbate analysis, for 30 minutes at room temperature, in dark. Each sample was diluted 1:10 by adding 180µL MilliQ water. Samples were kept in dark at 4ºC until injected (1 µL) or stored in -20 ºC until analysed. Data (245 nm) were processed using ChemStation on the Agilent Infinity.

4.1.12 Bioinformatics and statistical analysis

Amino acid and nucleotide sequences were collected from the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov) and Sol Genomics (https://solgenomics.net) database. Alignments and phylogenetic trees were performed with Geneious software v11.1.5 (Biomatters, http://www.geneious.com). The blast tool of Sol Genomics (https://solgenomics.net/tools/blast), using these options: expect (e-value) threshold $1e^{-10}$, max. hits to show 20, substitution matrix BLOSUM62 and output options basic, and blastn program (http://www.ncbi.nlm.nih.gov/blast) were used to perform homology researches of Solyc12g098480 in S. *pennellii* genome.

Data analysis was carried out using SPSS Software version 23. The t-Student test was calculated for the AsA analysis to verify if genotypes were statistically different from the wild type Money Maker.

4.2 Results

4.2.1 Study of an epimerase gene

The UDP-glucoronic-acid-4-epimerase (UGlcAE) Solyc12g09848, identified as candidate gene involved in the D-Glucoronate pathway, was studied as one gene present in the B27 introgression region that could contribute to the increase of AsA (Rigano et al., 2018). Indeed, the previous study of the UGlcAE expression in M82 and B27 Mature Green (MG), Breaker (BR) and Mature Red (MR) fruits carried out using real-time q-PCR, showed the expression of this gene only at MG and BR stages in fruits of the sub-line B27, while it was not expressed in M82 fruits during ripening (Rigano et al., 2018). Furthermore, the results of the RNA-seq analysis carried out during the year 2016, showed that Solyc12g098480 was overexpressed in the sub-line B27 respect to the cultivated genotype M82, therefore confirming this gene as candidate gene for the increase of AsA content. In the present thesis, using the Blast tool of Sol Genomics, the gene structure of the UGlcAE was investigated (Supplemental figure S4.1). Indeed, the analysis of the gene Solyc12g098480 showed an intronless fragment of 192 bp. The gene coded for a protein of 63 amino acids. Moreover, the use of the Blast function on the Sol Genomics platform to find homologous genes to the UGlcAE in the wild species *S. pennellii* (Figure 4.1a, Supplemental figure S4.2, S4.3), led to identify two identical intronless UDP-glucoronic-acid-4-epimerases on chromosome 12 of *S. pennellii* (Sopen12g033320 and Sopen12go33260), with a length of 966 bp and an identity of 3.5% (34 bp) with the Solyc12g098480. The two *S. pennellii* genes were discriminated from the cultivated genotype using molecular markers as described in Figure 4.1b.

A

Solyc12g098480.1.1 vs S. pennellii CDS

SubjectId	id%	Aln	evalue	Score	Description
Sopen12g033320	100.00	34/34	5e-11	67.9	NAD dependent epimerase/dehydratase family Length = 996
Sopen12g033260	100.00	34/34	5e-11	67.9	NAD dependent epimerase/dehydratase family Length = 996

B

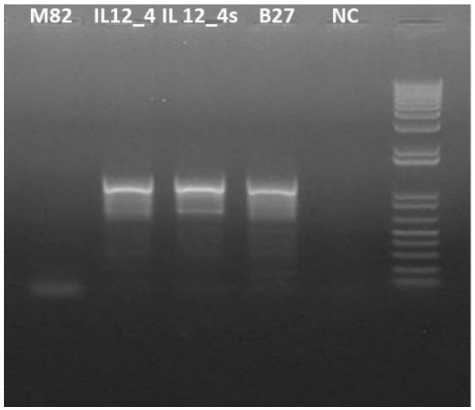


Figure 4.1-PCR of the Epimerase genes on the chromosome 12 of *S. pennellii.*

A) Blast result of the coding sequence (CDS) of Solyc12g098480 respect to the CDS of *S. pennellii*. B) Agarose gel of the PCR amplification used to discriminate the two *S. pennellii* genes.

Using the Sol Genomics website, additional nine *S. pennellii* genes were found that were homologous to *S. lycopersicum* intronless epimerases (Table 4.1) and are expressed in different tissues of the plant (Tomato functional genomic database, http://ted.bti.cornell.edu/). From the alignment (Supplemental figure S4.4) all *S. pennellii* genes showed approximately 70% of identity each other and the phylogenetic analysis carried out using amino acid sequences (Figure 4.2) exhibited differences among the genes. All proteins showed the amino acid sequence motif GxxGxxG involved in binding of the cofactor NAD(P) and the YxxxK motif that is part of the catalytic domain of short-chain dehydrogenases/reductases (accession number cl25409). The alignment among all the epimerases proteins of the *S. pennellii* genome (Figure 4.3), showed that the Sopen12g033320 and the Sopen12g033260 lost the motif GxxGxxG involved in binding of the cofactor NAD(P). Epimerases lacking this domain could be not active, however the BlastN analysis carried out using NCBI website and publicly available IL12-4 RNA-seq data allowed to identify the presence of an intron of 146 bp in the UGlcAE genomic sequence. Particularly, comparing the Sol Genomics and the NCBI tomato annotations, we found two identical epimerases (XM_015204087 and XM_015204086) on chromosome 12 of the wild species *S. pennellii*. Analysing the gene sequences alignment (Supplemental Figure S4.5) and their amino acid sequences (Figure 4.4), we found both the motifs of the dehydrogenase/reductase superfamily. The PCR amplification (Fig 4.5) of B27 genomic DNA and IL12-4 and B27 cDNA using as primers the start and end of the gene (Supplemental figure S4.6), differents from the first S. *pennelli* epimerases amplification (Figure 4.6), exhibited a genomic fragment of 1511 bp and a cDNA fragment of 1365 bp. Therefore, the NCBI epimerase sequence was considered as reference for the following analyses.

Table 4.1- Epimerase genes in *S. pennellii* and M82 genome.

S. pennellii gene	M82 gene	Tissue where the gene is expressed
Sopen01g036950	Solyc01g091200	Roots
Sopen03g023360	Solyc03g083550	Heinz unopened flower buds
Sopen05g029460	Solyc05g050990	Heinz unopened flower buds
Sopen05g032260	Solyc05g053790	Breaker
Sopen07g002320	Solyc07g006220	Mature Green
Sopen08g027810	Solyc08g079440	Roots
Sopen09g035230	Solyc09g092330	Leaves
Sopen10g007490	Solyc10g018260	Roots
Sopen12g005550	Solyc12g010540	Breaker+10
Sopen12g033260	Solyc12g098480	Heinz 2 cm fruits
Sopen12g033320	Solyc12g098480	Heinz 2 cm fruits

ID genes of all *S. pennellii* epimerases and their homologous genes in M82 with the tissue where each gene is expressed.

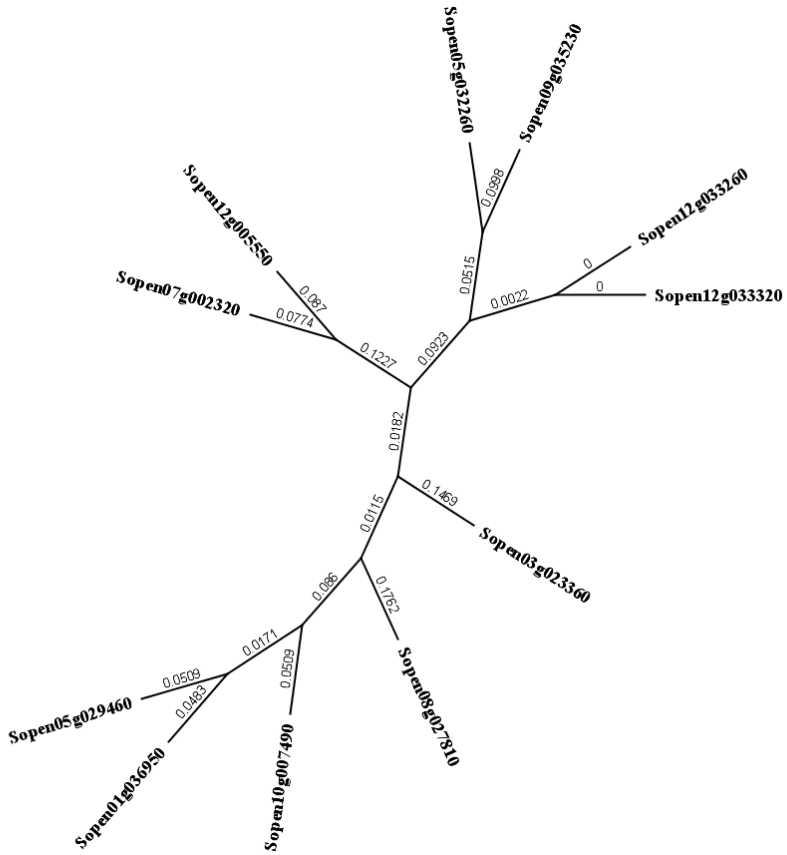


Figure 4.2-Phylogenetic tree of all epimerases in *S. pennellii* genome obtained using the Geneious software.

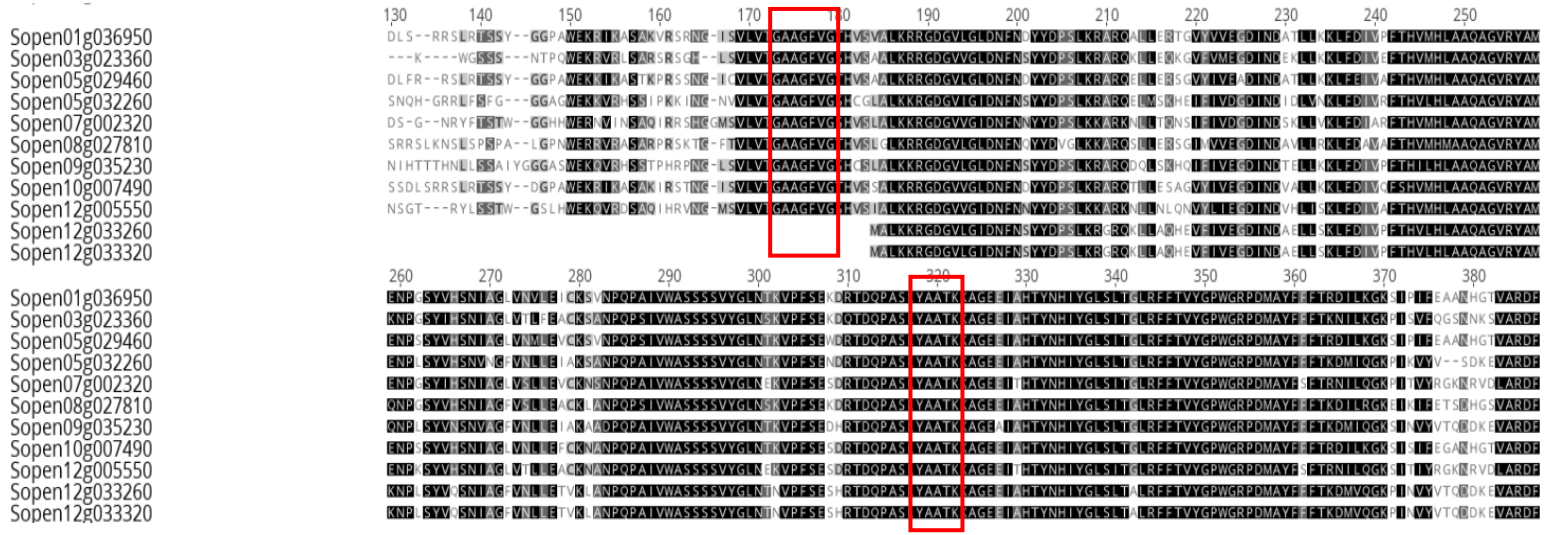


Figure 4.3- Part of the *S. pennellii* epimerase proteins alignment.

Red rectangles show the amino acid sequence motif GxxGxxG involved in binding of the cofactor NAD(P) and the YxxxK motif that is part of the catalytic domain of short-chain dehydrogenases/reductases. Sopen12g033260 and Sopen12g033320 lost motif GxxGxxG.

Figure 4.4- NCBI *S. pennellii* epimerase proteins alignment.

Red rectangles show the amino acid sequence motif GxxGxxG involved in binding of the cofactor NAD(P) and the YxxxK motif that is part of the catalytic domain of short-chain dehydrogenases/reductases.

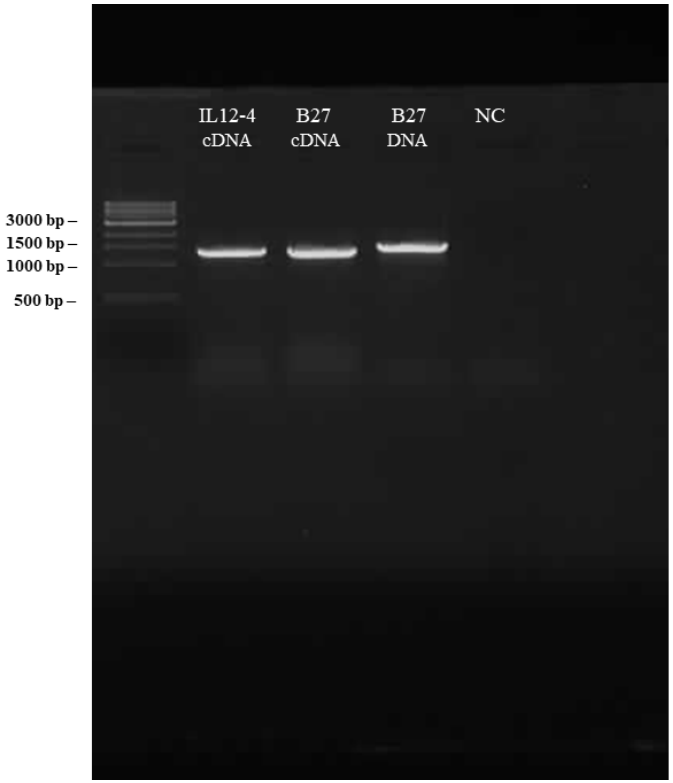


Figure 4.5 - PCR of epimerase genes on the chromosome 12 of *S. pennellii.*

Agarose gel of the PCR amplification using as primers the start and end of the gene to verify the presence of an intron *S. pennellii* epimerase genes, of the expected size 1365 bp, as reported in the NCBI database.

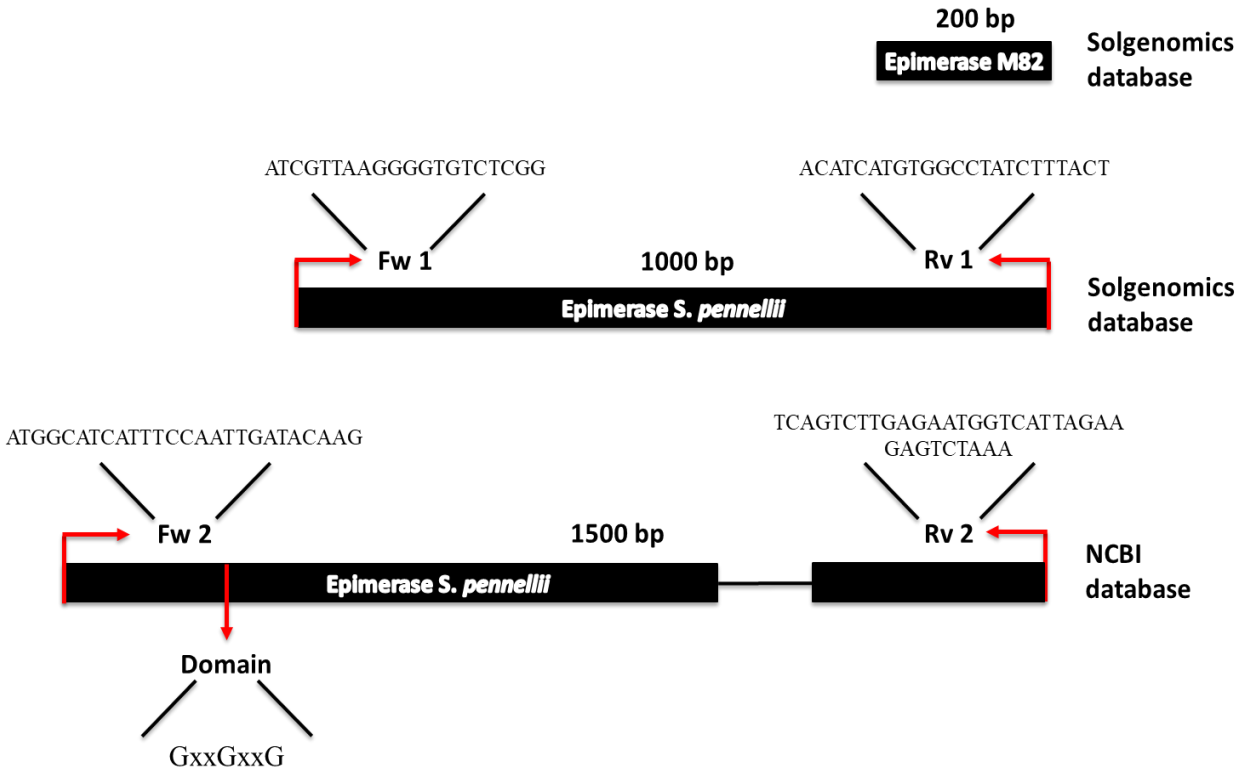


Figure 4.6– Structure of the Epimerase genes.

The primers FW1- RV1 were used for Sol Genomics S. *pennellii* epimerase genes discrimination, whereas the primers FW2- RV2 were used for the PCR amplification to verify the presence of an intron in *S. pennellii* epimerase genes as reported in NCBI.

4.2.2 Epimerase stable and transient overexpression

The effect of the *S. pennellii* epimerase overexpression was investigated to confirm the role of the UGlcAE as a gene involved in the increase of AsA content. Indeed, the XM_015204087 and the Solyc12g098480 were overexpressed using the Gateway Cloning Technology. The two genes, using the pDONR207 as entry vector and the pJAM2288 with the E8 promoter, specific for the fruit, were overexpressed in M82 plant for the stable transformation and in Money Maker fruits for studying their transient overexpression. As for the stable overexpression, M82 seedlings after one month from the co-cultivation showed both shoots and calluses; after two months shoots were placed in the rooting media using the Indole-3-butyric à-acid (IBA) as a rooting hormone (Figure 4.7 A and B). As for the transient overexpression, Money Maker fruits were infiltrated at breaker development stage as the E8 promoter involved in the biosynthesis of ethylene in the fruits. Fruits, collected two weeks after the agroinfiltration showed not statistically differences in term of AsA among the two overexpressed genes (Table 4.2), indeed the total ascorbate showed values from 2.05 to 2.16, whereas the oxidised ascorbate exhibited values from 2.07 to 2.20.

Figure 4.7- Stable overexpression of the M82 and *S. pennellii* epimerase in M82. A) Shoots and calluses in regeneration media after one month. B) Shoots in rooting medium after two months.

Table 4.2- Total and oxidised AsA in fruits of Money Maker transiently transformed with two epimerase genes.

	ng/µL AsA	
	Total	**Oxidised**
Solyc12g098480	2.05±0.04	2.07±0.05
XM_015204087	2.28±0.33	2.35±0.33
NC	2.16±0.08	2.20±0.09

Values of total and oxidised AsA content in Money Maker fruits infiltrated with the M82 epimerase gene Solyc12g098480, and the *S. pennellii* epimerase gene XM_015204087 genes and a negative control without infiltration (NC).

4.2.3 GGP plants genotyping

In the Cathie Martin's laboratory, the genes Solyc06g073320 and Solyc02g091510, coding for the GDP-L-galactosephosphorylase1 (GGP1) and GDP-L-galactosephosphorylase2 (GGP2), were knocked out using a genome editing approach. These two genes, involved in the L-Galactose pathway of the ascorbate biosynthesis, have an unusual long 5' UTR, 384 bp in the Solyc06073320 and 301 bp in the Solyc02g091510, respectively, containing a non-canonical uORF of 60 – 65 amino acid with an ATC initiation codon (Figure 4.7) previously studied in Arabidopsis as critical amino acid sequence in determining ascorbate repression of translation (Laing et al., 2015). To evaluate in tomato the function of the uORF as key peptide of ascorbate biosynthesis, *S. lycopersicum* cv Money Maker was transformed with a construct driving the expression of Clustered Regulatory Interspaced Short Palindromic Repeats (CRISPR) and CRISPR associated (Cas) genes expressing two single guide RNAs (sgRNA). This construct was designed to delete a region of the 5' UTR coding for the uORF of the two GDP-L-galactosophosphorylase. (Supplemental figure S4.7, S4.8).

uORF1- Solyc06g073320

```
1         10         20         30         40         50
|         |          |          |          |          |
ATCACGGCTCTTCTTGAATCTTTCGTTTGTATTCTCACAATTCATCATCA
CCGCAAAGTGTTGACCCTTAATCCAACTCTTCTGGTGGACGATAAGCACC
GGACCCCTTCCCCTCACGGAGGTAGGGGTGCCTCACCCGCTGAAGGCGGT
TGCCCCTCCGATCTCCTCTTCCTCGCCGGCGGCGGTCCAATTCTTCCTTT
CTCTTTTTCCTTCTCCTAA
```

uORF2- Solyc02g091510

```
1         10         20         30         40         50
|         |          |          |          |          |
ATCACGGCTATACACAAAGTAAACCGCCGACCACTTTTACATGTTCCAGC
AGTACGTCGTAAGGGTTGTGTAACAGCTACTAACCCTGCGCCGCACGGTG
GACGTGGCGCTTTGCCTTCTGAAGGTGGTAGTCCTTCCGACCTCCTCTTC
CTTGCCGGCGGCGGTTCTTTCCTCTCCTTCTCCTACTAG
```

Figure 4.7- DNA sequences of uORF1 and uORF2 of the two genes Solyc06g073320 and Solyc02g091510.

Most GGP plants, grown in the greenhouse at John Innes Centre, showed a phenotype similar to Money Maker, though three of them were phenotypically different respect to the parental genotype (Figure 4.8, Figure 4.9). As for the GGP1 plants, GGP1-2, GGP1-4, GGP1-5 were as Money Maker, whereas GGP1-1 and GGP1-3 exhibited a phenotype extremely different not only for leaves (thicker than Money Maker), but also the flower structure was unusual, indeed stamen pistil, petal shape was smaller than the normal and plants showed difficulties to produce normal fruits, as well as parthenocarpic fruits. As for GGP2, GGP2-1, GGP2-3, GGP2-4 and GGP2-7 were similar to Money Maker in phenotypic characteristics; although the flower structure was apparently similar to the wild type Money Maker, GGP2-7 showed difficulties to produce fruits (Supplemental figure S4.6).

As for genotyping, three fruits (A, B, C) were analysed for each plant, except for GGP2-3 (the plant produced only two fruits). Plants GGP1-1, GGP1-3 and GGP2-7 did not produce any fruit. As for GGP1 plants (Table 4.3), no differences were found in the uORF genomic and protein sequence respect to wild type gene Solyc06g073320. Only the fruit A of genotype GGP1-4 showed, on the first guide RNA, a single-nucleotide polymorphism (SNP) of adenine with thymine at position 41 from the start of the uORF. The analysis of this variation on the protein sequence using the software PROVEAN evidenced a neutral variation at position 14 from the start of the uORF of the glutamine (Q) with histidine (H).

As for GGP2 plants (Table 4.4), fruits of GGP2-1 showed different genotypes indeed the chromatogram of GGP2-1_A exhibited a wild type profile and a deletion on the first guide RNA of two nucleobases (guanine and cytosine) at positions 96 and 97 from the start of the uORF. This modification of the genomic sequence produced a deletion of the histidine at position 33 of the protein sequence and a premature stop codon generating an amino-acid sequence of 40 amino-acids. As for GGP2-1_C SNPs were found at positions 82, 89 and 180 from the start of the uORF (the first and the second on the first guide RNA and the third on the second guideRNA), however only the SNP at position 89 of the guanine for adenine produced a neutral variation at position 30 of the glycine (G) with aspartic acid (D).

As for GGP2-3, all fruits showed the same genotype, indeed looking at chromatograms was possible to identify a wild type profile and 4 SNP at positions 92 and 98 on the first guideRNA and two SNP at the position 186 on the second guideRNA. As for the first and second SNP we identify using PROVEAN software two predicted neutral substitution of at position 31 of glycine (G) with aspartic acid (D) and 33 of proline (P) with histidine (H) respectively, whereas the second two SNP at position 186 exhibited two predicted deleterious substitutions of arginine (A) with glycine (G).

As for the GGP2-4, the chromatogram of the fruit GGP2-4_A evidenced a heterozygous profile with two deletions on the first guideRNA at positions 95 -96 of an adenine and a guanine and at positions 94 -95 of two adenine, respectively. These two deletions of the lysine (K) at position 32 produced a premature stop codon and a peptide of 40 amino-acids.

For this fruit two insertions were also identified on the second guideRNA at position 181 of an adenine and a guanine respectively. As for GGP2-4_A and GGP2-4_C were found deletions at positions 95 – 96 on the first guide RNA and 182-183 on the second guide RNA. As for GGP2-6 plants no differences were found in genomic and protein sequence respect to Money Maker.

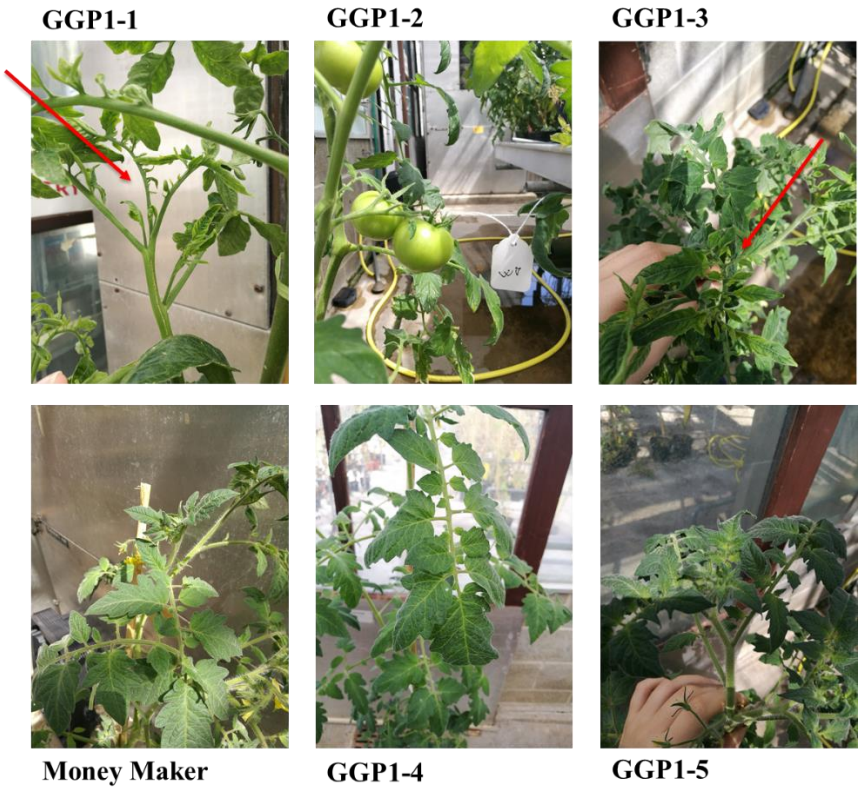


Figure 4.8 - Leaves and fruit of five GGP1 plants grown in the greenhouse.

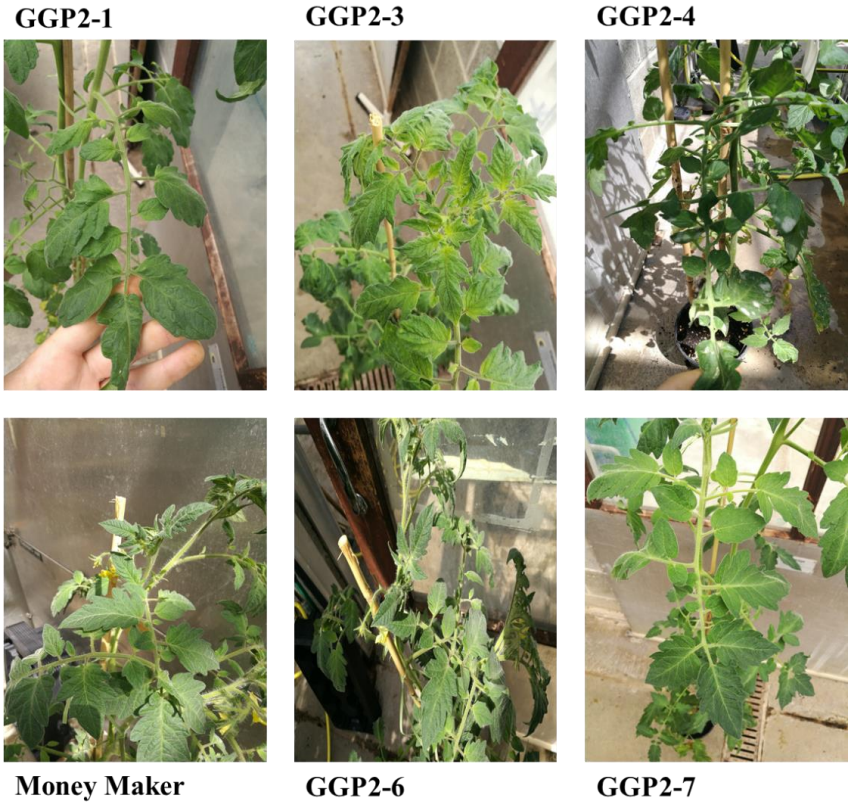


Figure 4.9-Leaves of five GGP2 plants grown in the greenhouse.

Table 4.3 - Genotyping of GGP1 fruits.

ID Plant	ID fruit	DNA mutations	Protein mutations
GGP1-2	GGP1-2_A	wt	wt
	GGP1-2_B	wt	wt
	GGP1-2_4	wt	wt
	GGP1-2_A	wt	wt
GGP1-4	GGP1-4-A	41T/A	H14Q
	GGP1-4-B	wt	wt
	GGP1-4-C	wt	wt
GGP1-5	GGP1-5_A	wt	wt
	GGP1-5_B	wt	wt
	GGP1-5_C	wt	wt

Table 4.4- Genotyping of GGP2 fruits.

ID Plant	ID fruit	DNA mutations	Protein mutations
GGP2-1	GGP2-1_A	wt	wt
		del. 96-97 GC	del. H33
	GGP2-1_B	C82T	wt
		A89G	D30G
		C180G	wt
		NF	wt
	GGP2-1_C	C82T	wt
		wt	wt
cGGP2-3	GGP2-3_A	wt	wt
		A92G	D31G
		A98C	H33P
		T186C	G62R
		T186A	G62R
	GGP2-3_B	wt	wt
		A92G	D31G
		A98C	H33P
		T186C	G62R
		T186A	G62R
cGGP2-4	GGP2-4_A	del. 95-96 AG	del. K32
		ins. 181 A	
		ins. 181 G	
		del. 94-95 AA	del K32
		del. 183-184 CG	
	GGP2-4_B	del. 95-96 AG	del K32
		del. 182-183 GC	
	GGP2-4_C	del. 95-96 AG	del K32
		del. 182-183 GC	
cGGP2-6	GGP2-6_A	wt	wt
	GGP2-6_B	wt	wt
	GGP2-6_C	wt	wt

4.3.3 AsA analysis of GGP plants

To verify if protein variations of uORF of GGP1 and GGP2 were related with an increase of AsA content in each genotype, for each plant two fruits were analysed when the protein structure was similar to the wild type and three fruits when the proteins were different respect to Money Maker. The qualitative analysis to identify potential variations of AsA content among all genotypes transformed respect to Money Maker (Table 4.5 and Table 4.6) showed some differences in terms of total and oxidised AsA. As for GGP1 plants, only the GGP1-4_B fruit showed a significant reduction of total and oxidised AsA and compared with the wild type, whereas the fruit GGP1-5-B exhibited a significantly higher value than Money Maker. As for GGP2 plants, the three fruits of GGP2-4 showed the statistically highest value in term of total and oxidised AsA respect to Money Maker, indeed GGP2-1_C and GGP2-6A exhibited a statistical increase of total AsA and GGP2-3_A, GGP2-3_B and GGp2-6_A for oxidised AsA, however their values were lower than GGP2-4 fruits. Among all genotypes only GGP2-1_A and GGp2-1_B showed a significant reduction respect to Money Maker in total and oxidised AsA.

Table 4.5- Total and oxidised AsA in GGP1 fruits.

	ng/µL AsA	
	Total	**Oxidised**
Money Maker	2.85±0.07	3.14±0.06
GGP1-2_A	2.68±0.11	2.85±0.27
GGP1-2_B	2.60±0.17	2.29±0.20
GGP1-4_A	2.93±0.09	3.16±0.05
GGP1-4_B	2.32±0.01*	2.29±0.03*
GGP1-5_A	3.28±0.14	3.38±0.16
GGP1-5_B	4.01±0.27	4.11±0.22*

Values of total and oxidase AsA in Money Maker and of GGP1 fruits of plants with genome editing of Solyc06g073320 using the CRISP/Cas9 technique. The significance of differences of each genotype *vs* M82 was evaluated by the t-Student's test (*: $P<0.05$).

Table 4.6- Total and oxidised AsA in GGP2 fruits.

	ng/µL AsA	
	Total	**Oxidised**
Money Maker	2.8±0.24	2.81±0.01
GGP2-1_A	2.07±0.11*	1.97±0.05***
GGP2-1_B	1.86±0.11**	1.91±0.02***
GGP2-1_C	4.42±0.01**	4.17±2.88
GGP2-3_A	2.60±0.17	2.48±0.06***
GGP2-3_B	2.39±0.13	2.44±0.11**
GGP2-4_A	11.92±0.46***	11.87±0.29***
GGP2-4_B	11.25±0.61***	10.97±0.69***
GGP2-4_C	5.72±0.45***	5.61±0.27***
GGP-2-6_A	3.91±0.03*	3.73±0.19**
GGP2-6_B	2.76±0.35	2.82±0.01

Values of total and oxidase AsA in Money Maker and of GGP1 fruits of plants with genome editing of Solyc02g091510 using the CRISP/Cas9 technique. The significance of differences of each genotype *vs* M82 was evaluated by the t-Student's test (*: P<0.05; ** :P<0.01; ***:P<0.001).

Chapter 5 Discussion

Since the last centuries the cultivated genotype *S. lycopersicum* lost some important traits due to continuous rounds of evolution, domestication and breeding. The modern cultivars, indeed, as result of these genotypic and phenotypic changes, show lower amounts of flavour chemicals compared to older and wild varieties (Fernandez-Moreno et al., 2017, Tieman et al.,2017). However, though at beginning the breeders interest was focused especially on productivity traits as yield or other characteristics such as disease resistance (Ballester et al., 2016), in the last years the attention of tomato breeders has moved towards quality traits, such as flavour, antioxidant and total sugars content (Sacco et al., 2013, Tieman et al., 2017). The wild species, such as S. *pennellii*, are reservoir of genetic variations respect to the cultivated genotype and are an important source of new alleles to exploit for breeding selection (Liu et al, 2016, Bolger et al., 2014). Currently, the most important sources to identify genomic regions associated with quantitative traits (QTL) are introgression lines (ILs), which represents the full wild species genome divided in small introgression regions. Among the ILs populations, the 76 ILs of S. *pennellii* in the M82 genetic background are the most used worldwide (Eshed and Zamir, 1995, Fernandez-Moreno et al., 2017, Krause et al., 2018). Indeed, in the last decades *S. pennellii* ILs were used as genetic material to identify around 3000 quantitative trait loci (QTLs) (Alseekh et al., 2013) directly involved in productivity and quality traits, such as yield, metabolic compounds, °Brix, and stress tolerance (Alseekh et al., 2013, Fernandez-Moreno et al., 2017, Krause et al., 2018). QTL mapping is a powerful technique for dissecting the genetic basis of economically interesting traits (Fulop et al., 2016), and the study of QLTs allows to associate each phenotypic variation with the presence of the wild regions, increasing the capacity of QTLs identification and consequently to detect candidate genes for each trait.

In addition, using the genomic resources available for tomato, such as the sequencing of S. *pennellii* (Bolger et al., 2014) and ILs (Chitwood et al., 2013) genomes, the study of the transcriptomics data of the parental genotypes and ILs (Koenig et al., 2013, Ranjan et al., 2016), the information on the ILs metabolic profile

(Schauer et al., 2006) and bioinformatic approaches for studying metabolic pathways (Ruggieri et al., 2016), it is nowadays possible to identify possible candidate genes to improve quality traits.

In the present thesis, two *S. pennellii* ILs (IL7-3 and IL12-4), and a group of sub-lines derived from them, have been investigated in combination with some genomic resources in order to identify genes controlling steps of the AsA biosynthetic pathways that could increase ascorbate content in red ripe fruit. In particular, after selecting the superior sub-lines that stably exhibit high levels of AsA in the fruit, these lines were investigated by molecular markers and bioinformatic tools, besides by transcriptomic analysis, that comprehensively allowed identifying candidate genes increasing AsA. Finally, the functional validation of some candidate genes was also explored.

5.1 Selection and characterization of *S. pennellii* sublines

Among the *S. pennellii* ILs, the two IL7-3 and IL12-4 lines were previously selected for their high content of AsA (Di Matteo et al., 2010, Sacco et al., 2013; Rigano et al., 2014). These ILs were also crossed to combine the identified QTLs and to produce superior materials with an increased amount of antioxidant compounds (Rigano et al.,2014, Calafiore et al., 2018), but in some cases the pyramiding of wild regions determined also negative effects. Therefore, the selection of sub-lines carrying smaller introgression regions compared to those of IL7-3 and IL12-4, and their intercrosses, would allow combining in the hybrids the positive effects reducing the negative ones, in prevalence affecting yield. The pyramiding of genes coming from the sub-lines could therefore lead to get surprisingly results, as reported in hybrids obtained by Gur and Zamir (2015), which had smaller introgressed fragments than ILs and exhibited significantly and consistently improved °Brix and yield.

The IL7-3 and IL12-4 were crossed in breeding programs in order to break the *S. pennellii* introgression fragment in different sub-lines (Alseekh et al. 2013, Ruggieri et al., 2015, Calafiore et al., 2016).

As for the IL7-3, the seven introgression sub-lines investigated in the present thesis were genotyped by Calafiore et al. (2016) combining metabolic analysis and genomics tools to identify QTLs directly involved in the control of quality traits, whereas the IL12-4 sub-line called B27 was selected by Ruggieri et al (2015) for its high content of AsA.

Looking at the productivity of the sublines and their parental genotypes in the years 2016 and 2017 in different environments it was possible to evidence a yield reduction in IL7-3 compared with M82, as reported by Eshed and Zamir (1995) and Sacco et al (2013). Indeed, for studying in the IL 7-3 parameters linked to the production, the attention was focused on the number of flowers *per* inflorescence, fruit set, number of fruits and their fruit weight. Particularly, comparing IL7-3 with M82 fruit set was identified as critical stage for the lower production of the introgression line. Indeed, despite the value of the number of flowers *per* inflorescence was similar to the cultivated genotype, the fruit set and the other parameters contributed to reduce yield of the introgression line IL7-3.

On the contrary, almost all the introgression sub-lines exhibited yield values similar to the genotype M82, except R179 and R202, which showed in the year 2017 at Battipaglia field a significant lower, whereas in the other fields yield values were lower though not statistically significant. In R179 the reduction observed at Battipaglia was mainly due to a reduction of fruit number, whereas in the second genotype the yield reduction seems more affected by smaller fruit weight. These results are in accordance with Monamodi et al. (2013) and Gur and Zamir (2015), who suggested that both FW and NFR are important components to be considered in the selection of superior genotypes for yield. Among all genotypes, R181 and R182 showed the best performances for all yield-related traits in all the environments tested, as well as B27, which always exhibited productivity values comparable with M82, thus confirming this genotype as a good line for the improvement of the productivity traits.

As for AsA content, B27 showed in the two years and in all environmental conditions a significant higher value than the cultivated genotype in accordance with results of Ruggieri et al. (2016), confirming this sub-line as elite material for AsA content.

These values allowed to focus our interest in the study of the B27 introgression region to identify possible candidate genes involved in the AsA pathways. As for the IL7-3 and its sub-lines, most genotypes showed higher values of AsA respect to M82, though among all sub-lines R181 and R182 exhibited the best performance. In the year 2017 our interest was also focused in the analysis of AsA in leaves for all genotypes considered in the two years. Particularly, only B27 and R182 showed a higher content of this compound also in leaves. Moreover, considering the interaction among the AsA content in fruits and leaves, we found a significant negative correlation among AsA in fruits and leaves in Acerra -0.047, whereas a positive not significant correlation 0.4602 was found at Giugliano. Looking at the AsA GxE interactions in fruits and leaves, it is clear that the environmental conditions play an important role in determining AsA content. Indeed, as reported by Bulley and Laing (2016) ascorbate showed a diurnal change, being lower in the day than the night (Wang et al., 2013) probably due to more ascorbate consumed during the day. Moreover, important genes involved in ascorbate biosynthesis showed transcription changes to respond to some environment condition such as light intensity, temperature and time of day (Bulley and Laing, 2016). Indeed, it was demonstrated that transferring plants from shade to sunshine the vitamin C content was increased significantly around the 13 % respect to the 'full shade' (Dumas et al., 2003). In particular, there is a controversy concerning the content in ascorbic acid under conditions of environmental stress, indeed some studies have noted an increase of AsA content when tomatoes were exposed to high temperatures and strong solar radiation, whereas other authors have reported a reduction of AsA under similar conditions (Rosales et al., 2011). B27 and R182, though exhibited different absolute values of AsA content among the three fields in the two years, probably due to the different environmental conditions, always showed higher values respect to the cultivated control M82. This could confirm the genetic importance in the control of the AsA biosynthesis of these two sub-lines.

Finally, considering all the traits analysed, the sub-line R182 was the best performing among the IL7-3 sub-lines. The potentially higher nutritional value of sub-lines R182 and B27 fruits, combined with the other good traits, makes these sub-lines the elite ones selected to focus the interest in studying their introgressed regions with the aim of identifying possible candidate genes involved in the increase of AsA content in the fruit.

5.2 Identification of candidate genes controlling high AsA content in the selected sub-lines B27 and R182

Since the last three decades, *S. pennellii* ILs were studied for identifying QTLs involved in different biological and agronomic processes (Alseekh et al., 2013, Fulop et al., 2016, Fernandez-Moreno et al., 2017, Krause et al., 2018), such as yield, disease resistance and tolerance to abiotic stresses (Eshed and Zamir, 1995, Mutschler et al., 1996, Foolad et al., 1998) The ILs were identified for the first time through a succession of backcrosses in the year 1992 by Zamir (Eshed et al., 1992, Eshed and Zamir, 1994), who characterized them using a total of 146 SGN molecular markers and as reference genome S. lycopersicum cv Heinz. Then, ILs were again characterized, using 2506 SGN molecular markers and Heinz as reference genome (Figure 5.1) (Fulton et al., 2002).

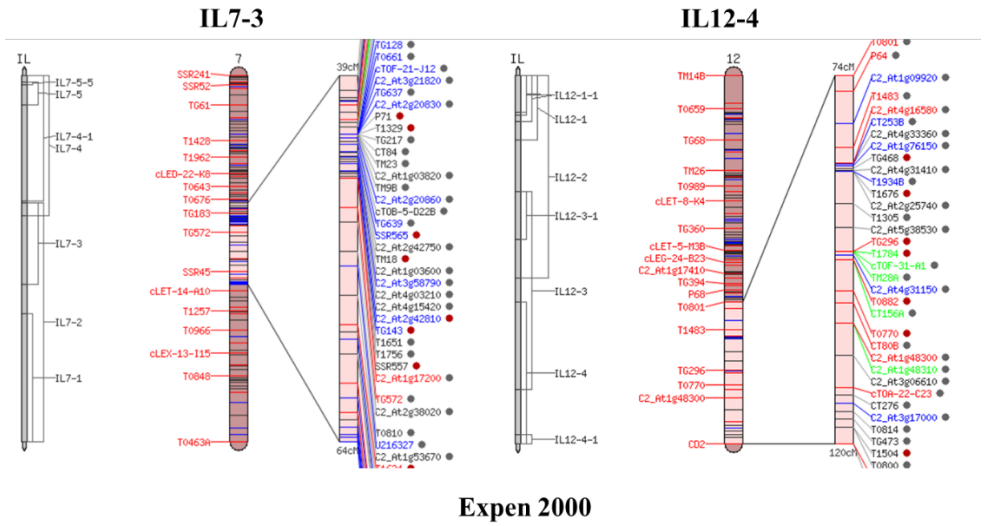


Figure 5.1- Example of characterization of IL7-3 chromosome 7 and IL12-4 chromosome using molecular markers and the tomato Expen 2000 map (https://solgenomics.net).

Moreover, Chitwood et al. (2013) studied the ILs defining them using around 750,000 SNP/indel and Heinz as reference genome (Figure 5.2).

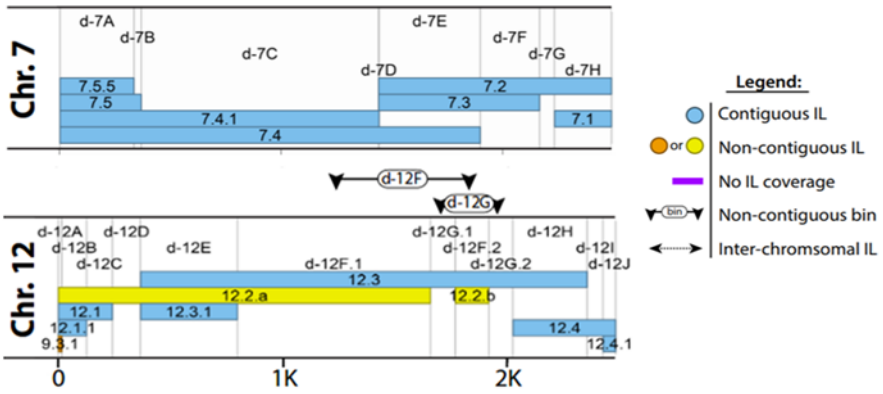


Figure 5.2- Example of characterization of the chromosomes 7 and 12 of ILs using SGN SNPs and Heinz as genome reference (Chitwood et al., 2013).

However, despite ILs characterization of Chitwood et al. (2013) was much more precise of that carried out by Fulton (Tomato-EXPEN 2000) (Fulton et al., 2002) in terms of number of genes and introgression size, both efforts relied on the use of the genome of Heinz as unique reference, though the ILs parental lines derive from *S. pennellii* and *S. lycopersicum* cv. M82. This over-simplification could theoretically produce erroneous results when trying to precisely identify borders and extension size of the introgressions since *S. pennellii* is a wild species with a genome quite different from *S. lycopersicum* in terms of size, co-linearity, transposable elements, etc (Bolger et al., 2014). Indeed, Bolger and colleagues (2014) highlighted that *S. pennellii* genome size was around 1.2 Gbp, about 30% larger/bigger than the *S. lycopersium* one. Also, they annotated around 32,273 and 44,966 protein coding genes according to two different approaches used (Bolger et al., 2014), which represent around +10% and around +25% genes when compared against *S. lycopersium* annotation (SL3.1). Finally, although the genotype M82 belongs to the same species of the reference genome (*S. lycopersicum* cv Heinz) more than 1,330,000 variants were observed between these two genotypes in a previous study (Bolger et al., 2014).

So a correct characterization of the introgression lines should take into consideration both parental genomes to identify either the wild regions inserted (from the *S. pennellii* donor) and the cultivated background region removed (from the *S. lycopersicum* M82). Since at present a vast collection of "omics" data are available for both parental lines, these were bioinformatically exploited in the present thesis, in order to fine characterize, re-assemble and annotate the two introgression lines IL7-3 and IL12-4 genomes. By exploiting public genomic resources available for *S. lycopersium* cv M82 (ERP005818) and *S. pennellii* (ERP005244), a set of specie-specific markers were obtained and used to correctly identify the breaking points and the size of each introgression on both parental lines allowing a precise identification of the genomic differences between each IL and the control parent M82. This will allow to correctly identify the key genes associated with the phenotypic changes (QTLs) observed for the ILs, and mainly related to the antioxidants content, and with the genetic differences underlying those traits. First step, in this process, was the reconstruction of the *S. lycopersicum* M82 genome that represents the background parent of the IL population. We observed more than 2,000,000 variants when comparing M82 and Heinz genotypes. This result was in accordance with Bolger et al. (2014). Those variants were used to "reconstruct" the M82 genome starting from the reference Heinz genome (SL3.0). The process was based on a complex bioinformatics workflow (Reconstructor, see M&M) where the variants of M82 (both SNPs and short INDELs) were reiteratively checked and in turns inserted in the genome. In order to improve the quality of the genetic annotation, two different annotations were used to annotate the reconstructed M82. The approach relied on the integration of two lift-over processes where both the ITAG3.1 and NCBI 102 annotations where transferred on the new genome. As a result of this integration we observed that our M82 annotation reported a higher number of genes, 7.5% of respect to the SGN SL 3.1 release and 7% respect to the NCBI v 102 release. Moreover, the IL7-3 and IL12-4 reassembly carried out using NGS RNA-seq and RAD-seq raw data showed some differences in term of introgression size and number of genes mapping in the introgression fragment. Comparing the two IL7-3 and IL12-4 characterizations (Table 5.1), it is evident that Chitwood et al. (2013) showed introgression region size values almost comparable respect the IL reassembly, whereas different is the situation

for the introgression genes number, where values are different among the two characterizations.

Table 5.1- Number of genes and introgression size of IL7-3 and IL12-4 as reported in Chitwood et al. (2013) and IL reassembly in the present thesis.

Chitwood study		IL reassembly in the present thesis			
Size Heinz introgr. region (Mbp)	**Heinz introgr. genes**	**Penn genes**	**M82 genes**	**Size Penn. region (Mbp)**	**Size M82 region (Mbp)**
6.67	745	716	817	7.14	6.66
3.11	447	433	485	3.4	3.09

Moreover, the border gene introgression line positions of Chitwood (ITAG 2.3) were used to find the new position on the ITAG 3.1 release. Looking at Table 5.2, it is possible to evidence some differences in the first and last introgressed genes of IL7-3 and IL12-4. In some cases we are in accordance with Chitwood et al. (2013) whereas in other cases the border genes have been validated using molecular markers to confirm the reliability of the reassembly.

Table 5.2- Break points of the IL7-3 and IL12-4 as reported by for Chitwood et al. (2013) and in the ILs reassembly carried out in the present thesis.

	Introgression line start				Introgression line end			
	Chitwood		IL reassembly		Chitwood		IL reassembly	
	Heinz	Pennellii	M82	Pennellii	Heinz	Pennellii	M82	Pennellii
IL7-3	Solyc07g047990	Sopen07g024420	Solyc07g048010	Sopen07g024420	Solyc07g063400	Sopen07g021980	Solyc07g063390	Sopen07g031590
IL12-4	Solyc12g088810	Solyc12g025390	Solyc12g088840	Sopen12g030710	Solyc12g100357	Sopen12g035030	Solyc12g100357	Sopen12g035030

The surprising quality and agronomic performances observed in the elite IL7-3 sub-line R182 carrying a small introgression fragment (around 384 Kbp) induced to study and better define the size and genes mapping in the wild region, which could be directly involved in the increase of AsA. Indeed, this sub-line could be an important source of candidate genes for quality and productivity traits, as well. However, though the R182 sub-line was previously characterized by Calafiore et al. (2016), the presence of wild/cultivated alleles mapping in this introgressed region was better investigated by designing additional molecular markers, using the new S. *lycopersicum* release and S. *pennellii* (Bolger et al., 2014) genome information available in the Sol Genomics Network database (https://solgenomics.net).

Particularly, SCAR markers were constructed not only to identify the first introgressed gene, but also to confirm the begin of the IL7-3 introgression region, as the comparison among the orthologous of S. *lycopersicum* and S. *pennellii* allowed to identify a mistake regarding the upper border. Indeed, the analysis carried out with molecular markers confirmed the results of Chitwood et al. (2013) and the first introgressed wild gene is Solyc07g047990 instead of the Solyc07g048010. The reason of this mistake is due to the expression of these genes, since using the Tomato functional genomics database (http://ted.bti.cornell.edu/), it was found that this gene is particularly expressed in roots and it has a very low expression (in term of RPKM) in the leaves. The RNA-seq data derived from the experiment of Chitwood et al. (2013) is based on RNA extracted from leaves. For this reason, no variants were detected during the bioinformatic analyses carried out to reconstruct the introgressed region of IL7-3. However, the analysis with molecular markers allowed to demonstrate that the sub-line R182 carries wild alleles from the gene Solyc07g04990, which corresponds to Sopen07g024210, to the gene Solyc0749310 corresponding to Sopen07g024640. Among all genes mapping in the introgressed region of R182, the attention was focused on Solyc07g049280/ Sopen07g024610 (Pyrophosphate-fructose 6-phosphate 1-phosphotransferase (PFP) subunit beta) and on two Major facilitator superfamily (MFS) membrane proteins

(Solyc07g049290/Sopen07g024620 and Solyc07g049310/Sopen07g024640) that could have a role in the increase of AsA in the fruit of R182. Indeed, the PFP is a key enzyme involved in the Glycolysis / Gluconeogenesis pathways (Duan, et al., 2016), whereas MFS perform the import or export of target substrates as sugars (Reuscher et al., 2014). The interest on the MFSs was confirmed by the analysis of the transcriptomics profile, carried out between the sub-line R182 and the cultivated genotype M82 at breaker and mature red stages, using an RNA-seq approach. Indeed, from these analyses it was pointed out that Solyc07g049290 and Solyc07g049310 were up-regulated. These two genes were therefore selected as good candidates for the increase of the AsA content in the R182 and their function will be validated by using genome editing techniques, such as the CRISP/Cas9 technology.

In a previous study, Ruggieri et al., (2015) identified 32 candidate genes in the introgressed fragment of the IL12-4 that might have a function related to AsA biosynthesis and accumulation, such as apoplastic L-ascorbate oxidases (Solyc12g094460, Solyc12g094470), glutathione dehydrogenases (Solyc12g094430, Solyc12g097080), and peroxidases (Solyc12g089360, Solyc12g089370, Solyc12g096530) involved in the mechanisms of AsA recycling; glyceraldehyde 3-phosphate dehydrogenase (Solyc12g094640), diphosphate-fructose-6-phosphate 1-phosphotransferase (Solyc12g095760), and phosphofructokinase (Solyc12g095880), involved in carbohydrate and sugar metabolism; pectinesterases (Solyc12g098340, Soly12g099230, Solyc12g099410), polygalacturonases (Solyc12g096730, Solyc12g096740, Solyc12g096750), UDP-D-glucoronate 4-epimerase (Solyc12g098480), glucosyltransferases (Solyc12g096080, Solyc12g096820, Solyc12g096830, Solyc12g096870, Solyc12g098580, Solyc12g098590, Solyc12g098600) and myo-inosytol oxygenase (Solyc12g098120), involved in the degradation of cell wall polymers (Ruggieri et al., 2015).

As for the IL12-4 sub-line B27, the analysis of the transcriptomics profile using the RNA-seq technique performed in the years 2016 and 2017 between the sub-line and M82 on mature red fruit, showed that the pectinesterase Solyc12g098380 was up-regulated in both the years, whereas the UDP-glucoronic-acid-4-epimerase Solyc12g098480 was up-regulated only in the year 2016.

Since in a previous study it has been demonstrated that the latter gene is over-expressed in mature green and breaker stages (Rigano et al., 2018), it was chosen as candidate gene to be further investigated by genetic transformation techniques.

5.3 Strategies to study the function of genes controlling AsA content in tomato fruits

In the last years, the functional validation of candidate genes has become a crucial step in new breeding programs (Zhang et al., 2018, De Oliveira Silva et al., 2018, Tang et al., 2018). Indeed, this validation allows to identify genes involved in the QTLs expression and to improve some traits in tomato. In the present thesis, it was decided to focus the attention on the Solyc12g098480 gene mapping in the introgression region of the sub-line B27 and also supposed as gene determining the content of AsA by Rigano et al. (2018). The B27 reassembly showed the presence of two identical S. *pennellii* UDP-glucoronic-acid-4-epimerases much more longer than the corresponding cultivated epimerase (Solyc12g098480), which was supposed to be truncated and not functional. Looking at the gene structure, as reported by Molhoj et al. (2004), the two S. *pennellii* epimerases were intronless, thus lacking the amino acid sequence motif GxxGxxG, which is involved in binding of the cofactor NAD(P). By contrast, considering the NCBI S. *pennellii* annotation, two identical UDP-glucoronic-acid-4-epimerases were found that have in their amino acid sequence the motif GxxGxxG involved in binding of the cofactor NAD(P) and the YxxxK motif, which is part of the catalytic domain of short-chain dehydrogenases/reductases. The only difference was that these epimerases showed the presence of an intron.

To verify and validate the function of these particular epimerases carrying an intron two different genome editing approaches were applied. The transient transformation was performed to investigate the AsA content when the cultivated genotype was transformed with vectors determining the over-expression of the S. *pennellii* epimerase and the S. *lycopersicum* epimerase. Unfortunately, no differences were detected in the content of AsA among the infiltrated tomatoes and the negative control probably because, when the agroinfiltration was performed, the Agrobacterium did not reach all parts of the fruits.

As consequence, results were not clear and the AsA content was very low when measured on the whole fruit. The other approach used for the epimerase genes functional validation was the stable transformation through Agrobacterium. This is a good tool to transfer the gene to next generations, though this technique is often inefficient and the transient transformation is preferred (Bond et al., 2016). The stable transformation experiments carried out in the present thesis at the John Innes Institute allowed to get ten transformed plants, whose genotyping and phenotyping is in progress.

A more recent alternative method to validate the function of a gene is the CRISP/Cas9 approach. This approach was used in the Cathy Martin's lab to study the function of two genes (Solyc06g073320 and Solyc02g091510) encoding for the GDP-L-galactosephosphorylase1 (GGP1) and GDP-L-galactosephosphorylase2 (GGP2). The two GGP genes are directly involved in the L-galactose pathway of the ascorbate biosynthesis (Figure 5.3). Indeed, the enzyme GGP catalyses an important step of the pathway, and its overexpression in plant showed an important increase of AsA content (Laing et al., 2007, Bulley et al., 2009, Bulley et al., 2012).

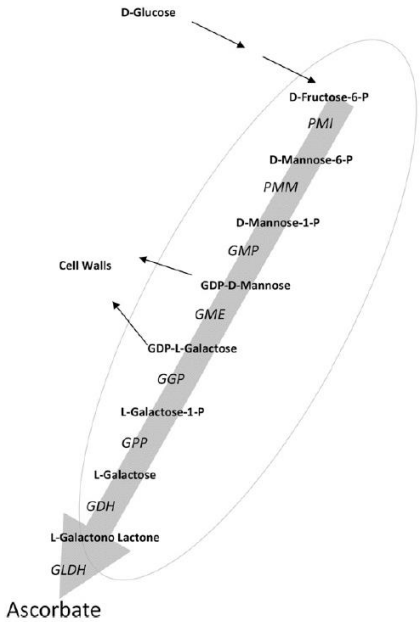


Figure 5.3- The L-Galactose Pathway of Ascorbate Biosynthesis (Laing et al., 2015).

As reported by Laing et al (2015), these two genes carry an unusual long uORF with uncanonical start codon, which could regulate the biosynthesis of ascorbate (Figure 5.4). There are two hypotheses on the role of uORF in the translation inhibition. The first one considers the uORF as directly involved in the efficiency of the translation, whereas the second one considers that the uORF peptide causes ribosome stalling during uORF translation (Fang et al., 2004, Tran et al., 2008, Laing et al., 2015).

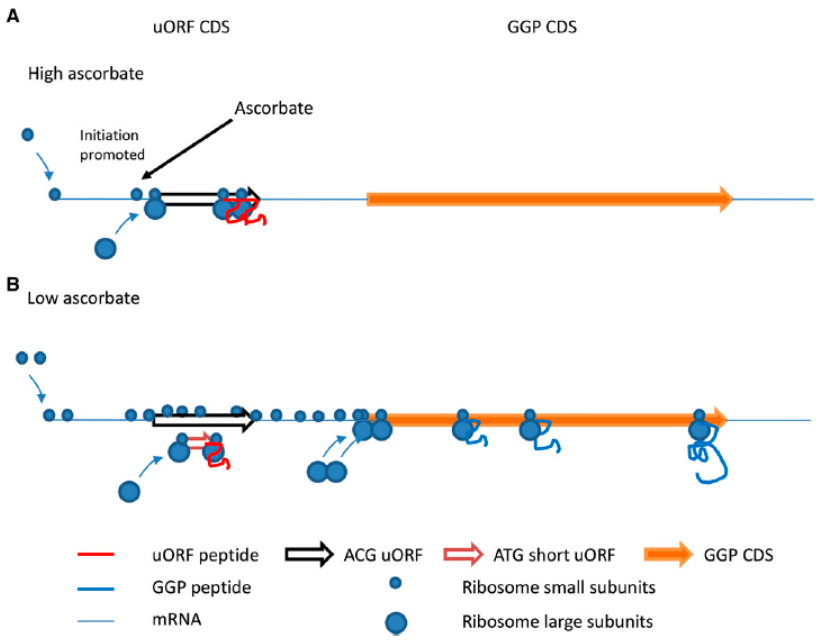


Figure 5.4- Simplified Scheme for Ribosome Interaction with the GGP mRNA at the uORF and CDS. The model assumes that high ascorbate (A) promotes recognition of the uORF ACG start codon, while under low ascorbate, the ribosomes skip the uORF (B) and start translation at the main GGP CDS. The likelihood that the short ATG uORF is translated is included in (B). (Laing et al., 2015).

To study if these uORF is involved in the AsA biosynthesis in tomato as in Arabidopsis, two guideRNA were designed on the uORF to inactivate the peptide. The phenotyping of the GGP1 and GGP2plants was confirmed by the results of the hairy roots genotyping (Table 5.3) carried out by Dario Breitel of the John Innes Centre. Indeed, were identified almost the same genetic variation among the two experiments, though hairy showed more alleles than plants.

Moreover, were identified some genotype knock out with a deletion on a single guideRNA or on both guide RNA. However, the analysis of AsA content in the fruits confirmed the work of Laing et al (2015) only when there was a deletion on both guideRNA in homozygous genotypes.

Therefore, this experiment shows that the concentration of AsA in tomato can be directly regulated by the presence of these uncanonical uORFs on GGP enzymes.

Table 5.3- Genotyping of GGP1 and GGP2 hairy roots.

Stable transformation of CRISPR casettes

ID uORF	ID line	allelle	Mutation DNA	Mutation protein
GGP1	cGGP1-1	1	del. 55-57 CGT del. 160-165 GGCGGT	del. R 19 del. GG 54-55
		2	del. 52-55 GTAC del. 157-160 GGCG	Change of ORF from residue 19 and introducing premature STOP codon after 22 residues
	cGGP1-2	X	wt	wt
	cGGP1-3	1	del. 56-57 GT del. 160-163 GGCG	Change of ORF from residue 18 and introducing premature STOP codon after 19 residues
		2	del. 52-57 GTACGT del. 160-165 GGCGGT	del. VR 18-19 del. GG 54-55
	cGGP1-4	X	wt	wt
	cGGP1-5	X	wt	wt
GGP2	cGGP2-1	1	wt	wt
		2	del. 91-97 GATAAGC	Change of ORF from residue 31, creating a new ORF into GGP2 coding sequence, affecting its sequence., creating a 108 aa peptide.
		3	del. 85-96 GTGGACGATAAG del. 182-189 GCGGTCCA	del VDDK 29-32; Change of ORF from residue 57 creating a chimeric protein with GGP2 in frame
		4	del. 94-96 AAG	del K32
		5	del. 95-109 AGCACCGGACCCCTT	Change of KHRTPS 32-37 to T32
	cGGP2-3	X	wt	wt
	cGGP2-4	1	del. 95-96 AG	Proper uORF up to residue 31, introducing premature stop codon, generating a 40 aa peptide
	cGGP2-6	X	wt	wt
	cGGP2-7	1	wt	wt
		2	del. 94-96 AAG	del K32
		3	del. 96 G	Change of ORF from residue 32, creating a new ORF into GGP2 coding sequence, affecting its sequence., creating a 110 aa peptide.

Chapter 6 Conclusions

The objective of this study was to exploit genetic and genomic resources to increase the ascorbic acid content in the tomato fruit. In particular, this study focused on the identification and validation of candidate genes directly involved in the AsA pathways exploring genomic regions of the wild species *S. pennellii*.

In the first chapter, the characterization for productive and qualitative traits of seven IL7-3 sub-lines and one IL12-4 sub-line, carried out in two years and in different environment conditions, allowed to select the genotypes R182 and B27 as elite sub-lines. Results obtained also demonstrated that introgression regions of small size coming from a wild species could confer positive traits to tomato fruits, such as the increased AsA content and firmness, without detrimental effects on other traits.

In the second chapter, combining public NGS RNA-seq and RAD-seq data, the reconstruction of M82 genome and the reassembly of the IL7-3 and IL12-4 genomes were obtained. In addition, using genetic and genomic information available for the two ILs, the size and wild S. *pennellii* genes mapping in were identified. Finally, data obtained from RNA-seq experiments carried out to compare the transcriptomic profile of B27 *vs* M82 and of R182 *vs* M82 allowed to identify one and three candidate genes for the increase AsA content, in the B27 and R182 sub-lines, respectively. The first one encodes for a UDP-glucoronic-acid-4-epimerase (UGlcAE), the other three encode for a Pyrophosphate-fructose 6-phosphate 1-phosphotransferase (PFP) subunit beta involved in the in the Glycolysis / Gluconeogenesis pathways and for two Major facilitator superfamily (MFS) membrane proteins that perform the import or export of target substrates as sugars.

In the third chapter, the validation of the B27 candidate epimerase gene was carried out to confirm its role in the increase of the AsA content in tomato fruits. The functional validation performed by a transient overexpression did not show any statistical difference of AsA content in the overexpressed fruits respect to the negative control.

Moreover, the stable overexpression of the cultivated and the wild *pennellii* epimerases mapping in the introgression region of B27 has been carried out to validate the epimerase function and hopefully would lead in the next future to obtain a high producing AsA cultivated line in a breeding program. Finally, the genotyping and phenotyping of some plants edited by the CRISP/Cas9 demonstrated that unusual uORFs on the GDP-L-galactosephosphorylase genes are involved in the regulation of the ascorbate biosynthesis also in tomato.

Comprehensively, in the present thesis the combination of genetic and genomic resources available for tomato led to obtain novel information related to the regulation of ascorbic acid content in red ripe fruit and to select superior genotypes. Both these achievements could be exploited in the next future to produce tomato fruit with a higher nutritional value.

References

Akram, N. A., Shafiq, F., & Ashraf, M. (2017). Ascorbic acid – a potential oxidant scavenger and its role in plant development and abiotic stress tolerance. *Frontiers in Plant Science*. 8, 613.

Alseekh, S., Ofner, I., Pleban, T., Tripodi, P., Di Dato, F., Cammareri, M., Mohammad, A., Grandillo, S., Fernie, A. R. & Zamir, D. (2013). Resolution by recombination: breaking up Solanum pennellii introgressions. *Trends Plant Sci.*18, 536-538.

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. (2016). Identification of Loci Affecting Accumulation of Secondary Metabolites in Tomato Fruit of a Solanum *lycopersicum* x Solanum *chmielewskii* Introgression Line Population. *Front Plant Sci.* 7, 1428.

Bassett, A. R., Tibbit, C., Ponting, C. P. & Liu, J. (2013). Highly efficient targeted mutagenesis of *Drosophila* within the CRISPR/Cas9 system. *Cell Reports*. *4*, 220−228.

Bertin, N., & Génard, M. (2018). Tomato quality as influenced by preharvest factors. *Scientia Horticulturae*. 233, 264–276.

Bolger, A., Scossa F., Bolger M. E., Lanz, C., Maumus, F., Tohge, T., Quesneville, H., Alseekh, S., Sorensen, I. & Lichtenstein, G. (2014). The genome of the stress-tolerant wild tomato species Solanum pennellii. *Nature Genetics.* 46, 1034-1038.

Bond, D. M., Albert, N. W., Lee, R. H., Gillard, G. B., Brown, C. M., Hellens, R. P. & Macknight, R. C. (2016). Infiltration-RNAseq: transcriptome profiling of Agrobacterium-mediated infiltration of transcription factors to discover gene function and expression networks in plants. *Plant Methods.* 12, 41.

Bulley, S.M., Rassam, M., Hoser, D., Otto, W., Schünemann, N., Wright, M., MacRae, E., Gleave, A. & Laing, W. (2009). Gene expression studies in kiwifruit and gene over-expression in Arabidopsis indicates that GDP-L-galactose guanyltransferase is a major control point of vitamin C biosynthesis. *J. Exp. Bot.* 60, 765–778.

Bulley, S., Wright, M., Rommens, C., Yan, H., Rassam, M., Lin-Wang, K., Andre, C., Brewster, D., Karunairetnam, S., Allan, A.C., & Laing, W.A. (2012). Enhancing ascorbate in fruits and tubers through over-expression of the L-galactose pathway gene GDP-L-galactose phosphorylase. *Plant Biotechnol. J*. 10, 390–397.

Bulley, S. & Laing, W. (2016). The regulation of ascorbate biosynthesis. *Curr Opin Plant Biol*. 33, 15-22.

Calafiore, R., Ruggieri, V., Raiola, A., Rigano, M. M., Sacco, A., Hassan, M. I., Frusciante, L. & Barone, A. (2016). Exploiting Genomics Resources to Identify Candidate Genes Underlying Antioxidants Content in Tomato Fruit. *Front Plant Sci.* 7, 397.

Calafiore, R., Rigano, M. M., Raiola, A., Aliberti, A. & Barone, A. (2018). Marker – assisted pyramiding of QTLs controlling fruit quality traits in tomato. *Acta Hort*. 1203, 71-77.

Canene-Adams, K., Campbell, J. K., Zaripheh, S., Jeffery, E. H. & Erdman, J. W. (2005). The Tomato as a Functional Food. *J. Nutr*. 135, 1226–1230.

Chitwood, D. H., Maloof, J. N. & Sinha, N. R (2013). Dynamic transcriptomic profiles between tomato and a wild relative reflect distinct developmental architectures. *Plant Physiol.* 162, 537-552.

Choo, A., Crisp, P., Saint, R., O'Keefe, L. R., & Baxter, S. W. (2017). CRISPR/Cas9-mediated mutagenesis of the *white* gene in the tephritid pest *Bactrocera tryoni. Journal of Applied Entomology*. 142, 1-7.

D'Ambrosio, C., Stigliani, A. L. & Giorio, G. (2018). CRISPR/Cas9 editing of carotenoid genes in tomato. *Transgenic Research*. 4, 367-378.

De Oliveira Silva, M. F., Lichtenstein, G., Alseekh, S., Rosado-Souza, L., Conte, M., Fuentes Suguiyama, V., Silvestre Lira, B., Fanourakis, D., Usadel, B., Lopes Bhering, L., DaMatta, F. M., Sulpice, R., Araújo, W. L., Rossi, M., de Setta, N., Fernie, A. R., Carrari, F. & Nunes‐Nesi, A. (2018). The genetic architecture of photosynthesis and plant growth-related traits in tomato. *Plant Cell Environ*. 41, 327–341.

De Vicente, M. C. & Tanksley, S. D. (1993) QTL analysis of transgressive segregation in an interspecific tomato cross. *Genetics*. 134:585–596.

Di Matteo, A., Sacco, A., Anacleria, M., Pezzotti, M., Delledonne, M., Ferrarini, A., Frusciante, L. & Barone, A. (2010). "The ascorbic acid content of tomato fruits is associated with the expression of genes involved in pectin degradation." *BMC Plant Biology*. 10, 163.

Diouf, I.A., Derivot, L., Bitton, F., Pascual, L., & Causse, M. (2018) Water Deficit and Salinity Stress Reveal Many 520 Specific QTL for Plant Growth and Fruit Quality Traits in Tomato. *Frontiers in Plant Science*. 9, 279.

Duan, E., Wang, Y., Liu, L., Zhu, J., Zhong, M., Zhang, H., Li, S., Ding, B., Zhang, X., Guo, X., Jiang, L. & Wan, J. (2016). Pyrophosphate: fructose-6-phosphate 1-phosphotransferase (PFP) regulates carbon metabolism during grain filling in rice. *Plant Cell Rep*. 35, 1321-1331.

Dumas, Y., Dadomo, M., Di Lucca, G. & Grolier, P. (2003). Review: effects of environmental factors and agricultural techniques on antioxidant content of tomatoes. *J. Sci. Food Agric*. 83, 369-382.

Eshed, Y., Abu-Abied, M., Saranga, Y., & Zamir, D. (1992). Lycopersicon esculentum lines containing small overlapping introgressions from L. pennellii. *Theor Appl Genet*. 83, 1027-1034.

Eshed, Y. & Zamir, D. (1994). Introgressions from Lycopersicon pennellii can improve the soluble-solids yield of tomato hybrids. *Theor Appl Genet*. 88, 891-897.

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

Fang, P., Spevak, C. C., Wu, C. & Sachs, M. S. (2004). A nascent polypeptide domain that can regulate translation elongation. *Proc. Natl. Acad. Sci. USA*. 101, 4059–4064.

Fang, T., Zhen, Q., Liao, L., Owiti, A., Zhao, L., Korban, S. S. & Han, Y. (2017). Variation of ascorbic acid concentration in fruits of cultivated and wild apples. *Food Chem*. 225, 132–137.

Fang, Z., Zhiu, D., Ye, X., Jiang, C. & Pan, S. (2016). Identification of candidate anthocyanin-related genes by transcriptomic analysis of "Furongli" plum (*Prunus salicina* Lindl.) during fruit ripening using RNA-seq. *Front. Plant Sci*. 7, 1338.

Fernandez-Moreno, J. P., Levy-Samoha, D., Malitsky, S., Monforte, A. J., Orzaez, D., Aharoni, A. & Granell, A. (2017). Uncovering tomato quantitative trait loci and candidate genes for fruit cuticular lipid composition using the Solanum pennellii introgression line population. *J Exp Bot*. 68, 2703-2716.

Food and Nutrition Board, Institute of Medicine. Dietary reference intakes for vitamin C, vitamin E, selenium and carotenoids. *Washington, DC: National Academy Press*. 2000, 186–283.

Foolad, M. R., Chen, F. Q. & Lin, G. Y. (1998). RFLP mapping of QTLs conferring salt tolerance during germination in an interspecific cross of tomato. *Theoretical and Applied Genetics.* 97**,** 1133–1144.

Friedman, M. (2013) Anticarcinogenic, cardioprotective, and other health benefits of tomato compounds lycopene, α-tomatine, and tomatidine in pure form and in fresh and processed tomatoes. *Journal of Agricultural and Food Chemistry*. 61, 9534−9550.

Frusciante, L., Carli, P., Ercolano, M.R., Pernice, R., Di Matteo, A., Fogliano, V., & Pellegrini, N. (2007). Antioxidant nutritional quality of tomato. *Mol. Nutr. Food Res*. 51, 609-617.

Fulop, D., Ranjan, A., Ofner, I., Covington, M. F., Chitwood, D. H., West D., Ichihashi, Y., Headland, L., Zamir, D., Maloof, J. N. & Sinha, N. R. (2016). A New Advanced Backcross Tomato Population Enables High Resolution Leaf QTL Mapping and Gene Identification. *G3 (Bethesda).* 6, 3169-3184.

Fulton, T., van der Hoeven, R., Eannetta, N., & Tanksley, S. (2002). Identification, Analysis and Utilization of a Conserved Ortholog Set (COS) Markers for Comparative Genomics in Higher Plants. *Plant Cell.* 14, 1457-1467.

Gallie, D.R. (2013). The role of l-ascorbic acid recycling in responding to environmental stress and in promoting plant growth. *Journal of Experimental Botany*, 63, 695–70.

Gur, A. & Zamir, D. (2015). Mendelizing all components of a pyramid of three yield QTL in tomato. *Front Plant Sci.* 6, 1096.

Hrdlickova, R., Toloue, M. & Tian, B. (2017). RNA-Seq methods for transcriptome analysis. *Wiley Interdiscip. Rev. RNA*. *8*, 1.

Hu, L., Li, L., Luo, Z., Yang, J. & Liu, W. (2012). Determination of trace vitamin C by ion-pair HPLC with UV detection in calcium gluconate and vitamin C compound oral solution. *J Chromatogr Sci.* 50, 102-107.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. *337*, 816–821.

Kamenetzky, L., Asís, R., Bassi, S., de Godoy, F., Bermúdez, L., Fernie, A. R., Van Sluys, M., Vrebalov, J., Giovannoni, J. J., Rossi, M. & Carrari, F. (2010). Genomic analysis of wild tomato introgressions determining metabolism- and yield-associated traits. *Plant Physiol.* 152, 1772–1786.

Khalil-Ur-Rehman, M., Sun, L., Li, C., Faheem, M., Wang, W. & Tao, J. (2017). Comparative RNA-seq based transcriptomic analysis of bud dormancy in grape. *BMC Plant Biol*. 17, 18.

Knapp, S., Bohs, L., Nee, M. & Spooner, D.M. (2004). Solanaceae-a model for linking genomics with biodiversity. *Comparative and Functional Genomics*. 5, 285–291.

Knapp, S. & Peralta, I. (2016). The tomato (*Solanum lycopersicum* L. Solanaceae) and its botanical relatives. In "The Tomato Genome", (M Causse, J Giovannoni, M Bouyazen, M Zouine ed.) pp. 7–21, Berlin: Springer.

Koenig, D., Jiménez-Gómeza, J. M., Kimuraa, S., Fulopa, D., Chitwood, D. H., Headlanda, L. R., Kumara, R., Covingtona, M. F., Devisettya, U. K., Tata, A. V., Tohgee, T., Bolgerf, A., Schneebergerb, K., Ossowskib, S., Lanzb, C., Xiongi, G., Taylor-Teeplesa, M., Bradya, S. M., Paulyi, M., Weigelb, D., Usadelf, B., Ferniee, A. R., Pengm, J., Sinhaa, N. R. & Maloofa, J. N. (2013). Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc. Natl Acad. Sci. USA*. 110, E2655–E2662.

Krause, K., Johnsen, H. R., Pielach, A., Lund, L., Fischer K. & Rose, J. K. C. (2018). Identification of tomato introgression lines with enhanced susceptibility or resistance to infection by parasitic giant dodder (Cuscuta reflexa). *Physiol Plant.* 162, 205-218.

Kruminis-Kaszkiel, E., Juranek, J., Maksymowicz, W., & Wojtkiewicz, J. (2018). CRISPR/Cas9 technology as an emerging tool for targeting Amyotrophic Lateral Sclerosis (ALS). *International Journal of Molecular Sciences*. 19, E906.

Laing, W.A., Wright, M.A., Cooney, J. & Bulley, S.M. (2007). The missing step of the L-galactose pathway of ascorbate biosynthesis in plants, an L-galactose guanyltransferase, increases leaf ascorbate content. *Proc. Natl. Acad. Sci. USA*. 104, 9534–9539.

Laing, W.A., Martínez-Sánchez, M., Wright, M. A., Bulley, S.M., Brewster, D., Dare, A. P., Rassam, M., Wang, D., Storey, R., Macknight, R. C. & Hellens, R. P. (2015). An upstream open reading frame is essential for feedback regulation of ascorbate biosynthesis in Arabidopsis. *Plant Cell.* 27, 772–786.

Li, X., Wang, Y., Chen, S., Tian, H., Fu, D., Zhu, B., Luo, Y. & Zhu, H. (2018)-Lycopene is enriched in tomato fruit by CRISPR/Cas9-mediated multiplex genome editing. *Front Plant Sci*. 9, 559.

Li, Y., Yang, K., Yang, W., Chu, L., Chen, C., Zhao, B., Li, Y., Jian, J., Yin, Z., Wang, T. & Wan, P. (2017). Identification of QTL and Qualitative Trait Loci for Agronomic Traits Using SNP Markers in the Adzuki Bean. *Frontiers in Plant Science.* 8, 840.

Liu, Z., Alseekh, S., Brotman, Y., Zheng, Y., Fei, Z., Tieman, D. M., Giovannoni J. J., Fernie, A. R. & Klee, H. J. (2016). Identification of a Solanum *pennellii* Chromosome 4 Fruit Flavor and Nutritional Quality-Associated Metabolite QTL. *Front Plant Sci.* **7**, 1671.

Locato, V., Cimini, S. & Gara, L. D. (2013). Strategies to increase vitamin C in plants: from plant defense perspective to food biofortification. *Frontiers in Plant Science*. 4, .152.

Lu, M., An, H. & Li, L. (2016). Genome survey sequencing for the characterization of the genetic background of *Rosa roxburghii* Tratt and leaf ascorbate metabolism genes. *PLoS One*, 11, e0147530.

Mølhøj, M., Verma, R., & Reiter, W.-D. (2004). The biosynthesis of D-Galacturonate in plants. functional cloning and characterization of a membrane-anchored UDP-D-Glucuronate 4-epimerase from Arabidopsis. *Plant Physiol*. 135, 1221–1230.

Monamodi, E.L., Lungu, D. M. & Fite, G. L. (2013). Analysis of fruit yield and its components in determinate tomato *(Lycopersicon lycopersci)* using correlation and path coefficient. Botswana Journal of Agriculture and Applied Sciences 9:29-40.

Mutschler, M. A., Doerge, R. W., Liu, S.-C., Kuai, J. P., Liedl, B. E. & Shapiro, J. A. (1996). QTL analysis of pest resistance in the wild tomato *Lycopersicon pennellii*: QTLs controlling acylsugar level and composition. *Theoretical and Applied Genetics*. 92, 709–718.

Orzaez, D., Mirabel, S., Wieland, W. H. & Granell. A. (2006). Agroinjection of tomato fruits. A tool for rapid functional analysis of transgenes directly in fruit. *Plant Physiol*. 140, 3–11.

Paran, I., Goldman, I. & Zamir, D. (1997) QTL analysis of morphological traits in a tomato recombinant inbred line population. *Genome*. 40, 242–248.

Raiola, A., Rigano, MM., Calafiore, R., Frusciante, L. & Barone, A. (2014). Enhancing the Health-Promoting Effects of Tomato Fruit for Biofortified Food. *Mediators of Inflammation*. 2014, 139873.

Ranjan, A., Budke, J.M., Rowland, S.D., Chitwood, D.H., Kumar, R., Carriedo, L., Ichihashi, Y., Zumstein, K., Maloof, J.N. & Sinha, N.R. (2016). eQTL regulating transcript levels associated with diverse biological processes in tomato. *Plant Physiol*. 172, 328–340.

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

Rigano, M. M., Raiola, A., Tenore, G. C., Monti, D. M., Del Giudice, R., Frusciante, L. & Barone, A. (2014). Quantitative trait loci pyramiding can improve the nutritional potential of tomato (Solanum *lycopersicum*) fruits. *J Agric Food Chem.* 62, 11519-11527.

Rigano, M. M., Lionetti, V., Raiola, A., Bellicampi, & Barone, A. (2018). Pectic enzymes as potential enhancers of ascorbic acid production through the D-galacturonate pathway in Solanaceae. *Plant Science.* 266, 55-63.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 26, 139-140.

Rosales, M. A., Cervilla, L. M., Sánchez-Rodríguez, E., delMar Rubio-Wilhelmi, M., Blasco, B., Ríos, J. J., Sorian, T., Castilla, N., Romero, L. & Ruiz, J. M. (2011). The effect of environmental conditions on nutritional quality of cherry tomato fruits: evaluation of two experimental Mediterranean greenhouses. *J. Sci. Food Agric*. 91, 152-162.

Ruf, S., Hermann, M., Berger, I.J., Carrer, H. & Bock, R. (2001). Stable genetic transformation of tomato plastids—high-level foreign protein expression in fruits. *Nat. Biotechnol.* 19, 870–875.

Ruggieri, V., Sacco, A., Calafiore, R., Frusciante, L. & Barone, A. (2015) Dissecting a QTL into candidate genes highlighted the key role of pectinesterases in regulating the ascorbic acid content in tomato fruit. *Plant Genome.* 8, 1-10.

Ruggieri, V., Bostan, H., Barone, A., Frusciante, L. & Chiusano, M. L. (2016). Integrated bioinformatics to decipher the ascorbic acid metabolic network in tomato. *Plant Molecular Biology.* 91, 397–412.

Sacco, A., Di Matteo, A., Lombardi, N., Trotta, N., Punzo, B., Mari, A. & Barone, A. (2013). Quantitative trait loci pyramiding for fruit quality traits in tomato. *Molecular Breeding*. 31, 217-222.

Sanseverino, W., Hénaff, E., Vives, C., Pinosio, S., Burgos-Paz, W., Morgante, M., Ramos-Onsins, S. E., Garcia-Mas, J. & Casacuberta, J. M. (2015). Transposon insertions, structural variations, and SNPs contribute to the evolution of the melon genome. *Mol Biol Evol.* 32, 2760–2774.

Scala, V., Grottoli, A., Aiese Cigliano, R., Anzar, I., Beccaccioli, M., Fanelli, C., Dall' Asta, C., Battilari, P., Reverberi M. & Sanseverino, W. (2017). Careful with that Axe, Gene, Genome perturbation after a PEG-mediated protoplast transformation in *Fusarium verticillioides. Toxins.* 9, 183.

Schauer N, Semel Y, Roessner U, Gur, A., Balbo, I., Carrari, F., Pleban, T., Perez-Melis, A., Bruedigam, C., Kopka, J., Willmitzer, L., Zamir, D. & Fernie, A. R. (2006). Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat Biotechnol.* 24, 447–54.

Smirnoff, N., Conklin, P. L., & Loewus, F. A. (2001). Biosynthesis of ascorbic acid in plants: a renaissance. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 52, 437.

Stevens, R., Buret, M., Garchery, C., Carretero, Y. & Causse, M. (2006). Technique for rapid, small-scale analysis of vitamin C levels in fruit and application to a tomato mutant collection. *J Agric Food Chem*. 54, 6159-6165.

Tang, X., Gong, R., Sun, W., Zhang, C. & Yu, S. (2018). Genetic dissection and validation of candidate genes for flag leaf size in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 131, 801–815.

Tieman, D., Zhu, G., Resende Jr., M. F., Lin, T., Nguyen, C., Bies, D., Rambla, J. L., Beltran, K. S., Taylor, M., Zhang, B., Ikeda H., Liu, Z., Fisher, J., Zemach, I., Monforte, A., Zamir, D., Granell, A., Kirst, M., Huang, S & Klee, H. (2017). A chemical genetic roadmap to improved tomato flavor. *Science* 355, 391-394.

Tomlinson, L., Yang, Y., Emenecker, R., Smoker, M., Taylor, J., Perkins, S., Smith, J., MacLean, D., Olszewski, N. E. & Jones, J. D. G. (2018). Using CRISPR/Cas9 genome editing in tomato to create a gibberellin-responsive dominant dwarf DELLA allele. *Plant Biotechnol J*.1-9.

Tran, M.K., Schultz, C. J. & Baumann, U. (2008). Conserved upstream open reading frames in higher plants. *BMC Genomics*. 9, 361.

Ueta, R., Abe, C., Watanabe, T., Sugano, S. S., Ishihara, R., Ezura, H., Osakabe, Y. & Osakabe, K. (2017). Rapid breeding of parthenocarpic tomato plants using CRISPR/Cas9. *Scientific Reports*. 7, 507.

Valderas-Martinez, P., Chiva-Blanch, G., Casas, R., Arranz, S., Martínez-Huélamo, M., Urpi-Sarda, M., Torrado, X., Corella, D., Lamuela-Raventós, R. M. & Estruch, R. (2016). Tomato sauce enriched with olive oil exerts greater effects on cardiovascular disease risk factors than raw tomato and tomato sauce: a randomized trial. *Nutrients.* 8, 170.

Valpuesta, V., & Botella, M. A. (2004). Biosynthesis of L-ascorbic acid in plants: new pathways for an old antioxidant. *Trends Plant Sci.* 9, 573–577.

Van Schalkwyk, A., Wenzl, P., Smit, S., Lopez-Cobollo, R., Kilian, A., Bishop, G., Hefer, C. & Berger, D. K. (2012). Bin mapping of tomato diversity array (DArT) markers to genomic regions of Solanum *lycopersicum* x Solanum *pennellii* introgression lines. Theoretical and Applied Genetics 124: 947–956.

Wang, J., Zhang, Z. & Huang, R. (2013) Regulation of ascorbic acid synthesis in plants. *Plant Signal Behav*. 8, 24536.

Wang, Z., Gerstein, M. & Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Rev. Genet*. 10, 57–63.

Wellburn, A.R. (1994). The Spectral Determination of Chlorophylls a and b, as well as Total Carotenoids, Using Various Solvents with Spectrophotometers of Different Resolution. *Journal of Plant Physiology* 144**,** 307-313.

Wheeler, G. L., Jones, M. A., & Smirnoff, N. (1998). The biosynthetic pathway of vitamin C in higher plants. *Nature*. 393, 365–369.

Wolucka, B. A. & Van Montagu, M. (2003). GDP-Mannose3', 5'-epimerase forms GDP-L-gulose, a putative intermediate forthe novo biosynthesis of vitamin C in plants. *J. Biol. Chem.* 278, 47483–47490.

Yadav, R., Rathi, M., Pednekar, A., & Rewachandani, Y. (2016). A detailed review on Solanaceae family. *European Journal of Pharmaceutical and Medical Research*. 3, 369–378.

Ye, J., Hu, T., Yang, C., Li, H., Yang, M., Ijaz, R., Ye, Z. & Zhang, Y. (2015). Transcriptome profiling of tomato fruit development reveals transcription factors associated with ascorbic acid, carotenoid and flavonoid biosynthesis. *Plos One.* 10, e0130885.

Zamir D. (2001). Improving plant breeding with exotic genetic libraries. *Nature Reviews Genetics.* 2, 983-989.

Zhang, S., Yu, H., Wang, K., Zheng, Z., Liu, L., Xu, M., Jiao, Z., Li, R., Liu, X., Li, J. & Cui, X. (2018). Detection of major loci associated with the variation of 18 important agronomic traits between *Solanum pimpinellifolium* and cultivated tomatoes. *The Plant Journal.* 95, 312–323.

Zouari, I., Salvioli, A., Chialva, M., Novero, M., Miozzi, L., Tenore, G. C., Bagnaresi, P. & Bonfante, P. (2014). From root to fruit: RNA-Seq analysis shows that arbuscular mycorrhizal symbiosis may affect tomato fruit metabolism. *BMC Genomics*. 15, 1-19.

Supplementary data

Figure S3.1 - Borders reliability check of the S. *pennellii* introgression fragment to identify the exact upper IL12-4 introgression border.

Figure S3.2 - Borders reliability check of the S. *pennellii* introgression fragment to identify the exact upper IL7-3 introgression border.

Figure S3.3 - Borders reliability check of the S. *pennellii* introgression fragment to identify the exact lower IL7-3 introgression border.

Figure S3.4 - Summary of the quality of the sequenced bases before trimming.

Figure S3.5 - Summary of the quality of the sequenced bases after trimming.

Figure S3.6 - Graphic representation of the HTSFilter analysis results of RNA-seq B27vs M82 for the year 2016 and 2017.

Figure S3.7 - Graphic representation of the HTSFilter analysis results of RNA-seq R182vs M82 BR and MR.

Table S4.1 - List of primers used in this study.

Figure S4.1 - Solyc12g098480 gene structure.

Figure S4.2 - S. *lycopersicum* and S. *pennellii* epimerases CDS alignment.

Figure S4.3 - S. *lycopersicum* and S. *pennellii* epimerase genes alignment.

Figure S4.4 - *S. pennellii* epimerases CDS alignment.

Figure S4.5 – Sol Genomics and NCBI S. *pennellii* epimerases alignment.

Figure S4.6 - NCBI S. *pennellii* epimerases pcr amplification.

Figure S4.7 - Solyc06g073320_cGGP1 uORF and gRNAs

Figure S4.8- Solyc02g091510_cGGP2 uORF and gRNAs

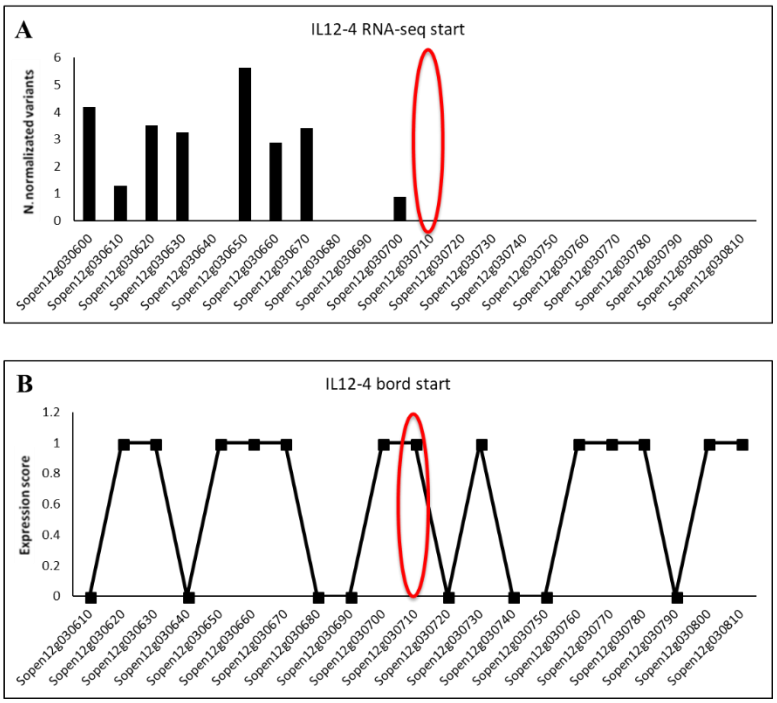


Figure S3.1 - Borders reliability check of the S. *pennellii* introgression fragment to identify the exact upper IL12-4 introgression border.

The red cycle shows the first gene added from S. *pennellii*. A) Number of normalized variations for each gene found using the IL check pipeline; B) Expression score of each gene. The value 1 indicates that the genes were expressed whereas the value 0 genes that they were not expressed, considering a threshold expression value of 0.05.

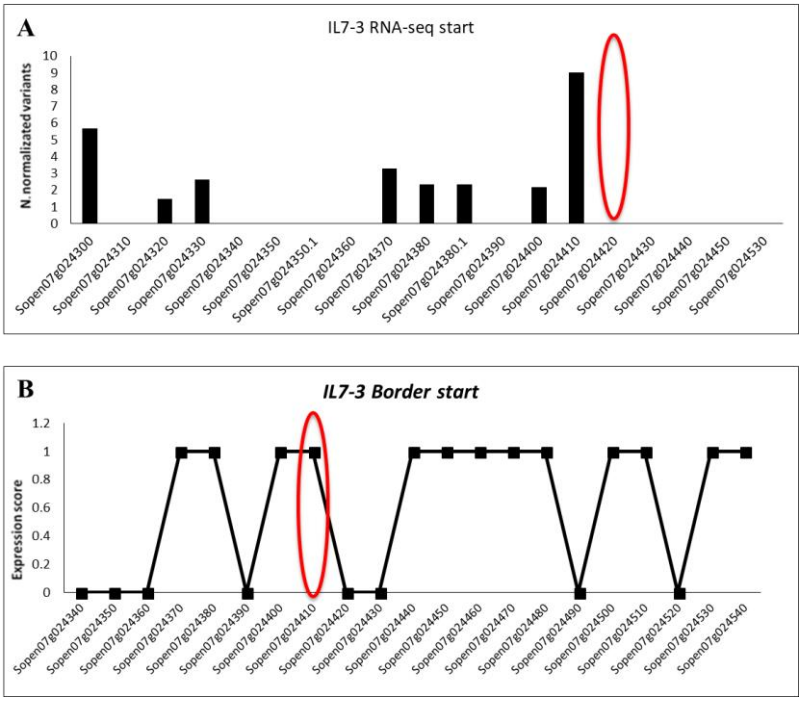


Figure S3.2 - Borders reliability check of the S. *pennellii* introgression fragment to identify the exact upper IL7-3 introgression border.

The red cycle shows the first gene added from S. *pennellii*. A) Number of normalized variations for each gene found using the IL check pipeline; B) Expression score of each gene. The value 1 indicates that the genes were expressed whereas the value 0 genes that they were not expressed, considering a threshold expression value of 0.05.

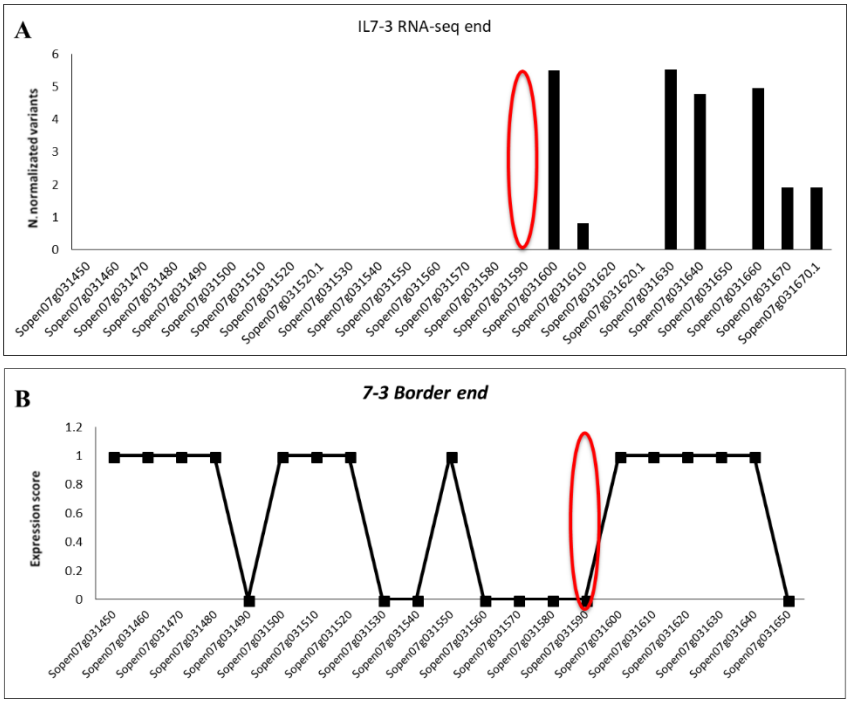


Figure S3.3 - Borders reliability check of the S. *pennellii* introgression fragment to identify the exact lower IL7-3 introgression border.

The red cycle shows the last gene added from S. *pennellii*. A) Number of normalized variations for each gene found using the IL check pipeline; B) Expression score of each gene. The value 1 indicates that the genes were expressed whereas the value 0 genes that they were not expressed, considering a threshold expression value of 0.05.

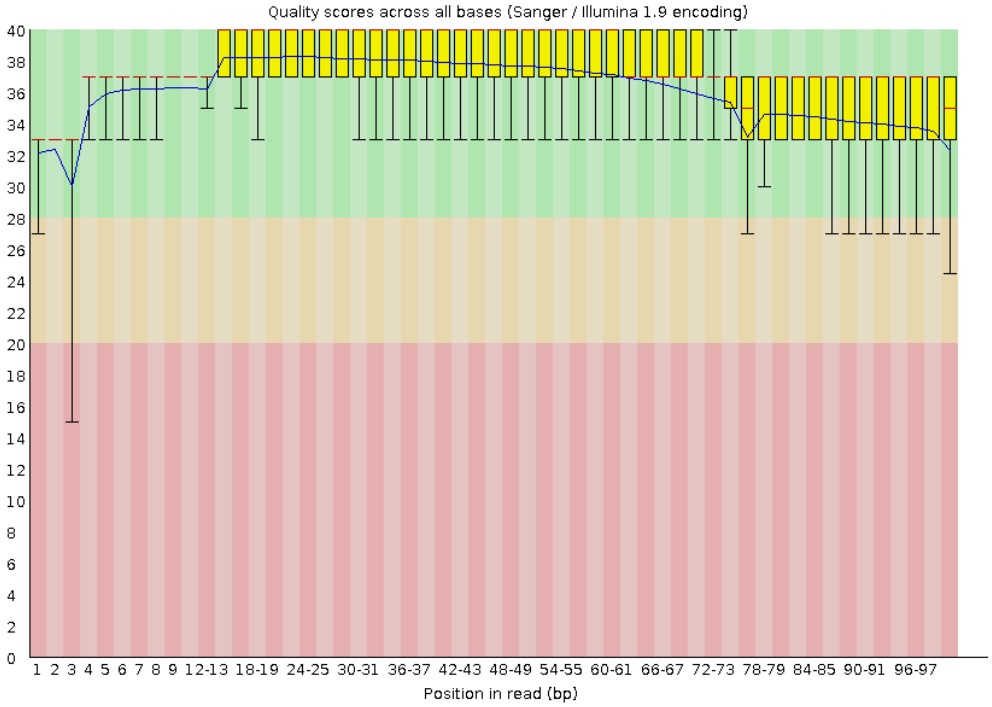


Figure S3.4- Summary of the quality of the sequenced bases before trimming. The x-axis reports the position along the reads, the y-axis reports the quality as Phred-Like score. The yellow boxes represent the interquartile range of the quality values at each position, the red bar is the median, the black line is the average quality value. The whiskers of the boxes represent the 10th and the 90th percentile. The scores in the green part of the chart are considered good.

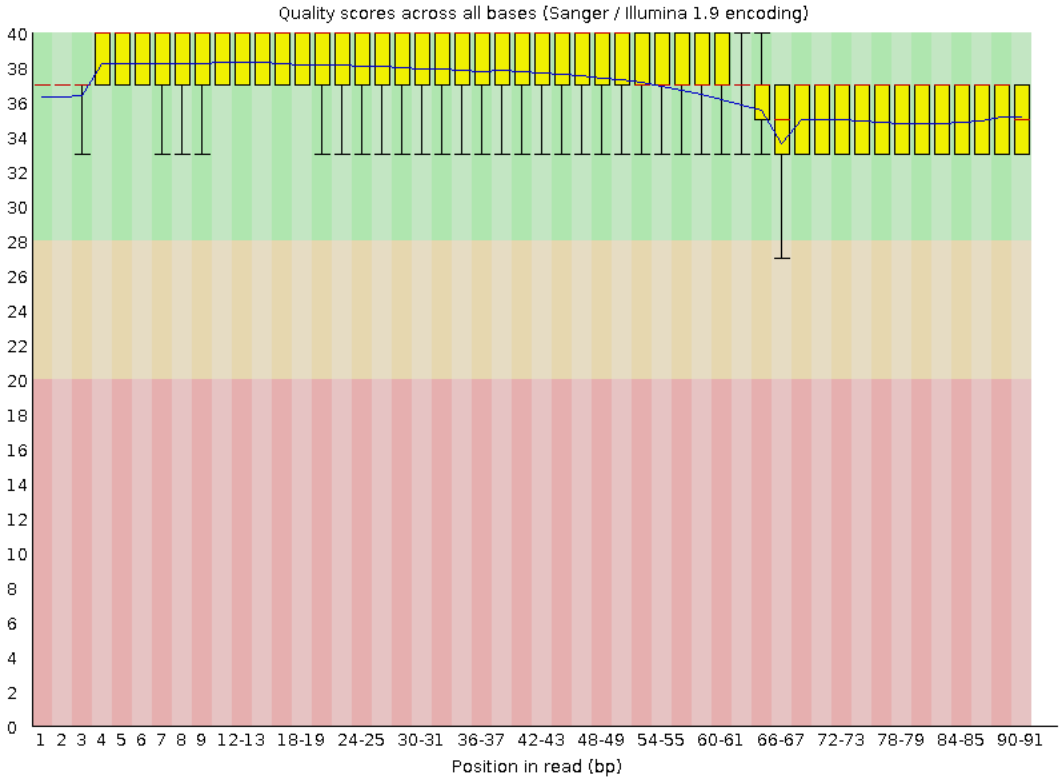


Figure S3.5- Summary of the quality of the sequenced bases after trimming. The x-axis reports the position along the reads, the y-axis reports the quality as Phred-Like score. The yellow boxes represent the interquartile range of the quality values at each position, the red bar is the median, the black line is the average quality value. The whiskers of the boxes represent the 10th and the 90th percentile. The scores in the green part of the chart are considered good.

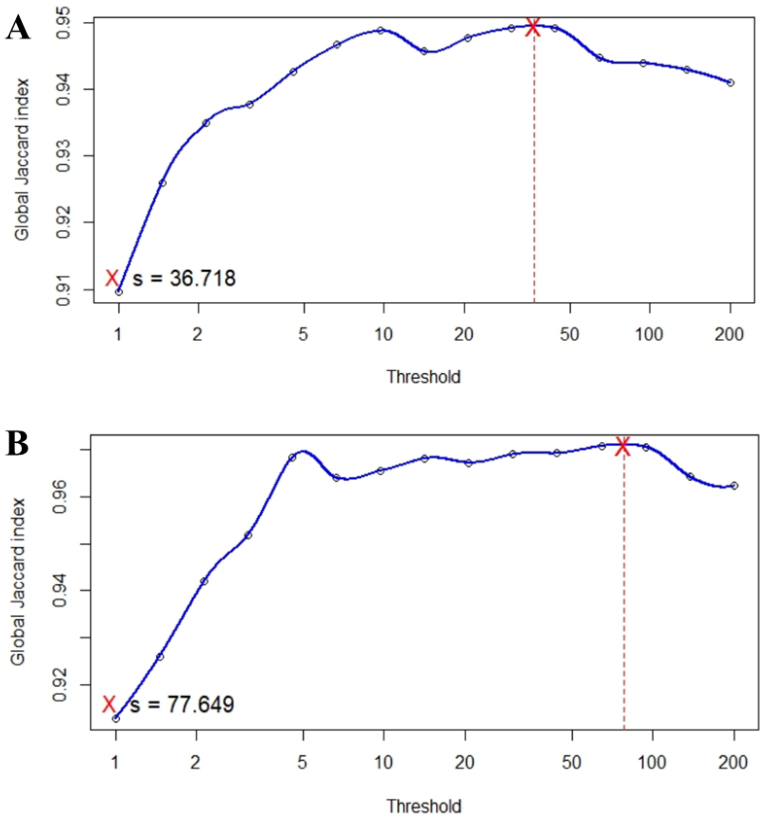


Figure S3.6- Graphic representation of the HTSFilter analysis results of RNA-seq B27vs M82 for the year 2016 (A) and 2017 (B).

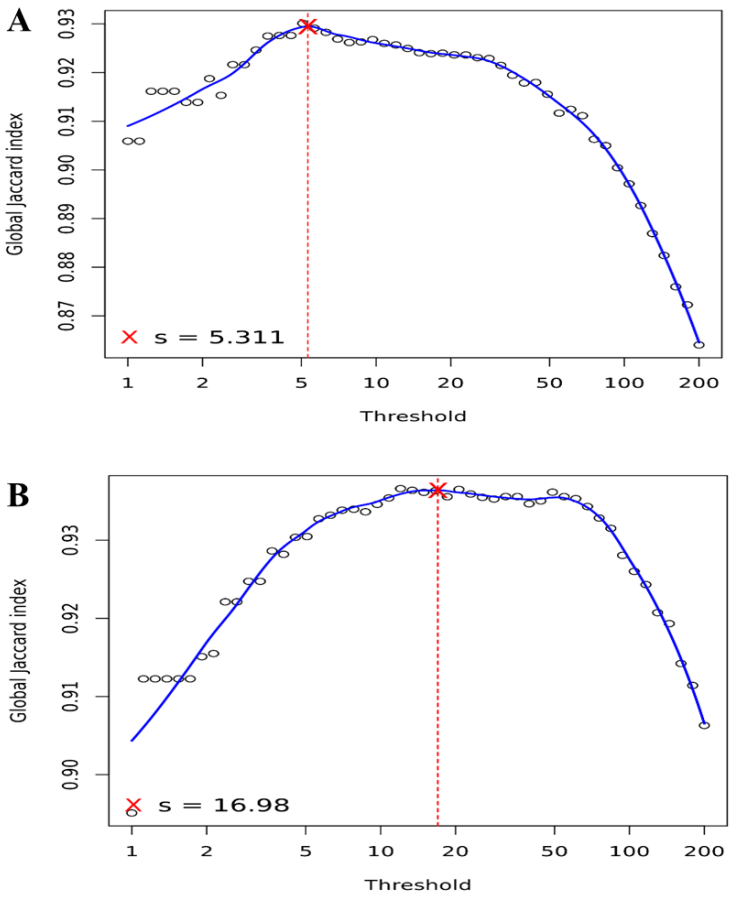


Figure S3.7- Graphic representation of the HTSFilter analysis results of RNA-seq R182vs M82 BR (A) and MR (B).

Table S4.1 - List of primers used in this study.

Gene	Forward primer (5' to 3')	Reverse primer (5' to 3')	Used for
	GGGGACA AGT TTG TAC AAA AAA GCA GGC TAA GGA GAT ATA ACA	GGGGAC CAC TTT GTA CAA GAA AGC TGG GTC	ATTB adapters
Solyc12g098480	ATGGCATCATTTCCAATTGATACAAG	TTAATCCACTTGACAAATCTGTGGTTG	Cloning in pDONR207
XM_015204087/ XM_015204086	ATGGCATCATTTCCAATTGATACAAG	TCAGTCTTGAGAATGGTCATTAGAAGAG TCTAAA	
XM_015204087/ XM_015204086	ATGGCATCATTTCCAATTGATACAAG	TCAGTCTTGAGAATGGTCATTAGAAGAG TCTAAA	PCR analysis
Sopen12g033320/ Sopen12g033260	ATCGTTAAGGGGTGTCTCGG	ACATCATGTGGCCTATCTTTACT	
GGP1/Solyc06g073320	AGGAAAAGGGGTTATTGTATTTTGC	GCTTTGAACCTGGAAGGCAGCAATTCC	
GGP2/Solyc02g091510	AACATCTTCCTCAAAATTATCGATG	CACAACCCATGACGTTACC	
pDONR207	TCGCGTTAACGCTAGCATGGATCTC	GTAACATCAGAGATTTTGAGACAG	Sequencing
PJAM2288	AGGGACTAAACTTAGAAGAGAAG	GAGACTGGTGATTTCAGCGTACCGA	
GGP1_Solyc06g073320	AGGAAAAGGGGTTATTGTATTTTGC	GCTTTGAACCTGGAAGGCAGCAATTCC	
GGP2_Solyc02g091510	AAGGATCGCTCCTCAGTAATTG	CACAACCCATGACGTTACC	

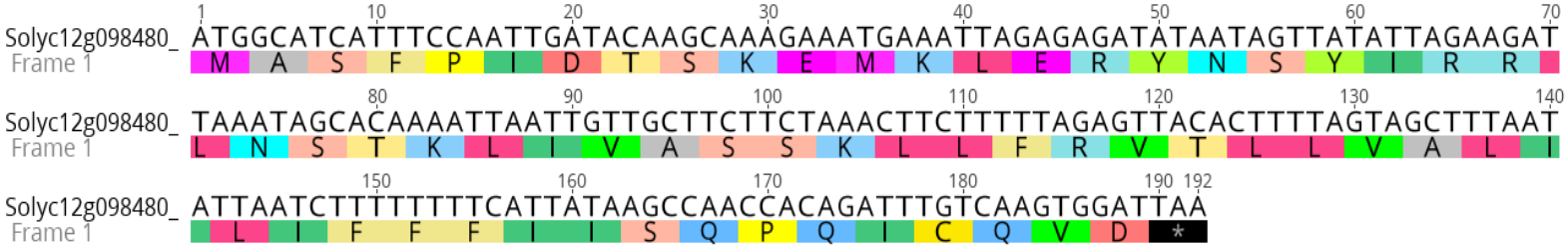


Figure S4.1- Solyc12g098480 gene structure.

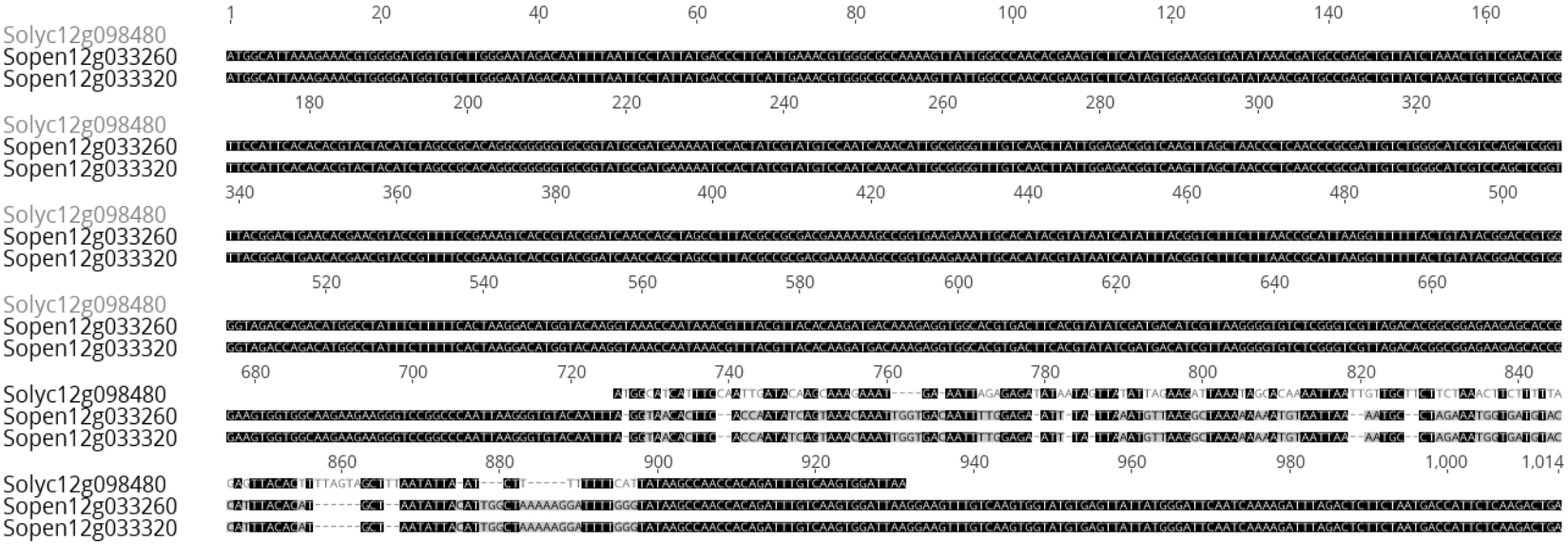


Figure S4.2- S. *lycopersicum* and S. *pennellii* epimerases CDS alignment.

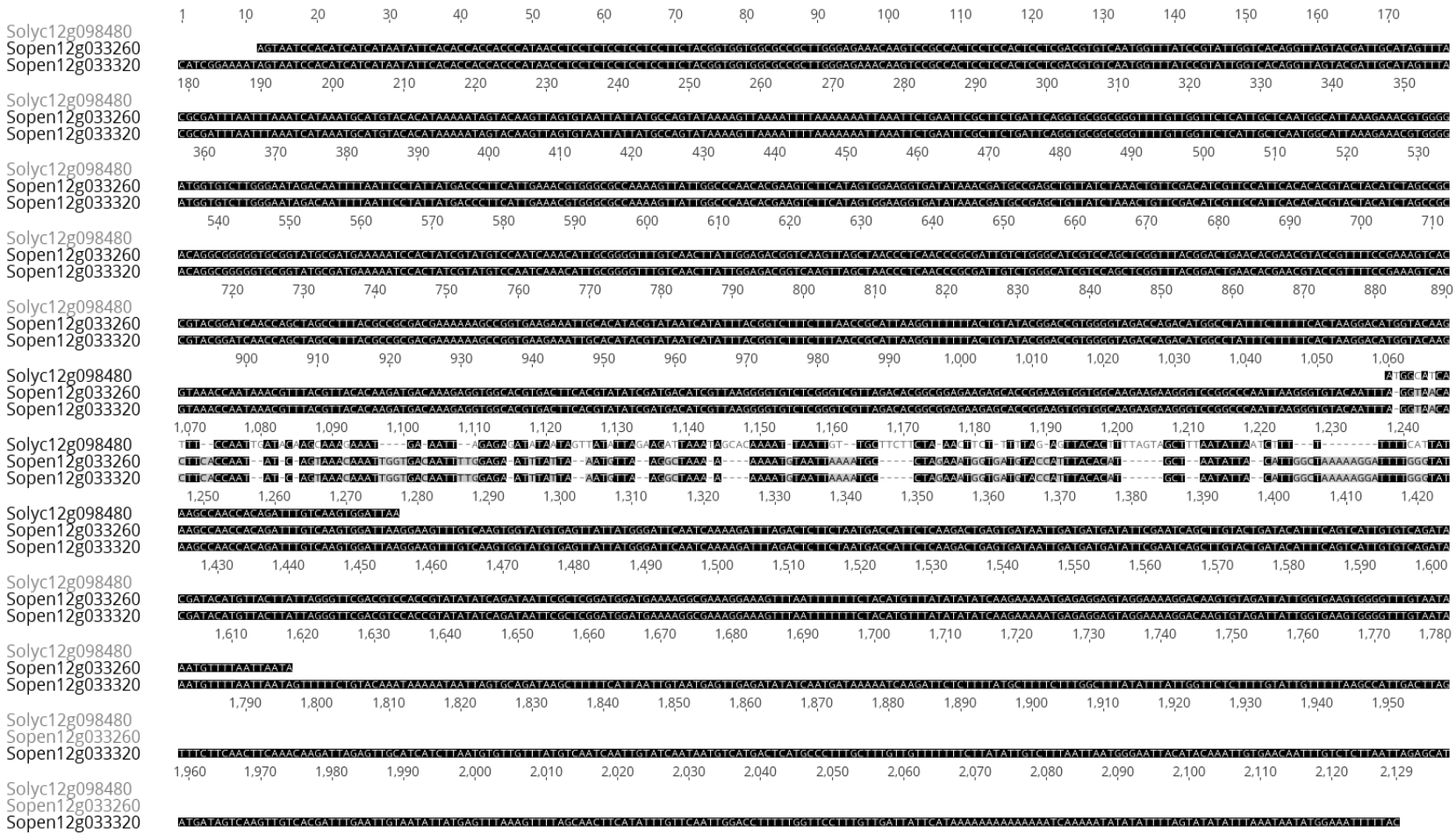


Figure S4.3 - S. *lycopersicum* and S. *pennellii* epimerase genes alignment

```
Sopen07g002320.1      ATGCCTTCGTTAGAGGAG---------------GAATTGTTTCCGTCTACACCGGGAAAA
Sopen12g005550.1      ATGAGAGGATTAGAAGAA---------------GAATTTATTCCATCAACACCAGGAAAA
Sopen12g033260.1      ATG-----------------------------------------------------------
Sopen12g033320.1      ATG-----------------------------------------------------------
Sopen05g032260.1      ATGGCTTCTCT-------------------------TTTTTCGATCGATAC--GAAAAA
Sopen09g035230.1      ATGAAAACTATGATGG--------------------------CATCGCCACCTGACACA
Sopen03g023360.1      ATG---------------------------------------GACAAGCACCGGAAATG
Sopen08g027810.1      ATGACTCAATTGAAGCCCATTCTTACGCATTTGGATTCCATTCCTTCCACCCCAGGAAAG
Sopen10g007490.1      ATGTCCTCAATGAAGCACGTT------------GACAATATTCCATCAACCCCAGGAAAG
Sopen01g036950.1      ATGTCCCAAATGATGCACATT------------GACAATATCCCATCAACCCCAGGAAAG
Sopen05g029460.1      ATGTCCCAAATGAAGCACATT------------GATAATACTCCATCAACCCCAGGAAAA
                      ***

Sopen07g002320.1      TTCAAGGATAGA---------------AATCGACATTTTCATCGA---------------
Sopen12g005550.1      TTCAAAGACAAAACGTATTACTCCGGCAACCGGCAATTCCACCGG---------------
Sopen12g033260.1      -----------------------------------------------------------
Sopen12g033320.1      -----------------------------------------------------------
Sopen05g032260.1      ATCAACAAAATTAGATC--------GTAATACTTATATTCATAAAGTTAACAGTAGCACA
Sopen09g035230.1      AGCAAAACCACAAAACTAGAGCGCTACAATAGCTATATTCGAAGAGTTAATAGTACAAAA
Sopen03g023360.1      GA------------------------------------------------------------
Sopen08g027810.1      TTTAAGCCTGATAAATCTTCACCCTATAATCTTTATCGTCTCCGTTTT------------
Sopen10g007490.1      TTCAAGGAAAAA------------TCCAATTACAATAGGCTTAGGCTT------------
Sopen01g036950.1      TTCAAGATGGAAAAGTCT---CCTTACA------ATAGGCTAAGGATG------------
Sopen05g029460.1      TTCAAGATTGAAAAATCT---CCTTATA------ATAGGCTTAGGCTA------------

Sopen07g002320.1      TGTTTTGCATCTACAAGCACCA-------TGTTTCTATGGGCATTGTTTTTAATAGCT--
Sopen12g005550.1      TGTTTCGCTTCAACAAGTACAA-------TGTTTTTATGGGCATTGTTTTTATTAGCT--
Sopen12g033260.1      -----------------------------------------------------------
Sopen12g033320.1      -----------------------------------------------------------
Sopen05g032260.1      AAACTTATCGTAGCTTCTATTAAACTTCTTTATTTTATCGCCATTTTAATTTCTGCGTTA
Sopen09g035230.1      CTCATCGCTGCATCGTCTAAGC-------TTTTATTTCGTGTCACTTTATTAGTGGCTCT
Sopen03g023360.1      CGTATTCCATTA----CCAAGC-------TTGTATTCTGGGCAACCATTTTTGCGGGT--
Sopen08g027810.1      CACCCTACTCTGTTTCCCAGGT-------TTACTCTTTGGTCCTTTTTTTTCATCTTT--
Sopen10g007490.1      CATTTTTCTGTAG---TCAAGC-------TTACATTTTGGTCATTTGTGTTCTTGGGT--
Sopen01g036950.1      CATTTTTCTCTAG---CAAAGC-------TCACATTTTGGTCATTTGTATTCTTGGGG--
Sopen05g029460.1      CAATTTTCTTTAG---CCAAGA-------TCATTTTTTGGTCACTTGTTTTTGTGGGG--

Sopen07g002320.1      TTAACAGCTTCATATTTGTGTTTTCAGTCGTTTATGGAT-TCCGGTAACCGGTATTTCAC
Sopen12g005550.1      TTAACGGCGTCGTATTTGAGTTTTCAATCCTTCATAAAC-TCCGGTACCCGGTATCTCTC
Sopen12g033260.1      -----------------------------------------------------------
Sopen12g033320.1      -----------------------------------------------------------
Sopen05g032260.1      TTGATCGGCTCATTCATATTAAATCACTTTTTATTCATT-TCTACCGACTCAAATC----
Sopen09g035230.1      TTTACTTATTTTCTTCTTCACTATAAATTACCCTCCGTTAACTTCCGAAAAATCATTTAA
Sopen03g023360.1      ---GCCTTTCTTTTCTTCTGTTTGCGTTCTTC--------TCCTCCGTCTCATTCTACAA
Sopen08g027810.1      TTCATCGTTTTGCTCATCTTTTTCTCATCTCCGTCCAAC-CCCACCGCCGGAAACAGCCG
Sopen10g007490.1      TTGATCTTTATATGTTT----TTTTAAATCAACATCTTC-ATCTTCATCCCCTGTATCTT
Sopen01g036950.1      TTGATCTTTGTATTCTT----CTACAGATCT---CCAGC-TTCTTCATCCCCTGTTTCTT
Sopen05g029460.1      TTAATCTTTGTATTCTT----TTACAGATCA---CCATC-ATCTTCCTCTCATGTTTCTT

Sopen07g002320.1      CTCTAC----------------------------------------------ATGGGGT-GG
Sopen12g005550.1      CTCCAC----------------------------------------------ATGGGGT-AG
Sopen12g033260.1      -----------------------------------------------------------
Sopen12g033320.1      -----------------------------------------------------------
Sopen05g032260.1      --------------------------AACACGGCCGCCGCCTGTTCTCCTTTGGTGGC-GG
Sopen09g035230.1      CAATAATATTCACACCACTACCCATAACC--TTCTATCCTCCGCGATTTACGGCGGT-GG
Sopen03g023360.1      CTATGAAATGGGGTAGTAG---------------------------------TAGTAAT-AC
Sopen08g027810.1      CCGAAGTCTGAAAAA------------------CTCTCTTTCGCCATCTCCCGCGCTCGG
Sopen10g007490.1      CAGATCTCTCAAGAAGATCTCTTAGAACAAGTTC----------------TTATGAT-GG
Sopen01g036950.1      CAGATCTCTCAAGAAGATCTCTCAGAACCAGCTC----------------CTACGGT-GG
Sopen05g029460.1      CAGATCTCTTCAGGAGATCTCTTAGAACAAGCTC----------------TTATGGT-GG

Sopen07g002320.1      TCATCATTGGGAGAGAAATGTTATTAATTCCGCTCAGATCCGCCGCTCTCACGGCGGAAT
Sopen12g005550.1      TCTTCATTGGGAGAAACAAGTTCGTGATTCCGCCCAAATCCACCGTGTTAACGGC---AT
Sopen12g033260.1      -----------------------------------------------------------
Sopen12g033320.1      -----------------------------------------------------------
Sopen05g032260.1      CGCGGGTTGGGAGAAAAAAGTCCGGCATTCTTCTATACCTAAGAAAATCAACGGG---AA
Sopen09g035230.1      GGCGTCGTGGGAGAAACAAGTTCGTCACTCCTCGACACCCCACCGCCCTAACGGG---CT
Sopen03g023360.1      CCCACAATGGGAAAAGCGTGTCAGATTATCTGCTCGTTCACGTTCCGGTCAT------TT
Sopen08g027810.1      CCCGAACTGGGAACGTCGGGTTCGGGCCTCGGCCCGACCAAGGTCTAAAACGGGC---TT
Sopen10g007490.1      TCCAGCTTGGGAGAAAAGGATTAAAGCCTCAGCAAAAATCAGGTCAACAAATGGT---AT
Sopen01g036950.1      CCCTGCTTGGGAAAAAAGGATTAAAGCTTCAGCTAAAGTAAGGTCACGTAATGGT---AT
Sopen05g029460.1      TCCAGCTTGGGAGAAAAAGATTAAGGCTTCAACAAAACCAAGGTCAAGTAATGGT---AT

Sopen07g002320.1      GTCGGTGTTAGTTACCGGCGCAGCCGGTTTCGTCGGTTCCCACGTCTCCCTCGCCTTGAA
Sopen12g005550.1      GTCCGTACTCGTCACCGGTGCCGCCGGTTTCGTCGGATCTCACGTTTCAATCGCATTGAA
Sopen12g033260.1      ----------------------------------------------------GCATTAAA
Sopen12g033320.1      ----------------------------------------------------GCATTAAA
Sopen05g032260.1      TGTAGTGTTGGTGACCGGTGCAGCTGGATTTGTTGGTTCTCATTGTGGATTAGCGTTGAA
Sopen09g035230.1      ATCCGTCCTCGTGACGGGAGCAGCTGGGTTCGTTGGTTCTCATTGCTCTTTAGCATTGAA
Sopen03g023360.1      ATCCGTTCTTGTTACCGGCGCCGCCGGGTTCGTTGGGTCTCACGTCTCCGCCGCACTCAA
Sopen08g027810.1      TACCGTTCTAGTCACTGGTGCTGCCGGTTTTGTGGGGACCCATGTGTCTCTAGGCCTGAA
Sopen10g007490.1      TTCTGTTTTGGTTACTGGTGCTGCTGGCTTTGTAGGTACACATGTTTCATCTGCGTTGAA
Sopen01g036950.1      TTCTGTATTGGTTACTGGTGCTGCTGGCTTTGTAGGAACTCATGTTTCAGTTGCTCTTAA
Sopen05g029460.1      TTGTGTGTTGGTAACTGGTGCAGCTGGTTTTGTAGGAACACATGTATCAGCTGCTTTAAA
                                                                    *   * **
```

```
Sopen07g002320.1      AAAACGAGGCGACGGAGTTGTAGGAATCGATAATTTCAACAATTACTACGATCCGTCGTT
Sopen12g005550.1      AAAACGTGGAGACGGTGTCGTCGGAATTGATAATTTCAACAATTACTATGACCCTTCGTT
Sopen12g033260.1      GAAACGTGGGGATGGTGTCTTGGGAATAGACAATTTTAATTCCTATTATGACCCTTCATT
Sopen12g033320.1      GAAACGTGGGGATGGTGTCTTGGGAATAGACAATTTTAATTCCTATTATGACCCTTCATT
Sopen05g032260.1      GAAACGCGGAGATGGAGTTATTGGAATTGATAATTTCAATTCGTATTATGATCCATCGTT
Sopen09g035230.1      GAAGCGTGGTGATGGTGTTTTAGGTTTAGATAATTTCAATTCATACTACGATCCTTCGTT
Sopen03g023360.1      ACGCCGTGGTGATGGGGTTGTTGGATTGGATAATTTCAACAGCTATTATGACCCATCGTT
Sopen08g027810.1      ACGCCGTGGAGATGGCGTTTTGGGGCTGGATAATTTTAATCAGTATTATGATGTCGGGCT
Sopen10g007490.1      ACGACGCGGTGATGGCGTTGTGGGGTTAGACAATTTCAATGATTATTATGATCCATCATT
Sopen01g036950.1      ACGCCGTGGCGATGGCGTATTGGGTTTGGATAATTTCAATGATTATTATGACCCTTCGCT
Sopen05g029460.1      ACGCCGCGGTGATGGCGTGTTGGGGTTGGATAATTTCAATGATTATTATGATCCCTCGCT
                          ** ** ** ** **  * **  * ** ***** **    ** ** **       *

Sopen07g002320.1      GAAAAAAGCTCGAAAGAATCTACTCACCCAGAATTCGATTTTCATCGTCGATGGTGATAT
Sopen12g005550.1      GAAAAAAGCTCGAAAAAATCTGTTGAATTTGCAAAATGTGTACCTCATTGAAGGAGATAT
Sopen12g033260.1      GAAACGTGGGCGCCAAAAGTTATTGGCCCAACACGAAGTCTTCATAGTGGAAGGTGATAT
Sopen12g033320.1      GAAACGTGGGCGCCAAAAGTTATTGGCCCAACACGAAGTCTTCATAGTGGAAGGTGATAT
Sopen05g032260.1      GAAACGTGCTCGTCAGGAATTAATGTCGAAACATGAGATTTTTATCGTCGATGGTGATAT
Sopen09g035230.1      GAAACGTGCTCGTCAAGATCAGTTATCGAAGCATCAGATTTTCATTGTAGAAGGTGATAT
Sopen03g023360.1      GAAAAGGGCCCGTCAGAAATTGTTGGAACAGAAAGGGGTTTTCGTCATGGAAGGTGATAT
Sopen08g027810.1      GAAAAAAGCCCGACAGAGCCTTCTTGAGCGTTCTGGGATTATGGTAGTTGAGGGTGATAT
Sopen10g007490.1      GAAAAGAGCCCGGCAAACTCTGTTAGAGAGTGCCGGGGTCTACATTGTAGAGGGTGACAT
Sopen01g036950.1      CAAAAGAGCACGACAAGCGCTCTTAGAGCGAACAGGGGTGTATGTTGTTGAGGGTGATAT
Sopen05g029460.1      CAAAAGAGCAAGGCAAGAGCTACTAGAGCGCTCTGGGGTGTACATTGTTGAGGCTGACAT
                       ***   *   *  *        *               *     *  * ** *  ** **

Sopen07g002320.1      CAACGATTCAAAATTGTTGGTAAAGTTGTTCGACATTGCTCGATTTACACATGTAATGCA
Sopen12g005550.1      TAACGATGTGCATCTGATATCAAAGCTATTCGATATCGTAGCGTTTACGCATGTAATGCA
Sopen12g033260.1      AAACGATGCCGAGCTGTTATCTAAACTGTTCGACATCGTTCCATTCACACACGTACTACA
Sopen12g033320.1      AAACGATGCCGAGCTGTTATCTAAACTGTTCGACATCGTTCCATTCACACACGTACTACA
Sopen05g032260.1      CAATGATATTGATTTGGTTAATAAATTGTTCGATATTGTTCGGTTTACTCATGTTCTTCA
Sopen09g035230.1      TAACGATACAGAGCTTCTGAAAAAGCTTTTCGACATTGTTCCTTTTACTCATATCCTTCA
Sopen03g023360.1      CAATGATGAAAAGCTTCTGAAAAAGCTTTTCGATATTGTTGAATTCACCCACGTAATGCA
Sopen08g027810.1      TAATGATGCTGTTTTGCTGAGGAAGCTGTTTGATGCTGTTGCTTTTACACATGTCATGCA
Sopen10g007490.1      CAATGATGTCGCCCTCTTAAAGAAATTATTCGATATTGTGCAATTTAGTCATGTTATGCA
Sopen01g036950.1      CAATGATGCCACCCTCTTGAAGAAACTTTTTGATATTGTTCCATTTACTCATGTAATGCA
Sopen05g029460.1      CAATGATGCCACCCTCTTGAAGAAACTTTTTGAAATTGTTGCATTTACTCATGTTATGCA
                       ** ***       *  *    **  * ** **    *     ** *  **  *  * **

Sopen07g002320.1      TCTGGCTGCTCAAGCTGGAGTTCGTTACGCCATGGAAAATCCAGGTTCGTATATACACAG
Sopen12g005550.1      TTTAGCTGCACAGGCCGGTGTTCGTTACGCCATGGAAAATCCCAAATCTTATGTTCATAG
Sopen12g033260.1      TCTAGCCGCACAGGCGGGGGTGCGGTATGCGATGAAAAATCCACTATCGTATGTCCAATC
Sopen12g033320.1      TCTAGCCGCACAGGCGGGGGTGCGGTATGCGATGAAAAATCCACTATCGTATGTCCAATC
Sopen05g032260.1      TTTAGCTGCACAGGCCGGCGTTCGTTACGCAATGGAGAATCCTCTTTCGTATGTACATTC
Sopen09g035230.1      TCTAGCTGCACAAGCAGGTGTTCGTTACGCGATGCAGAATCCTCTCTCTTATGTGAACTC
Sopen03g023360.1      CTTAGCTGCACAGGCCGGCGTTCGTTACGCAATGAAGAATCCTGGTTCTTACATCCATAG
Sopen08g027810.1      TATGGCAGCTCAAGCTGGAGTTAGATATGCAATGCAGAATCCAGGTTCTTATGTTCATAG
Sopen10g007490.1      TTTGGCTGCACAAGCTGGTGTTAGATATGCTATGGAAAATCCTAGCTCATATGTGCATAG
Sopen01g036950.1      TTTAGCTGCACAAGCAGGTGTGCGTTATGCCATGGAAAATCCTGGATCATATGTGCATAG
Sopen05g029460.1      TTTGGCTGCACAAGCCGGTGTGCGATATGCTATGGAAAATCCTAGCTCATATGTGCATAG
                        * ** ** ** ** ** **  * ** ** *** * *****    ** **  *  *

Sopen07g002320.1      TAACATCGCCGGCCTTGTTTCCCTGCTCGAAGTTTGTAAAAATTCCAACCCACAACCCGC
Sopen12g005550.1      TAATATCGCTGGGCTTGTTACTCTTCTCGAAGCTTGTAAAAATGCAAATCCTCAACCCGC
Sopen12g033260.1      AAACATTGCGGGGTTTGTCAACTTATTGGAGACGGTCAAGTTAGCTAACCCTCAACCCGC
Sopen12g033320.1      AAACATTGCGGGGTTTGTCAACTTATTGGAGACGGTCAAGTTAGCTAACCCTCAACCCGC
Sopen05g032260.1      AAATGTGAATGGATTTGTTAATTTGTTAGAAATTGCTAAATCTGCTAATCCACAACCCGC
Sopen09g035230.1      AAACGTAGCTGGGTTTGTAAATCTGTTAGAAATTGCTAAAGCTGCAGATCCACAACCGGC
Sopen03g023360.1      CAACATTGCAGGTCTCGTTACCCTTTTCGAGGCCTGCAAATCCGCTAACCCACAACCTTC
Sopen08g027810.1      TAATATTGCTGGTTTTGTTAGTTTGCTTGAAGCATGTAAATTGGCTAATCCACAACCTAG
Sopen10g007490.1      TAACATTGCTGGTTTAGTTAATCTTCTTGAATTTTGCAAAAATGCTAATCCTCAACCTGC
Sopen01g036950.1      TAACATTGCTGGTCTTGTTAATGTTCTTGAAATTTGTAAAAGTGTTAATCCTCAACCTGC
Sopen05g029460.1      TAATATTGCTGGACTTGTTAATATGCTTGAGGTTTGCAAAAGTGTTAATCCTCAACCTTC
                       **  *    **  * **     *  * **       **        * ** *****

Sopen07g002320.1      CATTGTTTGGGCGAGTTCAAGTTCTGTTTACGGGTTAAACGAAAAAGTACCGTTTTCCGA
Sopen12g005550.1      CATTGTTTGGGCAAGCTCGAGCTCAGTTTACGGGTTAAACGAAAAGGTACCGTTTTCTGA
Sopen12g033260.1      GATTGTCTGGGCATCGTCCAGCTCGGTTTACGGACTGAACACGAACGTACCGTTTTCCGA
Sopen12g033320.1      GATTGTCTGGGCATCGTCCAGCTCGGTTTACGGACTGAACACGAACGTACCGTTTTCCGA
Sopen05g032260.1      CATTGTTTGGGCTTCTTCAAGCTCTGTTTATGGGTTAAACACGAAGGTACCTTTTTCTGA
Sopen09g035230.1      GATAGTCTGGGCTTCATCGAGCTCTGTTTACGGATTGAACACCAAAGTTCCTTTCTCCGA
Sopen03g023360.1      AATTGTTTGGGCATCATCCAGTTCTGTTTATGGGCTCAATTCCAAAGTACCCTTCTCAGA
Sopen08g027810.1      TATTGTTTGGGCTTCATCAAGTTCTGTTTATGGATTGAATTCAAAAGTACCCTTTTCAGA
Sopen10g007490.1      TATTGTATGGGCATCATCAAGTTCTGTTTATGGATTGAATACAAAGGTTCCGTTTTCGGA
Sopen01g036950.1      TATTGTGTGGGCATCATCTAGTTCTGTATATGGATTGAATACTAAGGTACCTTTTTCAGA
Sopen05g029460.1      TATTGTGTGGGCGTCGTCTAGTTCTGTATATGGATTGAATACTAAGGTACCCTTTTCAGA
                       ** ** *****    ** ** ** ** ** **  * **    ** ** ** ** ** **

Sopen07g002320.1      ATCGGATCGGACTGATCAACCCGCTTCATTGTATGCCGCTACAAAAAAAGCAGGTGAAGA
Sopen12g005550.1      ATCGGATCGAACGGATCAACCCGCTTCGTTATATGCAGCAACGAAGAAAGCTGGTGAAGA
Sopen12g033260.1      AAGTCACCGTACGGATCAACCAGCTAGCCTTTACGCCCGCGACGAAAAAAGCCGGTGAAGA
Sopen12g033320.1      AAGTCACCGTACGGATCAACCAGCTAGCCTTTACGCCCGCGACGAAAAAAGCCGGTGAAGA
Sopen05g032260.1      AAATGACAGGACTGATCAACCAGCTAGCCTCTACGCTGCAACGAAAAAAGCCGGTGAAGA
Sopen09g035230.1      AGATCACAGAACAGATCAACCAGCTAGTTTATACGCTGCAACGAAGAAAGCAGGGGAAGC
Sopen03g023360.1      AAAAGATCAAACGGATCAACCAGCCAGTCTATACGCTGCCACAAAGAAAGCCGGTGAAGA
Sopen08g027810.1      AAAAGATAGAACTGATCAGCCTGCTAGTTTGTATGCAGCCACTAAAAAGGCTGGTGAAGA
Sopen10g007490.1      GTCTGATAGAACAGATCAACCAGCTAGTCTATATGCTGCAACTAAGAAAGCGGGCGAAGA
Sopen01g036950.1      GAAGGATAGAACAGATCAGCCTGCTAGTCTTTATGCTGCAACTAAAAAGGCTGGTGAAGA
Sopen05g029460.1      GTGGGATAGGACAGACCAGCCTGCTAGTTTATATGCTGCTACTAAGAAGGCTGGTGAGGA
                           *    ** ** ** ** **     * ** ** ** ** ** ** ** ** ** * *
```

```
Sopen07g002320.1      AATTACACACACGTATAATCATATTTACGGGTTATCGATAACCGGGTTAAGATTCTTCAC
Sopen12g005550.1      AATTACACATACTTATAATCACATTTACGGGTTATCAATAACCGGGTTGAGATTTTTTAC
Sopen12g033260.1      AATTGCACATACGTATAATCATATTTACGGTCTTTCTTTAACCGCATTAAGGTTTTTTAC
Sopen12g033320.1      AATTGCACATACGTATAATCATATTTACGGTCTTTCTTTAACCGCATTAAGGTTTTTTAC
Sopen05g032260.1      AATTGCTCACACGTATAACCATATCTACGGATTGTCATTAACTGGGTTACGTTTCTTCAC
Sopen09g035230.1      AATTGCACATACATATAACCATATCTACGGGCTTTCATTAACAGGTTTGAGATTTTTCAC
Sopen03g023360.1      AATTGCACATACGTACAATCACATCTACGGTCTTTCCATTACAGGGTTAAGATTCTTCAC
Sopen08g027810.1      AATTGCTCATACATATAACCATATATATGGTCTTTCGATTACTGGATTGCGGTTTTTCAC
Sopen10g007490.1      AATTGCTCATACATATAATCATATATATGGTCTTTCACTTACTGGATTGCGATTTTTCAC
Sopen01g036950.1      AATTGCTCATACATATAATCATATATATGGGCTTTCATTAACTGGATTGAGATTTTTCAC
Sopen05g029460.1      GATTGCTCATACCTATAATCATATATATGGGCTTTCGATAACTGGATTGAGGTTTTTCAC
                       *** * ** ** ** ** ** ** ** **  * **  * ** *  **  * ** ** **

Sopen07g002320.1      TGTCTACGGACCGTGGGGAAGACCCGACATGGCTTATTTCAGCTTTACCCGTAACATATT
Sopen12g005550.1      TGTTTATGGACCATGGGGAAGACCCGATATGGCTTATTTTTCGTTTACCCGGAATATCTT
Sopen12g033260.1      TGTATACGGACCGTGGGGTAGACCAGACATGGCCTATTTCTTTTTCACTAAGGACATGGT
Sopen12g033320.1      TGTATACGGACCGTGGGGTAGACCAGACATGGCCTATTTCTTTTTCACTAAGGACATGGT
Sopen05g032260.1      TGTTTATGGACCTTGGGGAAGGCCGGACATGGCGTATTTCTTCTTTACAAAGGATATGAT
Sopen09g035230.1      TGTTTACGGACCTTGGGGAAGACCAGACATGGCGTATTTCTTCTTCACAAAGGATATGAT
Sopen03g023360.1      TGTTTATGGACCTTGGGGTAGGCCAGATATGGCGTATTTCTTCTTCACAAAGAACATACT
Sopen08g027810.1      AGTTTATGGACCATGGGGTAGGCCTGATATGGCGTATTTTTTCTTTACAAAAGATATTTT
Sopen10g007490.1      TGTTTATGGACCATGGGGCAGGCCAGATATGGCGTATTTCTTTTTCACTAGGGATATTTT
Sopen01g036950.1      CGTTTACGGACCATGGGGTAGGCCAGACATGGCTTACTTCTTTTTCACAAGGGATATCTT
Sopen05g029460.1      AGTTTATGGACCGTGGGGGCGGCCAGATATGGCGTATTTCTTTTTCACTAGGGATATTTT
                       ** ** ***** *****  * ** ** ***** ** **    ** **     * **  *

Sopen07g002320.1      ACAAGGGAAACCGATTACGGTTTATCGAGGTAAGAATCGAGTCGATTTAGCTCGGGATTT
Sopen12g005550.1      ACAAGGTAAATCAATAACAATTTATCGAGGTAAGAATCGAGTTGATTTAGCACGTGATTT
Sopen12g033260.1      ACAAGGTAAACCAATAAACGTTTACGTTACACAAGATGACAAAGAGGTGGCACGTGACTT
Sopen12g033320.1      ACAAGGTAAACCAATAAACGTTTACGTTACACAAGATGACAAAGAGGTGGCACGTGACTT
Sopen05g032260.1      TCAAGGGAAGCCGATCAAGGTTTACGTATCGGATAA------GGAGGTGGCGCGTGACTT
Sopen09g035230.1      TCAGGGGAAATCGATTAACGTGTACGTCACTCAGGATGATAAAGAGGTGGCGCGTGACTT
Sopen03g023360.1      AAAAGGAAAGCCCATTTCAGTATTTCAAGGTTCAAACAACAAAAGTGTAGCTAGAGATTT
Sopen08g027810.1      GAGAGGGAAGGAAATCAAGATTTTTGAGACATCTGATCATGGTAGTGTTGCTAGGGATTT
Sopen10g007490.1      GAAGGGGAAGTCTATATCGATCTTTGAGGGAGCTAATCATGGGACTGTTGCTAGGGACTT
Sopen01g036950.1      GAAGGGAAAATCGATTCCTATATTCGAGGCTGCTAATCATGGCACGGTAGCTAGGGATTT
Sopen05g029460.1      GAAGGGAAAGTCAATTCCAATCTTTGAGGCAGCTAATCATGGCACGGTCGCGAGGGATTT
                          ** **    **     * *           *           * **  * ** **

Sopen07g002320.1      TACTTACATCGATGATGTGGTAAAAGGTTGTATCGGGTCACTTGATACTTCGGGTAAAAG
Sopen12g005550.1      TACGTACATAGATGATATTGTGAAAGGGTGTGTTGGGTCACTTGATACAGCAGGGAAGAG
Sopen12g033260.1      CACGTATATCGATGACATCGTTAAGGGGTGTCTCGGGTCGTTAGACACGGCGGAGAAGAG
Sopen12g033320.1      CACGTATATCGATGACATCGTTAAGGGGTGTCTCGGGTCGTTAGACACGGCGGAGAAGAG
Sopen05g032260.1      CACGTACATAGATGATGTGGTGAAAGGTTGTTTAGGTGCGATTGATACGGCGGAGAAAAG
Sopen09g035230.1      CACGTACATTGATGATATAGTAAAAGGATGCGTCGGCGCGTTGGATACGGCGGAGAAGAG
Sopen03g023360.1      TACCTACATTGATGATATAGTAAAAGGGTGTTTAGGGGCTTTGGATACAGCAGAGAAGAG
Sopen08g027810.1      TACTTATGTTGATGATGTGGTAAAGGGTTGTTTGGGAGCACTGGATACTGCGAAAAAGAG
Sopen10g007490.1      TACCTACATTGATGACATAGTAAAGGGTTGTTTGGGAGCGTTGGACACGGCTGAGAAGAG
Sopen01g036950.1      TACCTACATTGATGATATAGTAAAAGGATGTTTGGCAGCATTGGATACTGCTGAGAAGAG
Sopen05g029460.1      TACCTACATTGATGACATAGTTAAGGGTTGTTTGGGGGCATTGGACACGGCTAAGAAGAG
                       ** **  * *****  * ** ** ** **  * *   *  * ** **  *    ** **

Sopen07g002320.1      TACCGGGTCGGGTGGGAAGAAACGGGGACCCGCTCCATATCGGATATTTAATTTGGGTAA
Sopen12g005550.1      TACCGGGTCGGGTGGGAAAAAACGGGCCCCGGCTATGTTTCGGATATTTAATTTGGGTAA
Sopen12g033260.1      CACCGGAAGTGGTGGCAAGAAGAAGGGTCCGGCCCAATTAAGGGTGTACAATTTAGGTAA
Sopen12g033320.1      CACCGGAAGTGGTGGCAAGAAGAAGGGTCCGGCCCAATTAAGGGTGTACAATTTAGGTAA
Sopen05g032260.1      CACCGGTAGCGGCGGAAAGAAGAGAGGTCCGGCGCAGTTGAGAGTGTATAATTTGGGGAA
Sopen09g035230.1      CACCGGTAGCGGCGGAAAGAAGAGAGGTCCGGCGCAATTGAGGGTTTACAATTTGGGGAA
Sopen03g023360.1      CACAGGAAGTGGTGGTAAGAAGAAGAAGAATGCACAATTAAGAGTGTTTAATTTGGGGAA
Sopen08g027810.1      CACAGGAAGTGGGGGAAAGAAGAAGGGTGCAGCTCAGTTGAGGATTTTTAATTTAGGTAA
Sopen10g007490.1      CACAGGAAGTGGTGGGAAAAAGAAAGGTCCTGCTCAATTGCGCGTGTTCAATTTAGGAAA
Sopen01g036950.1      CACTGGAAGTGGTGGGAAGAAGAAAGGCGCTGCTCAACTGCGGGTGTTCAATTTAGGCAA
Sopen05g029460.1      CACCGGAAGTGGTGGGAAGAAGAAAGGTCCTGCTCAATTGAGGGTGTTCAATTTGGGCAA
                       ** **    ** ** ** **          **        *  * *  ***** ** **

Sopen07g002320.1      TACGTCGCCGGTGACAGTTCCGATGATGGTTGCGATGTTAGAGAAGCATTTAAAGGTAAA
Sopen12g005550.1      TACTTCGCCCGTTACGGTACCGATGATGGTTGCAATGTTAGAGAAGCATTTGAAAGTAAA
Sopen12g033260.1      CACTTCACCAATATCAGTAAACAAATTGGTGACAATTTTGGAGAATTTATTAAATGTTAA
Sopen12g033320.1      CACTTCACCAATATCAGTAAACAAATTGGTGACAATTTTGGAGAATTTATTAAATGTTAA
Sopen05g032260.1      TACTTCGCCGGTATCGGTGAAGAAGCTAGTGGCAATTTTGGAAAATTTACTCAACATTAA
Sopen09g035230.1      TACTTCACCAGTGTCGGTGAAGAAGTTGGTGGCAATTCTCGAAAATTTATTGAATCTTAA
Sopen03g023360.1      TACTTCACCAGTTCCAGTGACAAAGTTAGTTAGCATATTGGAGAAACTACTAAAGGTAAA
Sopen08g027810.1      TACAAAGCCGGTGCCTGTTGGGAGACTTGTTAGTATTTTGGAGAAGTTGTTGAAGGTGAA
Sopen10g007490.1      CACTTCTCCTGTGCCTGTTTCGGATCTTGTTACCATTTTGGAAAGGTTACTAAAGGTGAA
Sopen01g036950.1      CACATCTCCTGTTCCAGTTTCAGATCTTGTAGGCATTTTGGAGAGGTTGCTAAAGGTGAA
Sopen05g029460.1      CACTTCACCTGTCCCGGTATCTGATCTTGTCAGCATTTTGGAGAAGTTGTTAAAGGTAAA
                       **    **  *  * **        * **    **  * ** *      * **  * **

Sopen07g002320.1      AGCTAAGAAGAATTTCGTCGTAATGCCCGGAAACGGCGACGTTCCGTTTACTCATGCGAA
Sopen12g005550.1      AGCTAAAAAACATGTATTGGATATGCCTGGAAACGGCGACGTTCCGTTTACCCATGCGAA
Sopen12g033260.1      GGCTAAAAAAAATGTAATTAAAATGCCTAGAAATGGTGATGTACCATTTACACATGCTAA
Sopen12g033320.1      GGCTAAAAAAAATGTAATTAAAATGCCTAGAAATGGTGATGTACCATTTACACATGCTAA
Sopen05g032260.1      GGCTAAAAAGAATTTGATTAGAATGCCCCGAAACGGTGACGTTCCGTTTACTCATGCTAA
Sopen09g035230.1      GGCTAAAAAGAATGTTATTAAAATGCCACGAAACGGCGACGTTCCGTATACACATGCTAA
Sopen03g023360.1      GGCTAAAAGAAATGTATTGCCATTGCCAACAAATGGAGATGTGATGTTTACTCATGCTAA
Sopen08g027810.1      GGCAAAGAAGAAGGTTATCCAAATGCCGAGGAATGGGGATGTGCCATTTACGCATGCTAA
Sopen10g007490.1      AGCTAAACGAGCTGTGATGAAGTTGCCAAGGAACGGTGATGTTCAGTTTACTCATGCTAA
Sopen01g036950.1      GGCCAAGAGATTGGTGATGAAGTTGCCAAGGAATGGGGATGTGCCTTTTACTCATGCCAA
Sopen05g029460.1      GGCTAAGAGATTGATTATGAAGTTACCAAGGAATGGAGACGTGCAGTTTACTCATGCAAA
                       ** **        *  *     * **    ** ** ** **    * *** ***** **
```

```
Sopen07g002320.1      TATAAGTTCGGCCCGGAAGGAATTCGGGTATAAACCCACAACCGATTTACAAACCGGGTT
Sopen12g005550.1      TATTAGTTTGGCCCAAAAGGAACTCGGGTATAAACCGACAACCAATTTGCAAACCGGGTT
Sopen12g033260.1      TATTACATTGGCTAAAAAGGATTTTGGGTATAAGCCAACCACAGATTTGTCAAGTGGATT
Sopen12g033320.1      TATTACATTGGCTAAAAAGGATTTTGGGTATAAGCCAACCACAGATTTGTCAAGTGGATT
Sopen05g032260.1      CGTGAGCCTGGCGTATAGAGATTTTGGGTATAAGCCGACTACAAATTTGTCAAGTGGATT
Sopen09g035230.1      CGTGAGCCTGGCGTTGAGGGATTTTGGATATAAGCCAACAACTGATTTGTCAAGTGGGTT
Sopen03g023360.1      TATCAGTTATGCTCACAAGGAATTTGGATACAAACCCACTACAGATTTGCAGATGGGGTT
Sopen08g027810.1      TATTACTTTGGCACATACAGAGCTGGGATATAAGCCTACTACTGATTTGGAAATGGGGTT
Sopen10g007490.1      TATAAGCTTAGCACAAAGGGAACTTGGATATAAGCCTACAACAGATTTGCAGACAGGGCT
Sopen01g036950.1      TATAAGTTCAGCCCACAAGGAGCTTGGATATAAGCCTACGACGGATCTACAGACGGGATT
Sopen05g029460.1      TATTAGCTTGGCTCAGAAGGAGTTTGGGTATAAGCCTACCACAGATCTACAGACAGGATT
                        * *     **    *  **  * ** ** ** ** ** **  ** *    *  **  *

Sopen07g002320.1      GAAAAAGTTTGTTAAATGGTATCTCTCTTATTATGG-------CTATGATC---------
Sopen12g005550.1      GGGGAAATTCGTTAGGTGGTATCTCTCCTATTATGG-------TTATAGTC---------
Sopen12g033260.1      AAGGAAGTTTGTCAAGTGGTATGTGAGTTATTATGGGATTCAATCAAAAGATTTAGACTC
Sopen12g033320.1      AAGGAAGTTTGTCAAGTGGTATGTGAGTTATTATGGGATTCAATCAAAAGATTTAGACTC
Sopen05g032260.1      AAGGAAATTTGTTAAGTGGTATTTAAGTTATTATGGGATTCAAGCAAGGAT---------
Sopen09g035230.1      AAGGAAATTTGTGAAGTGGTATGTTAGTTATTATGGGATTCAATCAAGGGT---------
Sopen03g023360.1      GCACAAATTTGTTAATTGGTATCTTGATTACTATTC-----AGTAAGTGAA---------
Sopen08g027810.1      AAAGAAGTTTGTGAAGTGGTATGTTAGCTACTATGG-----TTCGAAGAAG---------
Sopen10g007490.1      TAAGAAATTCGTTCGATGGTACCTTAGCTACTATGG--------TGAGGGA---------
Sopen01g036950.1      GAAGAAATTTGTCCGATGGTACCTCAATTACTATGG--------TAATGGA---------
Sopen05g029460.1      GAAGAAATTTGTTCGATGGTACCTTAGCTACTATGG--------TAATGGA---------
                          ** ** **    *****  *    ** ***

Sopen07g002320.1      ----AAGGAAAGTT--------------GTAA
Sopen12g005550.1      ----AAGAAAAGTCTATGAAAG-----AATAA
Sopen12g033260.1      TTCTAATGACCATTCTCAAG-------ACTGA
Sopen12g033320.1      TTCTAATGACCATTCTCAAG-------ACTGA
Sopen05g032260.1      ----AA----------------------AATAG
Sopen09g035230.1      ----AAAAAAGGGAAGTGAAGGGGAAAAGTGA
Sopen03g023360.1      ----AAGAAGATTT-------------ATTGA
Sopen08g027810.1      ----AAGAAGAGTTCTT----------GGTGA
Sopen10g007490.1      ----AAGAAAAGCGCGC----------AGTGA
Sopen01g036950.1      ----AAGAAGAGTGCCC----------AGTGA
Sopen05g029460.1      ----AAGAAAAGCGCGC----------AGTGA
                          **                      *
```

Figure S4.4- *S. pennellii* epimerases CDS alignment.

```
Sopen12g033260      ------------------------------------------------------------
Sopen12g033320      ------------------------------------------------------------
XM_015204087.1      TATTCTCAAATAAAATTTCTTTCTACTTATTTTCCAAAAAAAATAATATTTGGTCATTTT
XM_015204086.1      TATTCTCAAATAAAATTTCTTTCTACTTATTTTCCAAAAAAAATAATATTTGGTCATTTT

Sopen12g033260      ------------------------------------------------------------
Sopen12g033320      ------------------------------------------------------------
XM_015204087.1      TCTATGATCAATGGCATCATTTCCAATTGATACAAGCAAAGAAATGAAATTAGAGAGATA
XM_015204086.1      TCTATGATCAATGGCATCATTTCCAATTGATACAAGCAAAGAAATGAAATTAGAGAGATA

Sopen12g033260      ------------------------------------------------------------
Sopen12g033320      ------------------------------------------------------------
XM_015204087.1      TAATAGTTATATTAGAAGATTAAATAGCACAAAATTAATTGTTGCTTCTTCTAAACTTCT
XM_015204086.1      TAATAGTTATATTAGAAGATTAAATAGCACAAAATTAATTGTTGCTTCTTCTAAACTTCT

Sopen12g033260      ------------------------------------------------------------
Sopen12g033320      ------------------------------------------------------------
XM_015204087.1      TTTTAGAGTTACACTTTTAGTAGCTTTAATATTAATCTTTTTTTTCATCATAAATTACCC
XM_015204086.1      TTTTAGAGTTACACTTTTAGTAGCTTTAATATTAATCTTTTTTTTCATCATAAATTACCC

Sopen12g033260      ----------------------AGTAATCCACATCATCATAATATTCACACCACCACCCA
Sopen12g033320      -----------CATCGGAAAATAGTAATCCACATCATCATAATATTCACACCACCACCCA
XM_015204087.1      TTCTTTTATCTCATCGGAAAATAGTAATCCACATCATCATAATATTCACACCACCACCCA
XM_015204086.1      TTCTTTTATCTCATCGGAAAATAGTAATCCACATCATCATAATATTCACACCACCACCCA
                                         ************************************

Sopen12g033260      TAACCTCCTCTCCTCCTCCTTCTACGGTGGTGGCGCCGCTTGGGAGAAACAAGTCCGCCA
Sopen12g033320      TAACCTCCTCTCCTCCTCCTTCTACGGTGGTGGCGCCGCTTGGGAGAAACAAGTCCGCCA
XM_015204087.1      TAACCTCCTCTCCTCCTCCTTCTACGGTGGTGGCGCCGCTTGGGAGAAACAAGTCCGCCA
XM_015204086.1      TAACCTCCTCTCCTCCTCCTTCTACGGTGGTGGCGCCGCTTGGGAGAAACAAGTCCGCCA
                    ************************************************************

Sopen12g033260      CTCCTCCACTCCTCGACGTGTCAATGGTTTATCCGTATTGGTCACAGGTTAGTACGATTG
Sopen12g033320      CTCCTCCACTCCTCGACGTGTCAATGGTTTATCCGTATTGGTCACAGGTTAGTACGATTG
XM_015204087.1      CTCCTCCACTCCTCGACGTGTCAATGGTTTATCCGTATTGGTCA----------------
XM_015204086.1      CTCCTCCACTCCTCGACGTGTCAATGGTTTATCCGTATTGGTCA----------------
                    *******************************************

Sopen12g033260      CATAGTTTACGCGATTTAATTTAAATCATAAATGCATGTACACATAAAAATAGTACAAGT
Sopen12g033320      CATAGTTTACGCGATTTAATTTAAATCATAAATGCATGTACACATAAAAATAGTACAAGT
XM_015204087.1      ------------------------------------------------------------
XM_015204086.1      ------------------------------------------------------------

Sopen12g033260      TAGTGTAATTATTATGCCAGTATAAAAGTTAAAATTTTAAAAAAATTAAATTCTGAATTC
Sopen12g033320      TAGTGTAATTATTATGCCAGTATAAAAGTTAAAATTTTAAAAAAATTAAATTCTGAATTC
XM_015204087.1      ------------------------------------------------------------
XM_015204086.1      ------------------------------------------------------------

Sopen12g033260      GCTTCTGATTCAGGTGCGGCGGGTTTTGTTGGTTCTCATTGCTCAATGGCATTAAAGAAA
Sopen12g033320      GCTTCTGATTCAGGTGCGGCGGGTTTTGTTGGTTCTCATTGCTCAATGGCATTAAAGAAA
XM_015204087.1      ----------CAGGTGCGGCGGGTTTTGTTGGTTCTCATTGCTCAATGGCATTAAAGAAA
XM_015204086.1      ----------CAGGTGCGGCGGGTTTTGTTGGTTCTCATTGCTCAATGGCATTAAAGAAA
                              ************************************************

Sopen12g033260      CGTGGGGATGGTGTCTTGGGAATAGACAATTTTAATTCCTATTATGACCCTTCATTGAAA
Sopen12g033320      CGTGGGGATGGTGTCTTGGGAATAGACAATTTTAATTCCTATTATGACCCTTCATTGAAA
XM_015204087.1      CGTGGGGATGGTGTCTTGGGAATAGACAATTTTAATTCCTATTATGACCCTTCATTGAAA
XM_015204086.1      CGTGGGGATGGTGTCTTGGGAATAGACAATTTTAATTCCTATTATGACCCTTCATTGAAA
                    ************************************************************

Sopen12g033260      CGTGGGCGCCAAAAGTTATTGGCCCAACACGAAGTCTTCATAGTGGAAGGTGATATAAAC
Sopen12g033320      CGTGGGCGCCAAAAGTTATTGGCCCAACACGAAGTCTTCATAGTGGAAGGTGATATAAAC
XM_015204087.1      CGTGGGCGCCAAAAGTTATTGGCCCAACACGAAGTCTTCATAGTGGAAGGTGATATAAAC
XM_015204086.1      CGTGGGCGCCAAAAGTTATTGGCCCAACACGAAGTCTTCATAGTGGAAGGTGATATAAAC
                    ************************************************************

Sopen12g033260      GATGCCGAGCTGTTATCTAAACTGTTCGACATCGTTCCATTCACACACGTACTACATCTA
Sopen12g033320      GATGCCGAGCTGTTATCTAAACTGTTCGACATCGTTCCATTCACACACGTACTACATCTA
XM_015204087.1      GATGCCGAGCTGTTATCTAAACTGTTCGACATCGTTCCATTCACACACGTACTACATCTA
XM_015204086.1      GATGCCGAGCTGTTATCTAAACTGTTCGACATCGTTCCATTCACACACGTACTACATCTA
                    ************************************************************
```

```
Sopen12g033260      GCCGCACAGGCGGGGGGTGCGGTATGCGATGAAAAATCCACTATCGTATGTCCAATCAAAC
Sopen12g033320      GCCGCACAGGCGGGGGGTGCGGTATGCGATGAAAAATCCACTATCGTATGTCCAATCAAAC
XM_015204087.1      GCCGCACAGGCGGGGGGTGCGGTATGCGATGAAAAATCCACTATCGTATGTCCAATCAAAC
XM_015204086.1      GCCGCACAGGCGGGGGGTGCGGTATGCGATGAAAAATCCACTATCGTATGTCCAATCAAAC
                    ************************************************************

Sopen12g033260      ATTGCGGGGGTTTGTCAACTTATTGGAGACGGTCAAGTTAGCTAACCCTCAACCCGCGATT
Sopen12g033320      ATTGCGGGGGTTTGTCAACTTATTGGAGACGGTCAAGTTAGCTAACCCTCAACCCGCGATT
XM_015204087.1      ATTGCGGGGGTTTGTCAACTTATTGGAGACGGTCAAGTTAGCTAACCCTCAACCCGCGATT
XM_015204086.1      ATTGCGGGGGTTTGTCAACTTATTGGAGACGGTCAAGTTAGCTAACCCTCAACCCGCGATT
                    ************************************************************

Sopen12g033260      GTCTGGGCATCGTCCAGCTCGGTTTACGGACTGAACACGAACGTACCGTTTTCCGAAAGT
Sopen12g033320      GTCTGGGCATCGTCCAGCTCGGTTTACGGACTGAACACGAACGTACCGTTTTCCGAAAGT
XM_015204087.1      GTCTGGGCATCGTCCAGCTCGGTTTACGGACTGAACACGAACGTACCGTTTTCCGAAAGT
XM_015204086.1      GTCTGGGCATCGTCCAGCTCGGTTTACGGACTGAACACGAACGTACCGTTTTCCGAAAGT
                    ************************************************************

Sopen12g033260      CACCGTACGGATCAACCAGCTAGCCTTTACGCCGCGACGAAAAAAGCCGGTGAAGAAATT
Sopen12g033320      CACCGTACGGATCAACCAGCTAGCCTTTACGCCGCGACGAAAAAAGCCGGTGAAGAAATT
XM_015204087.1      CACCGTACGGATCAACCAGCTAGCCTTTACGCCGCGACGAAAAAAGCCGGTGAAGAAATT
XM_015204086.1      CACCGTACGGATCAACCAGCTAGCCTTTACGCCGCGACGAAAAAAGCCGGTGAAGAAATT
                    ************************************************************

Sopen12g033260      GCACATACGTATAATCATATTTACGGTCTTTCTTTAACCGCATTAAGGTTTTTTACTGTA
Sopen12g033320      GCACATACGTATAATCATATTTACGGTCTTTCTTTAACCGCATTAAGGTTTTTTACTGTA
XM_015204087.1      GCACATACGTATAATCATATTTACGGTCTTTCTTTAACCGCATTAAGGTTTTTTACTGTA
XM_015204086.1      GCACATACGTATAATCATATTTACGGTCTTTCTTTAACCGCATTAAGGTTTTTTACTGTA
                    ************************************************************

Sopen12g033260      TACGGACCGTGGGGTAGACCAGACATGGCCTATTTCTTTTTCACTAAGGACATGGTACAA
Sopen12g033320      TACGGACCGTGGGGTAGACCAGACATGGCCTATTTCTTTTTCACTAAGGACATGGTACAA
XM_015204087.1      TACGGACCGTGGGGTAGACCAGACATGGCCTATTTCTTTTTCACTAAGGACATGGTACAA
XM_015204086.1      TACGGACCGTGGGGTAGACCAGACATGGCCTATTTCTTTTTCACTAAGGACATGGTACAA
                    ************************************************************

Sopen12g033260      GGTAAACCAATAAACGTTTACGTTACACAAGATGACAAAGAGGTGGCACGTGACTTCACG
Sopen12g033320      GGTAAACCAATAAACGTTTACGTTACACAAGATGACAAAGAGGTGGCACGTGACTTCACG
XM_015204087.1      GGTAAACCAATAAACGTTTACGTTACACAAGATGACAAAGAGGTGGCACGTGACTTCACG
XM_015204086.1      GGTAAACCAATAAACGTTTACGTTACACAAGATGACAAAGAGGTGGCACGTGACTTCACG
                    ************************************************************

Sopen12g033260      TATATCGATGACATCGTTAAGGGGTGTCTCGGGTCGTTAGACACGGCGGAGAAGAGCACC
Sopen12g033320      TATATCGATGACATCGTTAAGGGGTGTCTCGGGTCGTTAGACACGGCGGAGAAGAGCACC
XM_015204087.1      TATATCGATGACATCGTTAAGGGGTGTCTCGGGTCGTTAGACACGGCGGAGAAGAGCACC
XM_015204086.1      TATATCGATGACATCGTTAAGGGGTGTCTCGGGTCGTTAGACACGGCGGAGAAGAGCACC
                    ************************************************************

Sopen12g033260      GGAAGTGGTGGCAAGAAGAAGGGTCCGGCCCAATTAAGGGTGTACAATTTAGGTAACACT
Sopen12g033320      GGAAGTGGTGGCAAGAAGAAGGGTCCGGCCCAATTAAGGGTGTACAATTTAGGTAACACT
XM_015204087.1      GGAAGTGGTGGCAAGAAGAAGGGTCCGGCCCAATTAAGGGTGTACAATTTAGGTAACACT
XM_015204086.1      GGAAGTGGTGGCAAGAAGAAGGGTCCGGCCCAATTAAGGGTGTACAATTTAGGTAACACT
                    ************************************************************

Sopen12g033260      TCACCAATATCAGTAAACAAATTGGTGACAATTTTGGAGAATTTATTAAATGTTAAGGCT
Sopen12g033320      TCACCAATATCAGTAAACAAATTGGTGACAATTTTGGAGAATTTATTAAATGTTAAGGCT
XM_015204087.1      TCACCAATATCAGTAAACAAATTGGTGACAATTTTGGAGAATTTATTAAATGTTAAGGCT
XM_015204086.1      TCACCAATATCAGTAAACAAATTGGTGACAATTTTGGAGAATTTATTAAATGTTAAGGCT
                    ************************************************************

Sopen12g033260      AAAAAAAATGTAATTAAAATGCCTAGAAATGGTGATGTACCATTTACACATGCTAATATT
Sopen12g033320      AAAAAAAATGTAATTAAAATGCCTAGAAATGGTGATGTACCATTTACACATGCTAATATT
XM_015204087.1      AAAAAAAATGTAATTAAAATGCCTAGAAATGGTGATGTACCATTTACACATGCTAATATT
XM_015204086.1      AAAAAAAATGTAATTAAAATGCCTAGAAATGGTGATGTACCATTTACACATGCTAATATT
                    ************************************************************

Sopen12g033260      ACATTGGCTAAAAAGGATTTTGGGTATAAGCCAACCACAGATTTGTCAAGTGGATTAAGG
Sopen12g033320      ACATTGGCTAAAAAGGATTTTGGGTATAAGCCAACCACAGATTTGTCAAGTGGATTAAGG
XM_015204087.1      ACATTGGCTAAAAAGGATTTTGGGTATAAGCCAACCACAGATTTGTCAAGTGGATTAAGG
XM_015204086.1      ACATTGGCTAAAAAGGATTTTGGGTATAAGCCAACCACAGATTTGTCAAGTGGATTAAGG
                    ************************************************************

Sopen12g033260      AAGTTTGTCAAGTGGTATGTGAGTTATTATGGGATTCAATCAAAAGATTTAGACTCTTCT
Sopen12g033320      AAGTTTGTCAAGTGGTATGTGAGTTATTATGGGATTCAATCAAAAGATTTAGACTCTTCT
XM_015204087.1      AAGTTTGTCAAGTGGTATGTGAGTTATTATGGGATTCAATCAAAAGATTTAGACTCTTCT
XM_015204086.1      AAGTTTGTCAAGTGGTATGTGAGTTATTATGGGATTCAATCAAAAGATTTAGACTCTTCT
                    ************************************************************
```

```
Sopen12g033260      AATGACCATTCTCAAGACTGAGTGATAATTGATGATGATATTCGAATCAGCTTGTACTGA
Sopen12g033320      AATGACCATTCTCAAGACTGAGTGATAATTGATGATGATATTCGAATCAGCTTGTACTGA
XM_015204087.1      AATGACCATTCTCAAGACTGAGTGATAATTGATGATGATATTCGAATCAGCTTGTACTGA
XM_015204086.1      AATGACCATTCTCAAGACTGAGTGATAATTGATGATGATATTCGAATCAGCTTGTACTGA
                    ************************************************************

Sopen12g033260      TACATTTCAGTCATTGTGTCAGATACGATACATGTTACTTATTAGGGTTCGACGTCCACC
Sopen12g033320      TACATTTCAGTCATTGTGTCAGATACGATACATGTTACTTATTAGGGTTCGACGTCCACC
XM_015204087.1      TACATTTCAGTCATTGTGTCAGATACGATACATGTTACTTATTAGGGTTCGACGTCCACC
XM_015204086.1      TACATTTCAGTCATTGTGTCAGATACGATACATGTTACTTATTAGGGTTCGACGTCCACC
                    ************************************************************

Sopen12g033260      GTATATATCAGATAATTCGCTCGGATGGATGAAAAGGCGAAAGGAAAGTTTAATTTTTTT
Sopen12g033320      GTATATATCAGATAATTCGCTCGGATGGATGAAAAGGCGAAAGGAAAGTTTAATTTTTTT
XM_015204087.1      GTATATATCAGATAATTCGCTCGGATGGATGAAAAGGCGAAAGGAAAGTTTAATTTTTTT
XM_015204086.1      GTATATATCAGATAATTCGCTCGGATGGATGAAAAGGCGAAAGGAAAGTTTAATTTTTTT
                    ************************************************************

Sopen12g033260      CTACATGTTTATATATATCAAGAAAAATGAGAGGAGTAGGAAAAGGACAAGTGTAGATTA
Sopen12g033320      CTACATGTTTATATATATCAAGAAAAATGAGAGGAGTAGGAAAAGGACAAGTGTAGATTA
XM_015204087.1      CTACATGTTTATATATATCAAGAAAAATGAGAGGAGTAGGAAAAGGACAAGTGTAGATTA
XM_015204086.1      CTACATGTTTATATATATCAAGAAAAATGAGAGGAGTAGGAAAAGGACAAGTGTAGATTA
                    ************************************************************

Sopen12g033260      TTGGTGAAGTGGGGTTTGTAATAAATGTTTTAATTAATA----------------------
Sopen12g033320      TTGGTGAAGTGGGGTTTGTAATAAATGTTTTAATTAATAGTTTTTCTGTACAAATAAAAA
XM_015204087.1      TTGGTGAAGTGGGGTTTGTAATAAATGTTTTAATTAATAGTTTTTCTGTACAAATAAAAA
XM_015204086.1      TTGGTGAAGTGGGGTTTGTAATAAATGTTTTAATTAATAGTTTTTCTGTACAAATAAAAA
                    ************************************

Sopen12g033260      ------------------------------------------------------------
Sopen12g033320      TAATTAGTGCAGATAAGCTTTTTCATTAATTGTAATGAGTTGAGATATATCAATGATAAA
XM_015204087.1      TAATTAGTGCAGATAAGCTTTTTCATTAATTGTAATGAGTTGAGATATATCAATGATAAA
XM_015204086.1      TAATTAGTGCAGATAAGCTTTTTCATTAATTGTAATGAGTTGAGATATATCAATGATAAA

Sopen12g033260      ------------------------------------------------------------
Sopen12g033320      AATCAAGATTCTCTTTTATGCTTTTCTTTGGCTTTATATTTATTGGTTCTCTTTTGTATT
XM_015204087.1      AATCAAGATTCTCTTTTATG-------------------------------------------
XM_015204086.1      AATCAAGATTCTCTTTTA--------------------------------------------

Sopen12g033260      ------------------------------------------------------------
Sopen12g033320      GTTTTTAAGCCATTGACTTAGTTTCTTCAACTTCAAACAAGATTAGAGTTGCATCATCTT
XM_015204087.1      ------------------------------------------------------------
XM_015204086.1      ------------------------------------------------------------

Sopen12g033260      ------------------------------------------------------------
Sopen12g033320      AATGTGTTGTTTATGTCAATCAATTGTATCAATAATGTCATGACTCATGCCCTTTGCTTT
XM_015204087.1      ------------------------------------------------------------
XM_015204086.1      ------------------------------------------------------------

Sopen12g033260      ------------------------------------------------------------
Sopen12g033320      GTTGTTTTTTTCTTATATTGTCTTTAATTAATGGGAATTACATACAAATTGTGAACAATT
XM_015204087.1      ------------------------------------------------------------
XM_015204086.1      ------------------------------------------------------------

Sopen12g033260      ------------------------------------------------------------
Sopen12g033320      TGTCTCTTAATTAGAGCATATGATAGTCAAGTTGTCACGATTTGAATTGTAATATTATGA
XM_015204087.1      ------------------------------------------------------------
XM_015204086.1      ------------------------------------------------------------

Sopen12g033260      ------------------------------------------------------------
Sopen12g033320      GTTTAAAGTTTTAGCAACTTCATATTTGTTCAATTGGACCTTTTTGGTTCCTTTGTTGAT
XM_015204087.1      ------------------------------------------------------------
XM_015204086.1      ------------------------------------------------------------

Sopen12g033260      ------------------------------------------------------------
Sopen12g033320      TATTCATAAAAAAAAAAAAAAATCAAAAATATATATTTTAGTATATATTTAAATAATATGG
XM_015204087.1      ------------------------------------------------------------
XM_015204086.1      ------------------------------------------------------------
```

```
Sopen12g033260      ----------
Sopen12g033320      AAATTTTTAC
XM_015204087.1      ----------
XM_015204086.1      ----------
```

Figure S4.5– Sol Genomics and NCBI S. *pennellii* epimerases alignment.

Fw 33320/33260 : ATGGCATCATTTCCAATTGATACAAG

Rv#2 33320/33260 : TCAGTCTTGAGAATGGTCATTAGAAGAGTCTAAAT

>Spenn-ch12 strand = minus, start = 81918776, end = 81917181, size = 1365 bp

ATGGCATCATTTCCAATTGATACAAGCAAAGAAATGAAATTAGAGAGATATAATAGTTATATTAGAAGATTAAATAGCACAAAATTA
ATTGTTGCTTCTTCTAAACTTCTTTTTAGAGTTACACTTTTAGTAGCTTTAATATTAATCTTTTTTTTCATCATAAATTACCCTTCTTTTAT
CTCATCGGAAAATAGTAATCCACATCATCATAATATTCACACCACCACCCATAACCTCCTCTCCTCCTCCTTCTACGGTGGTGGCGCCG
CTTGGGAGAAACAAGTCCGCCACTCCTCCACTCCTCGACGTGTCAATGGTTTATCCGTATTGGTCACAGGTGCGGCGGGTTTTGTTGGT
TCTCATTGCTCAATGGCATTAAAGAAACGTGGGGATGGTGTCTTGGGAATAGACAATTTTAATTCCTATTATGACCCTTCATTGAAACG
TGGGCGCCAAAAGTTATTGGCCCAACACGAAGTCTTCATAGTGGAAGGTGATATAAACGATGCCGAGCTGTTATCTAAACTGTTCGAC
ATCGTTCCATTCACACACGTACTACATCTAGCCGCACAGGCGGGGGTGCGGTATGCGATGAAAAATCCACTATCGTATGTCCAATCAA
ACATTGCGGGGTTTGTCAACTTATTGGAGACGGTCAAGTTAGCTAACCCTCAACCCGCGATTGTCTGGGCATCGTCCAGCTCGGTTTAC
GGACTGAACACGAACGTACCGTTTTCCGAAAGTCACCGTACGGATCAACCAGCTAGCCTTTACGCCGCGACGAAAAAAGCCGGTGAA
GAAATTGCACATACGTATAATCATATTTACGGTCTTTCTTTAACCGCATTAAGGTTTTTTACTGTATACGGACCGTGGGGTAGACCAGA
CATGGCCTATTTCTTTTTCACTAAGGACATGGTACAAGGTAAACCAATAAACGTTTACGTTACACAAGATGACAAAGAGGTGGCACGT
GACTTCACGTATATCGATGACATCGTTAAGGGGTGTCTCGGGTCGTTAGACACGGCGGAGAAGAGCACCGGAAGTGGTGGCAAGAAG
AAGGGTCCGGCCCAATTAAGGGTGTACAATTTAGGTAACACTTCACCAATATCAGTAAACAAATTGGTGACAATTTTGGAGAATTTAT
TAAATGTTAAGGCTAAAAAAAATGTAATTAAAATGCCTAGAAATGGTGATGTACCATTTACACATGCTAATATTACATTGGCTAAAAA
GGATTTTGGGTATAAGCCAACCACAGATTTGTCAAGTGGATTAAGGAAGTTTGTCAAGTGGTATGTGAGTTATTATGGGATTCAATCA
AAAGATTTAGACTCTTCTAATGACCATTCTCAAGACTGA

Figure S4.6- NCBI S. *pennellii* epimerases pcr amplification.

>Solyc06g073320_GGP1

atcacggctatacacaaagtaaaccgccgaccacttttacatgttccagcagtacgtcgtaagggttgtgtaacagctactaaccct
gcgccgcacggtggacgtggcgctttgccttctgaaggtggtagtccttccgacctcctcttccttgccggcggcggttctttcctc
tccttctcctactag

>gRNA1

tgttccagcagtacgtcgta

>gRNA2

gcggcggttctttcctctc

Figure S4.7- Solyc06g073320_cGGP1 uORF and gRNAs.

>Solyc02g091510_GGP2

atcacggctcttcttgaatctttcgtttgtattctcacaattcatcatcaccgcaaagtgttgacccttaatccaactcttctggtggacg
ataagcaccggacccctttcccctcacggaggtaggggtgcctcacccgctgaaggcggttgcccctccgatctcctcttcctcg
ccggcggcggtccaattcttcctttctctttttccttctcctaa

>gRNA1

ttctggtggacgataagcac

>gRNA2

gcggcggtccaattcttcct

Figure S4.8- Solyc02g091510_cGGP2 uORF and gRNAs.