UNIVERSITA' DEGLI STUDI DI VERONA

SCUOLA DI DOTTORATO IN SCIENZE DELLA VITA E DELLA SALUTE

DIPARTIMENTO SCIENTIFICO E TECNOLOGICO

DOTTORATO DI RICERCA IN BIOTECNOLOGIE MOLECOLARI, INDUSTRIALI ED AMBIENTALI

Curriculum: biologia molecolare vegetale

CICLO XXII

Genetic evolution of the pathway controlling fruit carotenoid content in tomato (*S. lycopersicum*) and its wild relatives

Coordinatore: Ch.mo Prof. Roberto Bassi

Tutor interno: Ch.mo Prof. Roberto Bassi

Tutor esterno: Ch.mo Prof. Giovanni Giuliano

Dottoranda: Dott. ssa Giulia Falcone

Summary

Tomato (*Solanum lycopersicum*) is an economically and nutritionally valuable crop and constitutes a model plant for genetic research of the *Solanaceae* family thanks its compact genome (950 Mb), short generation time, the availability of a large set of mutants, routine transformation technology and availability of rich genetic and genomic resources.

The pre-release of the tomato genome sequence, produced combining a whole genome shotgun and a BAC-by-BAC approach, has been announced at the end of 2009 and a first genome annotation is available and is continuously updated. The genome of cultivated tomato, however, has a limited sequence variation due to bottlenecks during domestication and breeding.

The wild tomatoes are native to Western South America and grow in a variety of habitats, from near sea level along the arid Pacific coast to high mountains up to 3300 m. Thanks to their evolutionary and adaptive history, the wild tomato species contain useful traits that can be introgressed into cultivated tomato, such as tolerance to drought and salinity and resistance to multiple pathogens. In recent years there has been an increasing interest in analyzing various biological properties of natural genetic diversity and wild species provide a wealth of useful genetic traits to improve cultivated tomatoes.

The evolution of carotenoid pigments affecting berry colour is one of the variable characters within tomato and its wild relatives that have obtained much attention: the fruit colour of wild tomatoes varies from green to orange and red. Numerous studies have demonstrated that green fruit is the ancestral character. The reason

behind a transition to coloured fruits during the evolution of the group is still not clear, but probably is related to attraction of animals for seed dispersal. Accumulation of carotenoids as secondary metabolites in fruits and flowers occurs through up-regulation of the pathway at the gene expression level.

A candidate gene approach has been used, focusing our attention on the carotenoid biosynthetic pathway. Several genes of this pathway have been sequenced in the wild tomato species with different berry colour. Accessions of the *S. lycopersicum var. cerasiforme* as well as wild tomato species (*S. pimpinellifolium*, *S. cheesmaniae*, *S. neorickii*, *S. chmielewskii*, *S. chilense*, *S. habrochaites*, *S. pennellii*, *S. arcanum*) were selected to represent ancestral and closely related progenitor genotypes.

Thanks to the availability of the genome sequences and annotation data of tomato, potato and *A. thaliana,* a comparison was undertaken to investigate the presence/absence of microsynteny around the Phytoene Synthase genes among these organisms. Moreover, we have obtained structural information on the organization of the carotenoid genes object of this thesis. The analysis has highlighted the presence of three new genes, annotated respectively as putative LCY-b, putative CrtISO and putative LCY-b: from preliminary transcriptome data the genes codifying for these proteins seem to be expressed.

The sequencing of carotenoid genes from *PSY* down to *LCY-e* (α-branch) and *CHY1-2* (β-branch) has been completed for all the wild species studied in this thesis. The sequence analysis has highlighted the presence of numerous mutations. Some non-synonymous substitutions are candidate to be hypo- or hypermorphic alleles. Preliminary analysis on *in vitro* protein expression seems to confirm that one of

the green-fruited species analysed carries a hypermorphic allele of *LCY-e.*

Green-fruited species did not differ systematically from coloured-fruited ones in climacteric ethylene production or fruit softening, with the exception of S. arcanum, which shows an ethylene peak just before fruit abscission.

Carotenoid and gene expression profiles of the red-fruited species, *S. lycopersicum* and *S. pimpinellifolium*, are very similar, in agreement with their phylogenetic closeness. In contrast to expectations, orange-fruited *S. cheesmaniae* has the lowest carotenoid content among all species and has carotenoid and transcriptional profiles similar to the green-fruited species.

1. Introduction

1.1 Tomato

Tomato (*Solanum lycopersicum*) is a perennial vegetable, which is almost universally cultivated as an annual crop. It belongs to the *Solanaceae* family, which also includes potato, eggplant, pepper, tobacco and petunia (Fig. 1). Besides its economic importance, it is a model plant for research purposes [1] [2] it is easy to cultivate, has a short life cycle and lends itself to horticultural manipulation including grafting or cutting. This species has many features that distinguish it from other model plants: it is phylogenetically distant from Arabidopsis, maize, rice, *Medicago* or poplar, it grows as an indeterminate plant due to reiterate switches from vegetative to reproductive stages and is the most advanced model system among species having a fleshy berry type of fruit. Other properties of tomato, such as the relatively small genome size (950 Mb), among the smallest in the *Solanaceae* family, the availability of a large set of mutants and the development of genomic and sequencing resources (high-density genetic and physical maps [3] [4], EST databases [5], efficient transient and stable transformation [6] and microarrays [7] [8]). These resources have favoured an international sequencing project (The International Tomato Sequencing Project [9]). The pre-release of the tomato genome, produced by a whole genome shotgun approach using 454 and Sanger shotgun and paired end sequences, and BAC/fosmid end sequences has been announced at the end of 2009 (extensive details can be obtained at the following website: http://solgenomics.net/tomato/)

and a first genome annotation is available and is continuously updated.

The tomato mutants constitute an essential source of plant material for breeders and a valuable tool for isolating important genes which regulate developmental patterns, and whose functional roles are now being elucidated. The Tomato Genetic Resource Center (http://tgrc.ucdavis.edu) has preserved and characterized many spontaneous mutants [10]. Mutagenized populations, mainly generated through chemicals (generally ethylmethane sulfonate or EMS) and irradiation, are providing available screening populations and the possibility to identify new developmental genes (http:// zamir.sgn.cornell.edu/mutants, http://urgv.evry.inra.fr/UTILLdb).

1.1.2 Tomato anatomy

Tomatoes are short-lived perennials cropped as annuals. Although killed by frost outdoors, in the greenhouse plants can be cropped for 24 months or longer. The tomato plant can have a fibrous root system or a taproot system depending on how the plant was grown. If the plant is grown from a seed, it will exhibit taproot organization; when the plant is grown from cuttings, a fibrous root system will form [11]. Tomato varieties vary in growth habit from vines, which spread horizontally (*decumbent*), to bush-like (*erect*). The tomato plant has compound leaves. A compound leaf is made up of leaflets, which are distributed along the leaf rachis. While the entire leaf is connected to the stem by the petiole, the leaflets are connected to the rachis of the leaf by the petiolule. Some of the leaflets on this leaf are compound as well. Tomato phyllotaxy is termed *spiral* because only one leaf is

present at each node and each successive leaf is displaced approximately 137.5 degrees from the last. Thus, a line connecting successively older leaves would make a spiral. The shoot system of the tomato plant is made up of branching stems. A stem has a terminal bud at the tip or *apex* that is responsible for the increase in length of the stem. Branches grow out from axillary/lateral buds. In tomato, as in all typical dicot plants, lateral buds are found in the axil of the leaf where the leaf connects to the stem. While the terminal bud (or shoot apex) of the stem is growing vegetatively (producing leaves), the growth of lateral branches is somewhat inhibited ("apical dominance"). Eventually, however, the terminal bud stops growing because it has aborted, been cut off by pruning, or it forms an inflorescence (branches producing flowers rather than leaves). When this happens, the lateral branches grow out [11]. Tomato plants have perfect (hermaphroditic) yellow flowers that, in full bloom, are generally less than 2.5 cm in diameter (Fig. 2). The flowers can occur in a simple or a complex inflorescence. Simple flowers can appear, as well as simple or branched cymes. The number of flowers that occur in an inflorescence is dependent upon environmental factors such as temperature. They have 5 sepals and 5 petals; stamens sit inside the petals and each one consists of two elongated compartments. The individual stamens are fused together to form a yellow cylinder that surrounds the carpels. The tomato carpels are green, vary in number from cultivar to cultivar, but they are invariably fused together into a single bulb-like structure. The number of carpels in the tomato flower corresponds to the number of locules found in the fruits. Carpels contain the ovules, which will develop into seeds. The tomato fruit is a berry composed of

flesh (pericarp walls and skin) and pulp (placenta and locular tissue including seeds) (Fig. 3). The pericarp includes the inner wall or columella; the radial walls or septa; and the outer wall. The pericarp and the placenta comprise the fleshy tissue of the tomato. The skin (epicarp or exocarp) is formed by the cuticle, which is highly integrated to the cell wall of the epidermis. Tomatoes can be either bilocular or multilocular. Most cultivated varieties except cherry tomatoes have four to five locules. Fruits can be yellow, orange, pink, red, or even white (Fig. 4). The red colour comes from the pigment lycopene, while the orange and yellow colours come from β-carotene and xanthophyll pigments, or, in the *r* mutant, from flavonoids [12]. Yellow and orange tomatoes are equal or higher in nutrition to red tomatoes because lycopene has no particular nutritional value (apart from its antioxidant properties) while β-carotene is a vitamin A precursor. Pink tomatoes have the same interior colour as red tomatoes but have transparent rather than yellow skin [13].

Most of the fruit fresh weight is water, with solids constituting only 5 percent. These solids consist of water insoluble substances such as cell walls, and soluble components such as sugars and acids. The amount of sugar present (generally about half of the solids) and the amount of acids present (generally about one-eighth of the solids) and their ratio determine the flavour. High sugars and high acid are the best combination for good flavour.

1.1.3 Climatic requirements

Although we think of tomatoes as an adaptable crop, they are actually quite

sensitive to low light and adverse temperatures. Tomato plants do not tolerate frost, and grow as annuals in colder regions. In warmer regions, they are perennial, and flower regardless of day length. Tomatoes need at least 6 hours of direct sunlight to flower. Slender, non-fruiting tomato vines are a common complaint of home gardeners and are usually traceable to shading [14]. Although tomatoes grow well over a wide range of temperatures (19 to 30 °C), fruit set is very sensitive to high and low temperatures. Above 33 °C day or 21 °C night temperatures and below 10 to 13 °C, flowers may produce oddly-shaped (rough) fruit or may fall off without setting fruit at all. Malformed fruit are sometimes said to be 'catfaced' or to have open locules. These malformations are the result of incomplete separation of cells during the early stages of flower and fruit development. Adverse effects seem to be worst when both day and night temperatures are high or when both are low. Proper colouring of the fruit is also temperature dependent. Lycopene and carotenes are not synthesized above 30 C° and lycopene is not synthesized below 10 C°, precluding normal colour development in ripening fruit. Tomato plants need acidic (pH 5.0 to 6.0) soil, a good balance between the three primary fertilizer nutrients, and thrive in warm weather. Phosphorus, boron and chlorine are also necessary to balance vegetative growth, with the obtainment of more resistant tissues and more firm berries [15].

Tomatoes can suffer from all kinds of diseases and pests. Regarding leaf diseases, the most severe injuries are caused by the gray leaf spot and the early blight and late blight (all by fungy pests), which can also affect the fruits. *Verticillium*, a root fungus pest, is one of the more dangerous pests of tomato plants, cause it can

persist in the soil for many years with a dramatic reduction in tomato yield and quality [11].

1.1.4 Fruit ripening

Mature fruits can be classified generally as either dry or fleshy, which mainly differ in the mechanism used achieved to permit seed dispersal. In dry fruits, such as those of Arabidopsis, a senescence program leading to fruit dehiscence is needed before some external agent (such as wind, rain, and physical contact) can force seeds to be released from dry fruits. Instead, fleshy fruits such as the tomato fruits have evolved components (colour, flavour, aroma) making them attractive and edible for animals, which facilitate dispersion of the seeds. Tomato plants produce fleshy red fruits as result of a developmental process, which includes four main phases [16]. The first starts at anthesis and involves the development of the ovary. In the second phase, fruit growth is due primarily to cell division. In the third phase, cell division ends and fruit growth continues by cell expansion until the fruit reaches its final size (mature green). Once the fruit is fully developed and seeds are mature, respiration and ethylene synthesis are significantly increased, and a series of biochemical changes, collectively known as ripening, start.

Typical events occurring during fruit ripening include [17]:

a. Colour modification, through the breakdown of chlorophyll and the accumulation of the linear carotene, lycopene (Fig. 5);

b. Texture modification, through alteration of cell turgor and cell wall structure and/or metabolism;

c. Modification of sugar and acid content and volatile profiles, that affects nutritional quality, flavour, and aroma;

d. Enhanced susceptibility to opportunistic pathogens that is likely associated with the loss of cell wall integrity [17].

Fleshy fruit ripening can be divided in two classes of ripening: ripening of climacteric fruits, such as tomato or apple, involves increased respiration and synthesis of the gaseous hormone ethylene at the onset of the process, while in non-climacteric fruits, such as strawberry or grape, this phenomenon is not observed. Strawberry is the main model for non-climacteric fruits, whereas significant contributions in the fields of hormonal regulation of ovary growth, physiology of ripening, and genetic control of fruit size and shape have made tomato the main model for climacteric, and in general fleshy fruit development (Fig. 6) [17] [2] [18] [19].

Ethylene is the most important hormone involved in fruit ripening. In plant tissues its production results from methionine metabolism (Fig. 7) [20] [21]. The rate-limiting steps in fruit ethylene synthesis include the conversion of S-adenosylmethionine to 1-aminocyclopropane-1-carboxylic acid (ACC) *via* ACC synthase (ACS) and the subsequent metabolism of ACC to ethylene by ACC oxidase (ACO) [22]. In most characterized plants, both steps are encoded by multigene families. Two systems of ethylene biosynthesis have been proposed to operate in climacteric plants. System 1 is functional during normal vegetative growth, where ethylene production auto-inhibitory and is responsible for producing basal ethylene levels that are detected in all tissues including those of non-climacteric

fruits. System 2 operates during the climacteric and senescence of some petals, when ethylene production is autocatalytic [23]. Ripening usually commences in one region of a fruit, spreading to neighbouring regions as ethylene diffuses freely from cell to cell and integrates the ripening process throughout the fruit. In addition to ripening, ethylene is also known to be involved in other processes such as pathogen and wound responses, leaf senescence and abiotic and biotic stress responses [17]. Ethylene binds to different receptors [22], [7], [24] and the binding triggers a signal cascade by MAP kinases, until activation of a first class of ethylene transcription factors (such as EIN3 and EIN3-like (EIL) proteins), which bind in a sequence-specific manner to ethylene-response elements [25]. ERF1, a member of the Ethylene Response Element Binding Protein (EREBP) family of DNA binding proteins directly activates transcription of a wide variety of ethylene-responsive pathogenesis-related genes or fruit ripening genes, by binding to GCC-box elements located in their promoters [23].

Studies on ethylene signalling have been conducted in Arabidopsis and tomato, the latter supported by a series of natural mutants in fruit ripening such as *nr*, *rin*, *nor* and *Gr* [2]. The *nr* mutant represents a lesion in an ethylene receptor and this mutation prevents ripening in tomato *via* ethylene insensitivity even when ethylene is applied exogenously [26]. *rin* (*ripening-inhibitor*) and *nor* (*non-ripening*) mutants fail to ripen in response to exogenous ethylene and yet display signs of ethylene sensitivity and signalling, including induction of some ethylene-regulated genes: the genes for both mutations were cloned and they revealed to be implied in the regulation of gene transcription upstream the ethylene signaling

cascade [27], [28]. Finally, *Gr* is a mutant in an evolutionary conserved protein of unknown biochemical function that is associated with ethylene signalling [29], [30].

1.1.5 The nutritional value of tomato

Based on the volume of consumption per person, tomato is a very good source of provitamin A and vitamin C in the Western diet. On a weight basis, tomato ranks 16th among all fruits and vegetables as a source of vitamin A, 13th in vitamin C, and when adjusted for consumption, is the most important provider of these two vitamins in the diet [31]. It also contains significant amounts of dietary fiber, β-carotene, iron, lycopene, magnesium, niacin, potassium, phosphorus, riboflavin, sodium and thiamine [32] [33].

Tomato is low in saturated fat, cholesterol and sodium [31] [34]. Unlike most foods, cooking or processing of tomato is beneficial to health, since it increases the bioavailability of lycopene. This is because heating up tomato breaks down its cell walls and releases more lycopene. Eating tomatoes has more benefits (with all of its other ingredients) than taking lycopene alone [35].

Tomatoes are an excellent source of lycopene; in fact lycopene is the major carotenoid contained in tomatoes. Similar to β-carotene, lycopene is a potent anti-oxidant and numerous studies have confirmed that people who consume increased amounts of tomato products experience marked reductions in cancer risk [36] [37] [38].

1.2 Evolution and taxonomy of tomato and its wild relatives

1.2.1 Evolution

The tomato clade is an evolutionarily young group that has diversified to occupy different kinds of habitats. The genus *Solanum* is estimated to be ~12 million years old based on ribulose-bisphosphate carboxylase large subunit (*rbcL*), nuclear (18S) rDNA and ATP Synthase B (*atpB*) sequences [39]. The tomato clade radiation has been estimated at ~7 million years before present, based on four nuclear genes [40].

The origins of cultivated tomato (*S. lycopersicum*) can be traced to South America, but native wild varieties are less attractive in shape and size than the domesticated cultivars taken to Europe in the mid-sixteenth century. Rick [41] points to Mexico as the probable region of domestication and the word "tomatl" in the Nahuatl language of Mexico is doubtless the origin of the modern name.

This ecotype still grows in the wild state in Central America, producing small, cherry-like fruits. Thus, it is known commonly as the cherry tomato. Since domesticates were known to be cultivated in Central America, the lack of a genetically similar ecotypes in South America suggests that domestication took place probably only in the north (Mexico). Taken together, it seems well founded that initial domestication of tomato occurred in Central America.

The wild ecotype *S. lycopersicum var. cerasiforme* was extensively cultivated

throughout Central America when the first conquistadores arrived to Yucatan. *S. lycopersicum var. cerasiforme* is thought to be the direct ancestor of cultivated tomato, based on its wide presence in Central America and the presence of a shortened style length in the flowers [42].

The morphology of the cherry and the modern domesticated tomato differs primarily by the larger fruit size of the latter. The wild *var. cerasiforme* is naturally self-pollinated like cultivated tomato. Self-pollination in the cherry tomato is promoted by a stigma and style that protrudes only a short distance or not at all, beyond the end of the anther cone [43] [44]. The gradual shortening of style length favouring self-pollination, is believed be one of the major features of domestication. Besides the uniformity associated with inbreeding, a shorter style and recessed stigma position tend to favour fruit setting ability and therefore the yield.

Early cultivars of North America tended to have slightly exserted stigmas. This fact caused reduced fruit setting, particularly in the absence of appropriate insects for the pollination or under conditions of excessive temperature (that generally reduces the pollen fertility). Selection for less exserted stigmas resulted in stigmas approximately flush with the mouth of the anther tube, as seen in many early European cultivars. Full enclosure of the pistil by the anthers was accidentally selected during breeding of mechanically harvested processing tomatoes for California conditions, apparently because it ensured a reliable, concentrated fruit set. This improvement virtually guarantees self-pollination and has been bred into many modern cultivars [45].

1.2.2 Taxonomy

The tomato was originally named *Solanum lycopersicum* by Linnaeus in 1753. In 1768 Philip Miller in *The Gardeners Dictionary* used *Lycopersicon esculentum* and this became the accepted name until very recently. Various lines of evidence had suggested that the tomato belonged to the genus *Solanum* and molecular studies provided the compelling evidence that has led to the readoption of *S. lycopersicum* [46]. The other species of *Lycopersicon* have also been assigned or re-assigned to *Solanum*. More recent studies on the tomato variability, including genetic and molecular markers [47] [48] [49] have provided additional evidence for the inclusion of the tomato in the *Solanum* genus. Therefore, *Lycopersicon esculentum* Mill. has been renamed *Solanum lycopersicum* L. The tomatoes are placed in groups very close to the tuberous and non-tuberous potatoes. This has not led to major changes in tomato taxonomy, but it does make very clear the relationship of tomatoes and potatoes.

1.2.3 Phylogenetic relationships

Based on phylogenetic relationships, characters of morphology, and geographic distribution, the cultivated tomato (*Solanum lycopersicum*) and its wild relatives form a small clade of 13 closely related species, the *Solanum* sect. *Lycopersicon*. Four other *Solanum* spp. are the closest outgroup species to the tomato clade: *S. juglandifolium*, *S. lycopersicoides*, *S. ochranthum*, *S. sitiens* (Fig. 8) [50]. From the highly polymorphic green-fruited species *L. peruvianum* (old classification) have

been segregated four species (*S. arcanum*, *S. huaylasense*, *S. peruvianum*, and *S. corneliomulleri*) (Tab. 1). The first two have been described by Peralta [49] as new species from Perú, while the latter two had already been named by Linnaeus (1753) and MacBride (1962), respectively. From *L. cheesmaniae*, two new yellow- to orange-fruited species, *S. galapagense* and *S. cheesmaniae*, both endemic to the Galápagos Islands, were segregated [51] [52].

Despite numerous studies, resolving evolutionary relationships within the tomato group has been difficult due to its young age, which increases the probability of shared polymorphism among groups, and the low levels of variation within and between the self-compatible taxa. Recent results based on sequences of the structural gene *GBSSI* [50], morphology [49] and AFLPs [53] illustrate the current phylogenetic relationships within the group that is schematically represented in Fig. 8.

a. The tomatoes belonging to the sections *Lycopersicoides*, *Juglandifolia* and *Lycopersicon* have a common ancestor and are sisters to the potatoes.

b. Section *Lycopersicoides* (formerly recognized as a subsection of sect. *Lycopersicon*) is monophyletic and sister to sect. *Juglandifolia* + sect. *Lycopersicon*.

c. Section *Juglandifolia* is monophyletic and sister to sect. *Lycopersicon*.

d. Within sect. *Lycopersicon*, *S. pennellii* in most cases appears at the base of the trees as a polytomy with *S. habrochaites*, or sometimes forms a clade with this species. This relationship is yet unresolved, but the morphological data suggest that *S. pennellii* is sister to the rest of the tomatoes (sect.

Lycopersicon); it is the only species that lacks the sterile anther appendage character shared with the outgroups.

e. The self-compatible green-fruited species *S. chmielewskii* and *S. neorickii* are closely related to *S. arcanum* and constitute a monophyletic group.

f. The four species with coloured fruit (*S. lycopersicum, S. pimpinellifolium, S. cheesmaniae* and *S. galapagense*) unambiguously form a closely related monophyletic group.

1.2.4 Wild tomatoes

The most important public tomato germplasm collections are in the U.S. The collections are maintained in Geneva, N.Y by the USDA (United States Department of Agriculture: http://www.usda.gov/wps/portal/usdahome) and in Davis, California by the TGRC (C.M. Rick Tomato Genetics Resource Center: http://tgrc.ucdavis.edu/). The USDA and TGRC maintain together approximately 9,355 accessions of tomato (among which wild species, mutants, ecotypes, cultivars etc.) [54]. In general, the status of wild tomato germplasm is considered good and vastly better than that of many other crop plants. Each species is represented by many accessions; the collections are, however, poor in accessions coming from certain hard to access areas.

The wild tomato species are perennial and most of them flower regardless of day length, although in their natural habitats some wild tomatoes behave as annuals, probably because frost or drought kills the plants after the first growing season [43]. Is interesting to note that some of them can flower only in short days if

growth at mid-latitudes (http://tgrc.ucdavis.edu/spprecommed.html) (Tab. 2).

Domesticated species represent only a small fraction of the variability available among their wild relatives. Plant domestication represents an accelerated evolution form, producing changes in the tissues and organs of greatest interest to humans (for example tubers, seeds and roots). One of the most extreme cases has been the evolution of the tomato fruit. The modern cultivated tomato plants produce fruits up to 1,000 times larger than those of their wild progenitors [55] [18] (Fig. 9).

The wild tomatoes and their relatives are native to the western South America (Fig. 10). In general they are plants of dry areas, with the exception of the members of the sect *Juglandifolia* (*S. jugandifolium* and *S. ochrantum*), those are found in cloud forests (Fig. 11). Species of the sect *Lycopersicon*, which includes the modern tomato, and *Lycopersicoides* are dry habitat plants and occur in the inter-Andean valleys that are subject to severe rain shortage (*S. arcanum*, *S. chmielewskii*, *S.neorickii*, *S. huaylasense*), in the extremely dry high-elevation deserts of the western Andean slope (*S. lycopersicoides*, *S. sitiens*, *S. pennellii*, *S. corneliomullerii, S. chilense*) and in the unique *lomas* habitat along the Pacific coast of Peru and north of Chile (*S. chilense*, *S. habrochaites*, *S. pennellii*, *S. peruvianum*, *S. arcanum*, *S. pimpinellifolium*). The *lomas* formations are small areas of vegetation, occurring like islands in a sea of hyper-arid desert.

In the Galapagos Islands, *S. cheesmaniae* occurs from sea level to the volcanic peaks in dry rock areas, instead *S. galapagense* is found in lower elevation habitats [56] [57].

Thanks to their evolutionary and adaptive history, the wild tomato species contain useful traits that can be introgressed into cultivated tomato, such as tolerance to drought and salinity (*S. pennellii* [58] [59]) and sucrose accumulation (*S. habrochaites* {Miron, 1991 #164). In particular, resistance to multiple pathogens has been discovered in many wild species [60] [42] [61] such as *S. cheesmaniae* [62], *S. neorickii* [63], *S. peruvianum*, *S. chilense*. Notably *S. lycopersicoides* is known for its broad suite of disease resistance [64], including *Botrytis cinerea*, all races of *Xanthomonas campestris*, *Phytophthora parasitica*, cucumber mosaic virus (CMV) and tomato yellow leaf curl virus (TYLCV).

Many prebred populations (Introgression Lines and Recombinant Inbred Lines), that capture the genomes of related wild species via overlapping chromosome segments in the genetic background of the cultivated tomato, have been developed and accessioned by TGRC. Recombinant inbred lines (RILs) are produced by crossing two parents, followed by repeated self-mating, to produce a new inbred line whose genome is a mosaic of the genomes of the parents. On the other side the Introgression Lines (ILs) are originated by crossing the F_1 generation with one of the two parents (usually the cultivated species) to generate a set of recurrent parent lines with single introgressed segments of one donor (usually the wild species) on a genetic background of the other parent.

The first resource was a library of *S. pennellii* ILs (Fig. 12), synthesized by Eshed and Zamir [65] that capture the entire donor genome, S. pennellii in 50-75 lines. Similar, though less complete, populations are also available for the genomes of *S. habrochaites* (RILs) [66], *S. lycopersicoides* (ILs) [64], *S. pimpinellifolium* (ILs) [67],

S. cheesmaniae (RILs) [68], *S. chmielewskii* (ILs) [69] and *S. neorickii* (ILs) [70].

1.2.5 Character evolution

1.2.5.a Genetic diversity in wild tomatoes: the evolution of fruit colour

Wild tomato species have tiny berries with different size depending on the species: the diameter of the fruits varies from 0.6 cm of *S. galapagense* to 1.15 cm of *S. habrochaites*. The berries are in general globose, bi-locular, sometimes with densely sparsely pubescent trichomes (such as *S. habrochaites* and *S. pennelli*) [56] (Fig. 13).

The evolution of carotenoid pigments affecting berry colour is one of the variable characters within the tomato clade that have obtained much attention, to understand the evolutionary dynamics of the group. The fruit colour of wild tomatoes varies from green, yellow and orange to red. Numerous studies have demonstrated that green fruit is the ancestral character and that coloured fruits arose once within the clade [71] [72] [73] [74] [50] and their origin is believed to be recent, about ~1million years ago [40]. The four tomato species with coloured berry are *S. lycopersicum*, *S. pimpinellifolium*, *S. cheesmaniae*, and *S. galapagense*. The first two species produce red berries, and accumulate lycopene; instead the fruit colour of *S. cheesmaniae* and *S. galapagense,* is commonly attributed to the accumulation of β-carotene and can vary from yellow to deep orange.

The reason behind a transition to coloured fruits during the evolution of the group is still not clear, but probably is related to attraction of animals for seed dispersal.

In literature not many data are available regarding the biochemical composition of

the wild berries and, above all, the carotenoid composition. Accumulation of carotenoids as secondary metabolites in many fruits and flowers occurs by up-regulating the pathway at the gene expression level.

A hypothesis exists to explain the different berries colour of these species. This hypothesis has been formulated on the basis of interspecific crosses between modern tomato and wild species [75] [76] [77]. The green-fruited species *S. pennelli* was found to carry the recessive *r* allele, which causes a deficiency in carotenoid accumulation because of a deficiency in *PSY1* (the chromoplast-specific phytoene synthase). In this species the dominant *Del* allele is also present: this allele encodes a hypermorphic form of lycopene ε-cyclase (LCY-e). The introgression of *Del* in *S. lycopersicum* causes orange fruit production because of a δ-carotene accumulation [78].

In the wild species *S. cheesmaniae,* the dominant *B* allele encodes a chromoplast-specific lycopene β-cyclase (CYC-b), which is up-regulated at the mRNA level with respect to *b,* the *S. lycopersicum* allele. The introgression of *B* in *S. lycopersicum* causes β-carotene accumulation in fruits [79].

The origin of many alleles responsible for different fruit colours (*r*, *B*, *Del*) from wild tomato species is a starting point for research on fruit colour evolution before and during domestication. The current hypothesis indicates that in the most ancient species, those with green berries, the berry carotenoid composition is similar to leaves. Subsequently, hypomorphic mutations in *LCY-e* and one of the *Chy* genes could have led to the downregulation of these genes, and to the appearance of the orange fruited species (*S. cheesmaniae* and *S. galapagense*)

accumulating β-carotene. A further hypomorphic mutation in *CYC-b*, could have caused the accumulation of lycopene and the emergence of the red-berry species (*S. pimpinellifolium* and *S. lycopersicum*). Not many data are available on the metabolic profiles of the berries of the wild species and, above all, on their carotenoid composition.

The biochemical composition of some wild species has been investigated recently: Schauer {Schauer, 2005 #173} has led a GC-MS based survey of primary metabolite levels in leaves and fruit of *S. lycopersicum* and its five wild interfertile species (*S. pimpinellifolium*, *S. neorickii*, *S*. *chmielewskii*, *S. habrochaites* and *S. pennellii*).

The main features analysed were the contents of organic acids, sugars, sugar alcohols and amino acids, but the carotenoid composition has not been investigated. The results show that the leaves and the fruits of wild species present a tremendous variability in metabolite content and dramatically higher levels of secondary metabolites with respect to modern tomato.

The overall antioxidant content (AO: including lycopene and vitamin C) and the overall antioxidant activity (AOA) have been investigated in *S. pimpinellifolium* [80]. The results shown that the analysed accessions of *S. pimpinellifolium* had an antioxidant activity (ARP: anti-radical power), an inhibition of lipid peroxidation (ILP), and a concentration of lycopene, ascorbic acid, phenolics, and soluble solids higher then the modern tomato cultivars analysed. On the other hand, it is known that the fruit size is negatively correlated with the ARP, ILP and total soluble solids content indicating that it is difficult to fuse large fruit size and high overall antioxidant activity [81].

1.2.5.b Mating system evolution

All members of the genus have perfect (hermaphroditic) flowers (Fig. 14). *S. lycopersicum*, the modern tomato, is self fertile, while a large part of the wild tomatoes is self-incompatible (SI) but with some exceptions.

The length variation of the flower style is, apart from increased fruit size, one of the major features of domestication. This organ shows a gradual shortening from very long and prone to outcrossing, to very short and outcrossing-inhibitive. The full enclosure of the pistil by the anthers is a trait that practically guarantees self-fertilization. The self-incompatibility and allogamy have played an important role in the evolution of this clade and the loss of SI has occurred independently several times. In particular, SI was lost in some populations of *S. pennellii* and *S. habrochaites* [82] [83], the coloured-fruited wild tomatoes *S. pimpinellifolium*, *S.cheesmaniae*, *S. galapagense* and finally *S. chmielewskii and S.neorickii* [84]. In this latter group it seems that the loss of SI has a monophyletic origin [72], [85]. Lower levels of genetic variation are expected in self-compatible species as compared to their SI relatives. Various molecular markers (allozymes [73], SSRs [74], nuclear DNA sequences [86] [87] etc.) consistently have demonstrated the lower levels of genetic variation in SC wild tomato species.

1.3 The tomato genome sequencing project

Sequencing of a genome is an important step towards its understanding, and represents a valuable tool to find genes, to understand how the genome as a whole works, that is how genes work together to direct the growth, development and maintenance of an entire organism, and to know the part of the genome outside the genes, which includes the regulatory regions that control how genes are turned on an off.

In the last 15 years, the genomes of a large number of organisms have been sequenced, from viruses up to humans. Moreover, the employment of new sequencing approaches, based on massively parallel sequencing [88], [89], has dramatically improved throughput. The range of plant genome sizes is very large, extending from 150 Mb (Arabidopsis) to over 100 Gb for some ornamental flowers. Despite this hurdle a number of plant genomes have been sequenced and many other are in progress (http://www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html).

The tomato genome is approximately 950 Mb in size, is diploid and it is divided in 12 chromosomes. The genome is constituted by a gene-poor pericentromeric heterochromatin (about 75%) and distal, gene-rich euchromatin [90]. The heterochromatic part, consists mostly of repetitive sequences and it is difficult to assemble. Therefore the initial strategy was to sequence the euchromatic portions of the genome, which is estimated to make up one-quarter (220 Mb) of the tomato genomic sequence, including >90% of the genes [91].

The tomato sequencing project, launched in November 2003, is part of a larger initiative called the "International *Solanaceae* Genome Project (SOL): Systems

Approach to Diversity and Adaptation": an international consortium of 10 countries, including Italy, began this initiative using initially a BAC-by-BAC approach to generate a high quality sequence for the euchromatic portions.

This strategy has involved the anchoring of BACs or contigs of BACs to a reference genetic map. These anchored BACs have been sequenced and the sequence information has been used to extend the BAC contigs further (BAC walking). Gaps between BAC contigs have been closed by targeting novel markers or BACs to these gaps, followed by further rounds of BAC walking. The reference genetic map is the F_2–2000 map [92] based on 80 F_2 individuals from a *S. pennellii* - *S. lycopersicum* cross [3]. BAC libraries have been constructed from the Heinz 1706 tomato line using different enzymes *(Hin*dIII, *Eco*RI and *Mbo*I). A fosmid library is also available.

However, the BAC-by-BAC strategy is more expensive and slower than the whole genome shotgun sequencing (WGS) approach; moreover "next generation" sequencing technologies are now available for the sequencing of complex genomes. For this reason, at the end of 2008 a "Next Gen Sequencing Initiative" was launched, in order to produce a whole genome shotgun draft sequence of the tomato combining BAC Sequences (Sanger sequences) 454 reads (shotgun and paired-end runs), SOLID and Illumina reads. In little more than one year, an assembly of the tomato genome was released, including over 90 Gb of sequence data and a high quality annotation prediction (http://solgenomics.net/index.pl).

The tomato genome is not the only sequenced genome belonging to the *Solanaceae* genus. In fact, the potato genome has been also released at the end of 2009

(www.potatogenome.net). Due to the high level of macro and micro-synteny in the *Solanaceae*, both genomes can be used to investigate specific genetic regions or biosynthetic pathways. Our laboratory is part of both international consortia: it had an active role during the sequencing and now takes part in the validation of the generated scaffolds, and in the annotation. A preliminary paper on the tomato genome has been published in the course of 2009 [9].

1.4 Carotenoids and apocarotenoids

Carotenoids are isoprenoid pigments that are naturally synthesized in photosynthetic organisms like plants and algae, as well as some fungi, bacteria and archaea (Fig. 15). Animals are incapable of synthesizing carotenoids, and must obtain them through their diet. Food carotenoids are usually C40 tetraterpenoids built from eight C5 isoprenoid units, joined so that the sequence is reversed at the center. The basic linear and symmetrical skeleton has lateral methyl groups separated by six C atoms at the center and five C atoms elsewhere and it can be cyclized at one or both ends. Cyclization and other modifications (such as isomerization, double-bond migration, hydrogenation, dehydrogenation, chain shortening or extension rearrangement, introduction of oxygen functions etc.) result in a myriad of structures.

The more stable isomeric form of carotenoids in nature is the *all-trans* form, but *cis* isomers do occur of biosynthetic precursors, along with derivatives of the main

components. Carotenoids made up of only carbon and hydrogen are called carotenes (hydrocarbon carotenoids such as α-carotene, β-carotene and lycopene) while carotenoids that contain oxygen are termed xanthophylls.

The colour of these molecules ranges from pale yellow (zeaxanthin and lutein) through bright orange (α-, β- and ϒ-carotene) to deep red (lycopene and astaxanthin). This variability is due to their structure and to the fact that double bonds involving adjacent carbons interact with each other (the conjugation process) which allows electrons in the molecule to move freely across these areas of the molecule. As the system of conjugated double bonds (the cromophore) increases in length, the electrons associated with it require less energy to change states. This, in turn, increases the absorption in the short wavelength region of the visible spectrum, and the compounds acquire an increasingly red appearance [93].

Another very important group of compounds is that of apocarotenoids [94]: these metabolites derived from the enzymatic (oxidative) cleavage of carotenoid molecules [95]. Apocarotenoids are important flavours in diverse food products [96]. For example, β-damascenone, in addition to tomato, is found in berries, apples, and grapes (as well as wine). Safranal, that gives the aroma in saffron (with picrocrocin and crocin which give, respectively, its bitter taste and colour), grapefruit, and green tea, is derived from the carotenoid zeaxanthin. Likewise, dihydroactinidiolide and 4-oxoisophorone are flavour components of carotenoid origin found in tea, tobacco, lemon balm and saffron [97]. Many apocarotenoids are produced in plant tissues at specific developmental stages, for example, during flowering, ripening, or maturation. Although a single fruit or vegetable synthesizes

several hundred volatiles, only a small subset generates the “flavour fingerprint” that helps animals and humans recognize appropriate foods and avoid poor or dangerous food choices.

1.4.1 Biological functions of carotenoids

Most of the functions and the applications and uses of carotenoids are a consequence of the light-absorbing properties of the polyene chromophore. Natural roles in colouration, photosynthesis and photoprotection are well established and of major biological importance (Fig. 15).

Carotenoids are the most widespread group of pigments found in nature. In plants, carotenoids are involved in numerous physiological processes and have different functional roles in development, photosynthesis and membrane stability [98]. The mainly functions are:

a. Ecological: carotenoids and apocarotenoids provide flowers and fruits with distinct colours and scents that are designed to attract animals promoting pollination or seed dispersion [99] or antagonists of plant pathogens [100]. Such apocarotenoid flavours have high commercial value, like safranal and picrocrocin from saffron [101] and β-damascenone from rose [97].

b. Photosynthetic (photoreception and photoprotection): in photosynthetic organisms, carotenoids play vital role in the photosynthetic apparatus. They participate in the light energy-harvesting and energy-transfer process, and protect the reaction centre from auto-oxidation. Specifically, xanthophylls

are accessory pigments in the light-harvesting antennae of the chloroplast [102], which are capable of transferring energy to the chlorophylls, capturing light between 400nm and 500nm. They also quench triplet excited states in chlorophyll molecules by dissipating the excess excitation energy in a non-radiative manner, a process known as nonphotochemical quenching (NPQ). This function is crucial to protect against chlorophyll bleaching [103], [104]. The carotenoid complement of light-harvesting complexes (LHCs) (neoxanthin, violaxanthin and lutein) is well conserved during plant evolution. Zeaxanthin and violaxanthin are an integral part of the xanthophyll cycle, which is involved in the dissipation of excess excitation energy as heat. Excess light energy absorbed by the antennae chlorophyll in photosystem II is safely dissipated as heat in a process that depends on zeaxanthin, whose synthesis is triggered by the proton gradient across the thylakoid membranes, that activates de-epoxidation of violaxanthin to zeaxanthin [104].

c. Metabolic and developmental: carotenoids are precursors of the hormone ABA [105] and of apocarotenoids, which can act as signalling molecules within the plant and with other plants or in aroma production. Notably a group of apocarotenoids, named strigolactones, was known to trigger the germination of parasitic plant seeds and stimulate symbiotic fungi. Recent studies [106] [107] [94] have demonstrated that these molecules are a new class of plant hormones that are implicated in plant architecture and they have also a function in underground communication with other

neighbouring organisms.

d. Defence against biotic/abiotic stresses: apocarotenoids have been implicated during plants and fungi interactions [108]. Tobacco plants treated with β-ionone are protected against the infection caused by the pathogenic fungus Peronospora [109]. Moreover blumenin, a glycosylated C13 apocarotenoid, inhibits fungal colonization and arbuscule formation during the initial steps of mycorrhiza development in barley and wheat [110].

1.4.2 Nutritional properties

The most important nutritional role of carotenoids, (specifically of α- and β-carotene), is as vitamin A precursors [111] [112]. The involvement of the vitamin A aldehyde, retinal, as a chromophore of the visual pigments in the eye is central to the process of vision. Vitamin A deficiency is still a major nutritional problem in many developing parts of the word where its consequences, xerophthalmia, blindness and premature death, are still too common, particularly in children [113]. Vitamin A also has important systemic functions in maintaining growth and reproductive efficiency and in the maintenance of epithelial tissues and prevention of their keratinization. The importance of the latter effect has led to the synthesis of a wide range of related compounds, the retinoids, and to the evaluation of these substances for therapeutic use to treat skin problems such as acne also as cancer-prevention agents.

More recently, both statistical correlation and model experiments with animal

systems have provided evidence that carotenoids may give protection against some life-threatening conditions such as cancer [114] [115] [116], heart diseases [117] and AIDS [118]. It is suggested that these effects are not related to the formation of vitamin A, so major dietary carotenoids such as lycopene and lutein, which do not have pro-vitamin A activity, may be as effective as β-carotene. The commercial demand for carotenoids is mainly met by chemical synthesis and to a minor extent by extraction from natural sources. However, the supply is restricted to few carotenoids. Some important dietary carotenoids are not abundant in our food and cannot be taken as supplements. Zeaxanthin, for example, is a rare carotenoid, which, together with lutein, is the essential component of the macular pigment in the eye [119], [120]. A low level of carotenoid intake increases the risk of age-related macular degeneration.

1.4.3 Carotenoid Biosynthesis

Biosynthesis of carotenoids occurs in all photosynthetic organisms, as bacteria, algae and plants, as well as in some non-photosynthetic bacteria and fungi. In plants, it takes place in chloroplasts (in which they play key roles in photosynthesis) and in chromoplasts (where they act as secondary metabolites) (Fig. 16). The intermediate steps in the carotenoid biosynthesis pathway were postulated several decades ago by standard biochemical analyses using labelled precursors, specific inhibitors and characterization of mutants. The enzymes of this pathway exist in minute amounts and are very labile upon purification. These characteristics and the lack of genuine *in vitro* assays for any of the enzymes have

hindered the usage of conventional biochemical investigation. In plants, carotenoids are synthesized within the plastids by enzymes that are nuclear encoded. A unique exception to this rule has been discovered in the green alga *Haematococcus pluvialis* in which the last steps in the synthesis of the ketocarotenoid astaxanthin take place in cytoplasmic lipid vesicles [121].

In recent years, molecular-genetic approaches to the study of carotenogenesis have provided a wealth of information and new perspectives of both the enzyme activities and the regulation of the pathway [99] [122].

Like all other isoprenoids, carotenoids are built from the 5-carbon compound isopentenyl diphosphate (IPP). In plastids, IPP is produced via in the 'DOXP pathway' from pyruvate and glyceraldehyde-3-phosphate [123], [124], [125]. The first enzyme in this pathway is 1-deoxyxylulose 5-phosphate (DOXP) synthase (DXS), which is encoded by a gene that has been cloned from different species: Arabidopsis and tomato. DXS is impaired in the *chilling-sensitive5* (*chs5*) mutant of Arabidopsis. At the restrictive temperature, chlorotic leaves develop in young leaf tissues of this mutant but not in mature leaves, indicating that DXS functions preferentially at an early stage of leaf development [126]. It has been suggested that DXS could potentially be a regulatory step in carotenoid biosynthesis during early fruit ripening in tomato [127]. DOXP is converted to 2C-methyl-D-erythritol 2,4-cyclodiphosphate via 2C-methyl-D-erythritol 4-phosphate, 4-diphosphocytidyl-2C-methyl-D-erythritol (DPME) and 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate. These steps are catalyzed by the enzymes DOXP reductoisomerase (DXR) [128], [129], DPME synthase (ispD [YGB-P]), DPME

kinase (ISP-E) and 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (ISP-F), respectively [130], [125]. Additionally, an enzyme encoded by the *Isp-H* gene, catalyzes the subsequent reaction that affects the ratio of IPP to dimethylallyl diphosphate [131]. IPP is isomerized to dimethylallyl diphosphate by the enzyme IPP isomerase (which is encoded by the *Ipi* genes). There are two *Ipi* genes in plants and one of them is specific to the plastids [132] [133]. The sequential addition of three IPP molecules to dimethylallyl diphosphate, which is catalyzed by a single enzyme, geranylgeranyl diphosphate synthase (GGPS), gives the 20-carbon molecule GGPP.

The first committed step in the carotenoid pathway is the condensation of two GGPP molecules to produce 15-cis phytoene, which is catalyzed by a membrane-associated enzyme, phytoene synthase (PSY in eukaryotic cells and CrtB in bacteria) (Fig. 17) [134]. PSY shares amino-acid sequence similarity with GGPP synthase and other prenyl-transferases. Partial purification of PSY from tomato indicated that the enzyme is associated with the isoprenoid biosynthesis enzymes IPI and GGPS in a protein complex that is larger than 200 kDa [132]. In tomato, there are two genes for PSY: *PSY1*, which encodes a fruit- and flower-specific isoform, and *PSY2*, which encodes an isoform that predominates in green tissues [135], [136]. PSY is a rate-limiting enzyme of carotenoid biosynthesis in ripening tomato fruits [137], [138]. Two structurally and functionally similar enzymes, PDS and ζ-carotene desaturase (ZDS), convert phytoene to lycopene via ζ-carotene. These FAD-containing enzymes each catalyze two symmetric dehydrogenation reactions that require plastoquinone and a plastid terminal oxidase as electron

acceptors [139], [99]. When co-expressed in *E. coli*, PDS and ZDS from Arabidopsis convert phytoene to poly-*cis* lycopene, which is also called 'pro-lycopene', whereas the bacterial phytoene desaturase (CrtI) produces all-*trans* lycopene [140]. The next step in the plant pathway involves a carotene isomerase (CrtISO) [141], [142]. This enzyme has an activity of *cis*-to-*trans* isomerization, which requires redox-active components, suggesting that isomerization is achieved by reversible redox reaction acting at specific double bonds. CrtISO isomerizes adjacent *cis*-double bonds at C7 and C9 pairwise into the *trans*-configuration, but is incapable of isomerizing single *cis*-double bonds at C9 and C9'. This means that CrtISO functions in the carotenoid biosynthesis pathway in parallel with ζ-carotene desaturase, by converting 7,9,9'-tri-cisneurosporene to 9'-*cis*-neurosporene and 7'9'-di-*cis*-lycopene into all-*trans*-lycopene.

Cyclization of lycopene marks a branching point in the pathway: one branch leads to β-carotene and its derivative xanthophylls, whereas the other leads to α-carotene and lutein. Lycopene β-cyclase (LCY-b in eukaryotic cells and CrtY in bacteria) catalyzes a two-step reaction that creates one β-ionone ring at each end of the lycopene molecule to produce, firstly, ϒ-carotene and, finally, β-carotene, whereas lycopene ε-cyclase (LCY-e) creates one ε-ring to give δ-carotene. LCY-b even acts to produce α-carotene (β, ε-carotene) from δ-carotene. There is a high degree of structural resemblance, 30% identity in amino-acid sequence, between LCY-b and LCY-e in both tomato and Arabidopsis. The two enzymes contain a characteristic FAD/NAD(P)-binding sequence motif at the amino termini of the mature polypeptides. In tomato, there are two lycopene β-cyclase enzymes, LCY-b

[143] and CYC-b (chromoplast-specific lycopene cyclase) [79], whose amino-acid sequences are 53% identical. LCY-b is active in green tissues, whereas CYC-b functions only in chromoplast-containing tissues.

Interestingly, the amino-acid sequence of CYC-b is more similar (86.1% identical) to that of capsanthin-capsorubin synthase (CCS) from pepper, an enzyme that converts antheraxanthin and violaxanthin to the red xanthophylls capsanthin and capsorubin, respectively, than to that of LCY-b. A deletion mutation in the *Ccs* gene (locus *y*) that results in the accumulation of violaxanthin is responsible for the recessive yellow fruit phenotype of pepper [144]. CCS exhibits low lycopene β-cyclase activity when expressed in *E. coli* [132]. Similarities in function, gene structure and map position strongly suggest that the genes *Ccs* and *CYC-b* orthologs and that they are both paralogs to *Lcy-b*. Tomato CYC-b has retained its original catalytic function, whereas the second cyclase in pepper acquired a new enzymatic activity of a similar biochemical nature during evolution [145]. Conservation of amino-acid sequences as well as their similar mechanisms of catalysis suggests that all plant cyclases, including CCS and perhaps also neoxanthin synthase, have evolved from a common ancestor, most probably the cyanobacterial *CrtL* [99].

Hydroxylation of cyclic carotenes at the 3C, 3'C positions is carried out by two types of enzymes: ferredoxin dependent enzymes (β-carotene hydroxylase 1 and 2) and cytochrome p450 enzymes (LUT1 and LUT5) [146], [147]. The β-carotene hydroxylases (CHY-b 1 and 2 in plants; CrtZ in bacteria) require iron and are characteristic of enzymes that exploit iron-activated oxygen to oxygenate

carbohydrates [148]. Consequently, in the β, β- branch, they convert β-carotene to zeaxanthin via β-cryptoxanthin. There are two β-carotene hydroxylases in both Arabidopsis and tomato. In tomato, one hydroxylase is expressed in green tissues (CHY1) while the other is expressed in the flower (CHY2) [149] [99] [150].

Cytochrome p450 enzymes were isolated in Arabidopsis. LUT1 is involved in the conversion of zeinoxanthin into lutein [151]; the double *chy1chy2* Arabidopsis mutant can still produce zeaxanthin while the triple mutant *chy1chy2lut5* abolishes totally β, β- xanthophylls, indicating the role of LUT5 in β-ring hydroxylation [147]. In this triple mutant, hydroxylation of the β-ring of α-carotene is still observed, suggesting the existence of an additional hydroxylase, acting on the β-ring of α-, but not β-carotene: a candidate for this activity is LUT1.

The enzyme ZEP converts zeaxanthin to violaxanthin via antheraxanthin by introducing 5,6-epoxy groups into the 3-hydroxy- β-rings in a redox reaction that requires reduced ferredoxin [152]. In leaves, violaxanthin can be converted back to zeaxanthin by violaxanthin deepoxidase (VDE), an enzyme that is activated by low pH, which is generated in the chloroplast lumen under strong light. Zeaxanthin is effective in the thermal dissipation of excess excitation energy in the light-harvesting antennae and plays a key role in protecting the photosynthetic system from damage by strong light. The inter-conversion of zeaxanthin and violaxanthin is known as the 'xanthophyll cycle'. The *Vde* gene was originally cloned from lettuce. The amino acid sequences of ZEP and VDE indicate that they are members of the lipocalins, a group of proteins that bind and transport small hydrophobic molecules [99].

Genes for neoxanthin synthase (*NXS*) from potato (*Solanum tuberosum*) and tomato were cloned [153] [154]. Surprisingly, the amino-acid sequence of NXS from tomato is identical to that of the lycopene β-cyclase CYC-b. This led to the proposal that CYC-b is a bi-functional enzyme that is capable of converting both lycopene to β-carotene and violaxanthin to neoxanthin. In this case, another neoxanthin synthase must exist in tomato, because neoxanthin is synthesized in the mutants *old-gold* and *old-gold-crimson*, which carry null mutations in the *CYC-b* (*B*) gene.

1.4.4 Fruit colour tomato mutants

A large body of genetic data exists for simply inherited genes that influence carotenoid content in tomato. More than 20 genes have been characterized in tomato those influence the type, amount, or distribution of fruit carotenoids (Fig. 18). Many of the available colour variants were first identified as spontaneous mutants in cultivars of *S. lycopersicum*, but also occur in wild tomato species. The *Beta* (*B*) allele was first characterized in orange-fruited segregants derived from a cross between *S. lycopersicum* and *S. habrochaites* (wild green-fruited species) and it is located on chromosome 6 [155] [156]. The *B* allele encodes a CYC-b cyclase that converts lycopene to β-carotene [79]. The introduction of *B* from accessions of *S. galapagense*, *S. pimpinellifolium, S. chilense*, and *S. chmielewskii* has also been described [157], [158], [159]. The recessive *crimson* allele (*c*), in contrast to *B,* enhances lycopene content at the expense of β-carotene [160]. Ronen et al. [79] demonstrated that *c* is an allele of *B* and that null mutations in the *B* gene are

responsible for the crimson phenotype. *og* is a second allele of *B*, and similar to *crimson* enhances the lycopene content of the berry. The *crimson* cultivar has a dark red pigmentation and it has been developed for its capably to accumulate high lycopene quantities.

Trans-lycopene is the principal form of lycopene in red tomato fruits. The *tangerine* (*t*) mutant is recessive and its orange berry accumulates poly-*cis*-lycopene, also referred to as prolycopene [161]. The fruit also exhibits elevated phytoene and phytofluene levels. Located on chromosome 10, the *tangerine* gene was shown to encode a carotenoid isomerase, *CrtISO*, required during carotenoid desaturation [141]. Analysis of two alleles of *t* demonstrated that in one case, loss of function in CrtISO was attributable to a deletion mutation in *CrtISO*, and in the second, expression of this gene was impaired. *CrtISO* is expressed in all green tomato tissues but is up-regulated during fruit ripening and in flowers. It is demonstrated that the *cis* isoform of lycopene is more bioavailable than the *trans* [162], [163]: this fact has brought to a considerable interest on this mutant in human nutrition-related studies.

The dominant *Delta* (*Del*) allele increases fruit δ-carotene and reduces lycopene content, resulting in reddish-orange coloured fruit [164]. Ronen et al [78] demonstrated cosegregation of the *Lcy-e* locus encoding ε-cyclase with the *Del* mutation located on chromosome 12. ε-cyclase converts lycopene to δ-carotene. The transcript for *LCY-e* was shown to increase 30-fold in ripening fruit of the *Del* mutant Additional orange colour variants include the *diospyros* (*dps*) mutant with dusky orange fruit.

The *high pigment-1* (*hp-1*) and *high pigment-2* (*hp-2*) [165] mutant alleles enhances total fruit carotenoid content 30 to 50% without significantly altering the relative percentage of different carotenoid constituents [166].

The *hp-2* allele encodes the tomato homolog of the nuclear protein *DEETIOLATED1* (*DET1*) from Arabidopsis that is involved in light signal transduction [167]. The *hp-1* allele is a mutation in a tomato *UV-DAMAGED DNA-BINDING PROTEIN 1* (*DDB1*) homolog whose Arabidopsis counterpart interacts with *DET1*. The *dark green* (*dg*) mutant is an allele of *hp-2* and it is light hypersensitive [168]. Additional alleles of *hp-1* and *hp-2*, *w* and *j*, respectively, have been identified and they exhibit varying photoresponsiveness [169]. Plants expressing both *crimson* and high pigment alleles produce fruit with lycopene levels three to four times that of conventional red-fruited tomatoes. Often undesirable pleiotropic effects are associated with these mutants, limiting their practical use.

The high pigment-3 (*hp-3*) mutant has been characterized recently; this allele exhibits higher concentrations of carotenoids and chlorophylls in leaves and fruits. The mutation is found to occur in zeaxanthin epoxidase, which converts zeaxanthin to trans-violaxanthin [170].

The recessive *r* allele is another fruit colour mutant. It is located on chromosome 3 and is responsible for yellow fruit flesh, resulting in greatly reduced levels of polyenes and very low levels of coloured carotenoids [171]. This locus has been suggested to correspond to a null mutation for *PSY1* (the chromoplast-specific phytoene synthase). Two additional colour mutants are the recessive *apricot* (*at)* locus on chromosome 5 [172] and the *sherry* (sh) locus on chromosome 10 [173]

that results in fruit that are characteristically yellow but with a pinkish/red blush at maturity.

A block in the desaturation of phytoene results in the *ghost* (*gh*) mutant (on chromosome 11), *gh* contains a lesion in a plastid terminal oxidase, an essential cofactor in phytoene desaturation [174].

2. Aim of the work

The investigation of physiological processes in species with limited variation due to bottlenecks from domestication and subsequent breeding, has limited resources; however, the study of the genetic diversity available within related species could provide a useful tool for the analysis of genes involved in such processes. Our approach implies the identification of polymorphisms deriving from both coding and non-coding sequences affecting the functionality of a locus. The presence of wild interfertile species and its commercial importance renders tomato a good model for these studies.

Accessions of the *S. lycopersicum var. cerasiforme* as well as wild tomato species (*S. pimpinellifolium*, *S. cheesmaniae*, *S. neorickii*, *S. chmielewskii*, *S. chilense*, *S. habrochaites*, *S. pennellii* and *S. arcanum*) were selected to represent ancient varieties and closely related progenitor species. Within the gene pools of these species, much variability exists for various morphological and biochemical characters connected with frui development and quality.

We focused our attention on the carotenoid biosynthetic pathway. This has been characterized in tomato, but the speciation event that led to the development of red-fruited species is still unclear. For this reason, some genes of this pathway have been analyzed in the wild tomato species characterized by different berry colour, that vary from red to green in relation with genetic distance (red, orange and green). Carotenoids and apocarotenoids are one of the most important groups of secondary metabolites. In plants they have a set of functions, ranging from photoprotection to attraction of pollinators to antagonism against phytophages, to

hormone precursors, to aroma production.

The aim of the project is to identify the genetic differences, generated during tomato speciation, responsible for the variation in mature berry colour, and associate these differences with specific fruit carotenoid compositions and gene expression profiles.

3. Materials and methods

Unless differently indicated, molecular biology methods are as described in [175].

3.1 Plant material growth conditions

The species and the relative accessions of *Solanum* selected for this thesis are listed in Tab. 3. Seeds were obtained from the true-breeding monogenic stocks maintained by the TGRC (University of California, Davis).

Plants were grown in a greenhouse under controlled environmental conditions; cross-species pollination was prevented by the absence of wind shear and insects. Seeds of wild species were seen to be more recalcitrant to germination than those of cultivated tomato, so they were soaked in 2.7% sodium hypochlorite for 30 minutes and rinsed for 15 minutes in water before placing them on wet blotting paper in plastic dishes at 25°C. Plantlets were then moved to greenhouse and planted in pots (diameter: 25 cm) with Vigorplant TN soil.

In the first month, in order to encourage the vegetative development, photoperiod was set at "long day" conditions (16 hours of light and 8 hours of dark), with temperature set at 25°C during the light period and at 20°C during the dark. Since some species were reported to shift from vegetative to reproductive phase only in "short day" conditions, the light period was later set to 8 hours. Allogamous species have been manually self-pollinated. Watering was reduced during fruit production, with the exception of *S. habrochaites*, where water reduction started from the early vegetative stage, in order to prevent leaf edema.

3.1.2 Sample collection

Tissues from 3 plants for each species were collected and used as fresh tissue or stored at -80°C for later experiments. Leaves were collected for genomic DNA and total RNA extractions. Berries were staged measuring days post-anthesis (DPA). Time points were set to correspond to the visible berry maturation stages of *S. lycopersicum.* The selected time points (44, 51, 58 and 65 DPA) correspond, in cultivated tomato, to the following developmental stages: mature green, breaker, red ripe and over ripe (Fig. 19). Tissues were dissected and snap-frozen in liquid nitrogen. Midribs were removed from leaves and seeds and locular tissue from fruits prior to freezing.

3.2 Nucleic acids protocols

3.2.1 Genomic DNA extraction

It was performed with a DNeasy Plant Kit (Qiagen) from fresh or frozen leaf tissue following the standard protocol (Qiagen Mini prep catalog n° 69104 and Qiagen Maxi prep catalog n°68163).

3.2.2 Analytical PCR primer design

Primers design was performed using Primer3 (http://frodo.wi.mit.edu/primer3/input.htm) on the sequences of candidate genes

availale on GenBank (http://www.ncbi.nlm.nih.gov/) (Tab. 4). The pairs were chosen to produce amplicons of 800-1000 bp, suitable for subsequent Sanger sequencing. In some cases only the CDS (Coding Sequence) was known, so primers were designed on 2 exons to amplify the included intron. Primers were then manually checked with NetPrimer (http://www.premierbiosoft.com/netprimer/) for the possibility of hairpins or dimers formation between the pairs. In some cases primer pairs needed to be re-designed, due to polymorphisms between *S. lycopersicum* and its wild relatives. For the aim of the thesis, promoter regions were sequenced as well.

The primers size varied between 18 and 22 bases and their annealing temperature was set between 56 and 60 °C with a maximum melting temperature difference in a pair not exceeding 4°C (Tm calculation: 2°C for each A/T, 4°C for each G/C).

3.2.3 PCR assay

The reaction was performed in 20 ul volume per sample with the following final component conditions:

- General Electric (GE) Taq Buffer: 1X
- dNTP: 200 uM
- primer up: 50 ng/sample
- primer down: 50 ng/sample
- GE Taq polymerase (28-9373-45): 0,25 U/sample
- 2d H2O: to 20 ul final volume

Template: genomic DNA (15-20 ng), plasmids (0.01-0.1 ng) or a tip of a bacterial

colony.

Prepare enough the mix for n samples + negative control + positive controls + 1

Termocycler settings:

- Initial Step: 3' at 95°C (denaturation);

- 30 cycles of: 45" at 95°C (denaturation) + 45" at Tm-2°C (annealing) + X" at 72°C (elongation); (X = 45" up to 0,5 Kb + 30" for each more 0,5 Kb)

Purification of the PCR product was performed on Sepharose G100 (GE catalog n° 17-0060-02) and amplicons were quantified using a Nanodrop spectrophotometer.

3.2.4 RNA Maxi-prep Extraction

RNA isolation was performed as described in [176].

Solutions:

- Extraction buffer:

1M Tris pH 7.5 (adjusted with borate)	75ml (final conc 150 mM)
0.5M EDTA	50ml (final conc 50 mM)
SDS	10g (final conc 2%)
2d H_20	till 500ml

Before use add mercaptoethanol 0,1%.

- Other solutions:

LiCl 6M (DEPC-treated and autoclaved)

2d H_20 (DEPC-treated and autoclaved)

5 M Potassium Acetate (DEPC-treated and autoclaved)

5M NaCl (DEPC-treated and autoclaved)

RNA-grade plasticware

RNA-grade Ethanol and 80% Ethanol

Phenol-Chloroform 1:1 (ultrapure, equilibrated at pH 8).

1° day (afternoon):

1. Take 5 g of frozen tissue in a 50-ml Falcon. Add extraction buffer to 15 ml and 5 ml phenol/chloroform;

2. Omogenize extensively with UltraTurrax (IKA, Staufen, Germany);

3. Add 3 ml di Et-OH and 1.5 ml di 5M KAc. Vortex. Transfer in a 50-ml Oak-Ridge tube. Incubate on ice 30'-45';

4. Centrifuge 10' 15000 rpm 4°C in JA-17 or similar rotor;

5. Place 15 ml of aqueous phase in a clean-sterile tube and add 15 ml of 6M LiCl. Vortex extensively. Incubate over night on ice.

2° day (morning):

6. Centrifuge 10' 15000 rpm 4°C JA-17 or similar rotor;

7. Decant supernatant and add 2 ml EtOH 80% to the pellet (RNA). Vortex;

8. Centrifuge 5' 15000 rpm 4°C JA-17 rotor. Decant supernatant and take out last drop with P1000;

9. Resuspend pellet in 500ml H_2O (DEPC). Vortex extensively. Read the concentration with the NanoDrop spectrophotometer. Assume DNA is 20% of the total and digest the samples with DNase for 30' at 37°C using an enzyme concentration of 1U/ug of DNA;

10. Add 25 ml NaCl + 1 ml Et-OH. Vortex extensively and incubate on ice for 30';

11. Centrifuge 5’ 15000xg. Decant supernatant and take out last drop with P200;

12. Add 200ml di Et-OH 80%. Mix gently. Centrifuge 5’ 15000xg. Decant supernatant and take out last drop with P200;

13. Resuspend pellet in 200 ul H_2O.

Read the concentration with a Nanodrop spectrophotometer. Dilute all the RNA’s at the same concentration and run them on a formaldehyde gel.

3.2.5 RT-PCR

PCR Buffer III:

On ice, add the following to a sterilized 1,5 ml microcentrifuge tube:

25 mM MgCl2	100 ul
10X PCR Buffer II (Kit Perkin Elmer)	50 ul
10 mM dAGCT	50 ul
20 mg/ml BSA (for molecular biology)	1 ul
100 mM DTT	2 ul

Store at -20 °C. Defreeze at room temperature.

Reverse Transcriptase Reaction (for 20 reactions)

On ice mix the following in a PCR tube:

PCR Buffer III	8 ul
Ribonuclease Inhibitor	1 ul
50 U/ul MuLV Reverse Transcriptase	1 ul
50 uM oligo d(T)16	1 ul
Total RNA	0.5 ug

Add H2O for a total volume of 20 ul

RT cycle: 25 °C x 10'; 42 °C x 30'; 99 °C x 5'.

Add 20 uL H2O bid and store at –20°C.

3.3 Plasmid constructs

3.3.1 Bacterial strains

The following strains of *E. coli* were used:

- *E. coli* XL-1 Blue strain: recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proAB lacIqZΔM15 Tn10 (Tetr)] (Stratagene);

- *E. coli* XL-1 Blue (pFabeR) (XL-pFabeR), stably transformed with the plasmid pFabeR coding for bacterial *CrtE, CrtI* and *CrtB* genes that lead to the accumulation of lycopene (a gift of P. Beyer, Uni-Freiburg, [177]).

3.3.2 Plasmid DNA extraction

Plasmid DNA was extracted by mini-prep method using Nucleobond AX PC 100 Kit (Nucleobond, catoalog n° FC140573M).

3.3.3 Bacterial transformation

Bacterial transformation was performed by electroporation with the Micropulser electroporator (BioRad) according to the instrument manufacturer's protocol.

3.3.4 Cloning strategy

Fragments containing the CDS of *LCY-e* of *S. lycopersicum* and *S. arcanum* but lacking the chloroplast transit peptide (cTP), predicted by ChloroP (http://www.cbs.dtu.dk/services/ChloroP/), were cloned in the expression vector pQE-50-hys (3500 bp) (Qiagen) for expression in *E. coli*.

- cDNA were prepared from leaf total RNA and subsequently amplified using a proofreading polymerase (Phusion® High-Fidelity DNA Polymerase, Finnzymes catalog n° F-530L) with primers tailed with recognition sites for restriction enzymes (*BamHI* and *KpnI*) to ensure in-frame cloning in expression vectors;

- The fragments were cloning firstly in the pBlueScriptSK+ vector (Stratagene) previously digested with EcoRV (NEB, USA) and re-sequencing (intermediate cloning).

- The pBSCDS constructs were digested with *BamHI* and *KpnI* and isolated (QiaPrep Spin Miniprep, Qiagen catalog n° 27104) to originate the insert to clone into the pQE-50 expression, ensuring that the finally construct was in-frame with the promoter of the vector by sequencing.

The obtained constructs were then inserted into *E. coli* strains XL-pFabeR.

3.3.5 Induction of expression vectors in E. coli

The *E. coli* strains overexpressing lycopene (XL-pFabeR) were singularly transformed with plasmids pQE50/LCY-e (*S. lycopersicum*) and pQE50/LCY-e (*S. arcanum*). Five transformants per construct were pooled and inoculated in 25 ml of LB medium with 1% Glucose and antibiotics (100 μg/ml of Ampicillin and 50

μg/ml Kanamycin) and grown at 37°C overnight under continuous shacking (220 rpm).

The following day, glucose was removed and aliquots were inoculated in LB plus relevant antibiotics but without glucose. Samples were subsequently split into 5ml replicas and grown to an OD_{600} = 0.4 (Optical Density). The sample were induced with Isopropylthio-β-galactoside (IPTG) 1mM, and let grow o/n as above. The following day the cells were pelleted and stored at -20°C for carotenoid extraction.

3.4 Real-time PCR

Two independent RNA extractions and four cDNAs (two from each RNA) were used for the analyses.

3.4.1 Primer test

1. Amplification of the fragment through PCR assay (as described before) except for the PCR cycle that is:

35 cycles of: 15" at 95°C (denaturation) + 1' a 60°C (annealing + elongation);

2. Load 3 ul of the PCR reaction in a 2% agarose gel and perform a long electrophoretic run. The amplicon should be single and of the expected size;

3. Purification of the amplicon in Sepharose G100 and quantify the amplicon using a Nanodrop spectrophotometer.

4. Make serial dilutions (factor 10) starting from 1ng/ul and until 1ag/ul. Use the dilutions in the range 10 pg-1 ag.

5. Prepare the PCR real-time mix to get the standard curve for each gene according

the following recipe:

Final reaction volume/sample: 15 ul

7.5 ul Master Mix 2X

0.3 ul 10 uM primer up (final conc.: 50 nM)

0.3 ul 10 uM primer dw (final conc.: 50 nM)

1.5 ul template

sterilized 2d H2O: 5.4 uL

Add two negative controls for each standard curve.

Use the following real-time PCR cycle:

50 cycles of: 15" at 95°C (denaturation) + 1' at 60°C

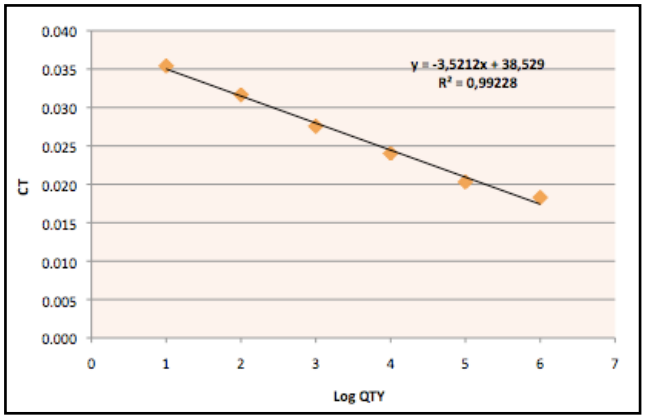
Real-Time PCR was performed using an ABI PRISM 7000 instrument and a SYBR Green Master Mix kit (Applied Biosystems, catalog n° 4364346).

6. Analysis of the Real-Time PCR to obtain the standard curve for each couple of primers: a. Threshold cycles in the replicas must be very close (not more than 0.2 cycles of distance); b. Each sample must show just one dissociation peak; c. All the samples of a standard curve must show a dissociation peak at the same temperature; d. The Tm of the dissociation peak must be over 74°C (lower temperatures reflect "primer dimer" formation);

7. Determination of the slope of the standard curve: export the data in an excel file and create a table as following:

Sample Name	Detector	CT	Std. Dev. CT	Quantity	Log qty	Average Ct
(Zds)100fg	SYBR	18.13	0.253	100000	5	18.31
(Zds)100fg	SYBR	18.49	0.253	100000		
(Zds)10fg	SYBR	21.84	0.139	10000	4	21.935
(Zds)10fg	SYBR	22.03	0.139	10000		
(Zds)1fg	SYBR	25.8	0.165	1000	3	25.68
(Zds)1fg	SYBR	25.56	0.165	1000		
(Zds)100ag	SYBR	29.55	-1	100	2	29.55

Create a chart with the slope (o trendline), the R^2 and the intercept. Example: in this chart the slope is y=-3.5, the intercept is 38.52 and R^2 is 0.992.



Slope values should be between -3.6 and -2.8 (the optimum is –3.2) and R^2 values should be close to 1 and never under 0.98.

3.4.2 Sample analysis

For quantitation of the unknown samples, use the slope, R^2 and intercept values of the related primer standard curve. Example: (ZDS A1 and ZDS A2 are replicas of the same cDNA).

Sample	Ct	St.Dev. Ct	INTER.	Y	A	Qty	Average Qty	StDev Qty
Zds A1	26	0.17	35.86	-3.46	2.84	698.91	645.34	75.76
Zds A2	26,3	0.17	35.86	-3.46	2.77	591.78		
Zds B1	34,3	0.731	35.86	-3.46	0.45	2.84	4.26	2.01
Zds B2	33,3	0.731	35.86	-3.46	0.75	5.68		

A= (Ct-INTER.)/Y

Qty= POWER (10;A)

Average Qty= Avg. of the Qty in replicates samples.

9. Normalization of the data: a. Avg. normalization: Divide the Qty value of a gene for the Qty value of the gene used for the normalization. St.Dev. normalization: divide the St.dev. value of a gene for the Qty value of the normalization gene. All data were normalized for the level of the α-*Actin* or β-*Tubulin.* Primers for Real Time experiments (Tab. 5) were designed using the Primer Express v2.0 software and validated with Amplify v3.1. Moreover primers were checked another time and were used only those blasted in conserved CDS regions with respect to sequences of the wild tomato species.

3.5 Sequencing pipeline

Purified PCR products were utilized as substrates for sequencing reactions.

Reactions were prepared in a total volume of 10 ul/sample:

1X BigDye Terminator sequencing buffer (AB, catalog n° 4336699),

2 pmoles of forward OR reverse primer,

0.33 uL of BigDye Terminator Ready Reaction mix (AB, catalog n° 4336611).

Thermocycler conditions:

30 cycles: 94 °C x 30 sec, 48 °C x 30 sec, 60 °C x 4 min.

Sequencing reaction products were purified on Sephadex G50 (GE, catalog n° 17004202). Samples were eluted in 5 ul of ABI Prism Hi-Di Formamide (AB, catalog n° 43311320) and loaded onto an ABI 3730 DNA Analyser.

3.6 Sequence analysis and bioinformatic tools

All the sequences were analysed with “CodonCode Aligner” (CodonCode Corp, USA) in order to produce contigs, define coding region and produce protein sequence for each gene in this thesis. “MegAlign” (DNAStar, USA) was used to produce alignment of the proteins with the “ClustalW” method (http://www.ebi.ac.uk/Tools/clustalw2/index.html).

Moreover, the sequences obtained were analysed with a series of bioinformatic tools for comparing and evaluation of various aspects and peculiarities of the sequences.

- Blast tools (http://blast.ncbi.nlm.nih.gov/Blast.cgi) were used to look for regions of similarity between sequences, both at nucleotide of amino acid levels.

Blastn: search a nucleotide database using a nucleotide query

Blastp: search protein database using a protein query

tBlastn: **s**earch translated nucleotide databases using a protein query

tBlastx: search translated nucleotide database using a translated nucleotide query

- TFSEARCH (http://www.cbrc.jp/research/db/TFSEARCH.html) and TESS (http://www.cbil.upenn.edu/cgi-bin/tess/tess) were used for predicting transcription factor binding sites in DNA sequences

- ChloroP (http://www.cbs.dtu.dk/services/ChloroP/) was used to predict the presence of chloroplast transit peptides (cTP) in protein sequences and the location of potential cTP cleavage sites.

- SIFT (Sorting intolerant from tolerant - http://sift.jcvi.org/) was used to predict if an amino acid substitution may affect protein function. The prediction is based on a sequence homology search among proteins that may share similar function to the

query (looking for the probability of each possible substitution from the alignment) and the physical properties of amino acids. SIFT values vary between 0 and 1. Values lower than 0.05 correspond to mutations that, theoretically, affect protein function. Because SIFT usually under-estimates the number of deleterious mutations, we empirically identified values between 0.05 and 0.1 as possible deleterious mutations. For each SIFT value, the software return a quality value that indicate the goodness of the prediction. This value depends on the quality of the alignment with the sequences present in the database. In order to get a good estimation on a wide database, the SWISS-PROT/TrEMBL database was chosen.

- Mega4 (http://www.megasoftware.net/) was used to establish the dN/dS ratio on a coding gene. This represents the ratio of nonsynonymous substitution frequency (dN) to the synonymous substitutions frequency (dS), that can be used as an indicator of selective pressure acting on a protein-coding gene [178]. Comparisons of homologous genes with a high dN/dS ratio are usually said to be evolving under positive selection.

3.7 Phylogenetic analysis

Phylogenetic relationships, obtained with Mega4, were performed using the "Neighbor Joining" method (NJ). The NJ method is a simplified version of the "minimum evolution" method, which uses distance measures to correct for multiple hits at the same sites, and chooses a topology showing the smallest value of the sum of all branches as an estimate of the correct tree. With the NJ method, the examination of different topologies is imbedded in the algorithm, so that only

one final tree is produced [179]. This method does not require the assumption of a constant rate of evolution so it produces an unrooted tree. To tests of the reliability of the tree we used the bootstrap test with 500 replicas and a random number of seeds. Sequence gaps are treated with the "pairwise deletion" option and the substitution model was the "p-distance".

3.8 Experimental repeatability

All experiments involved a minimum of 3 independent biological and 2 technical replicates.

3.9 Ethylene analysis in tomato fruits

1. Collect carefully tomato fruits at 44, 51, 58 and 65 days post-anthesis (DPA) severing the petiole with a scissor. Select 5 samples (with uniform size and pigmentation) for each line;

2. Put carefully each fruit in a Mason jar or in a chromatography vial (SIGMA catalog n° SU860099): the volume depends on the size of the fruits and varies from 11 ml, to 220 ml (Fig. 20);

3. Measurements: leave open overnight at room temperature, paying attention not to move the vial/jar, close tightly with the cover and after 2 hours (red fruits) or 4 hours (green fruits) insert a SGE Gas Tight Syringe, Fixed Needle 25 ul (SIGMA catalog n° 26257) in each jar and take out 25 ul of headspace after pipetting up and down 2-3 times;

4. Collect 4 headspace samples for each fruit and read the samples trough at gas-chromatography (Varian 3600 with FID, Transport gas: H2);

5. Calibrate using a standard with a known concentration of ethylene (eg 10 ppm).

3.10 Press-test assay

For tomato fruit firmness measurements, a durometer was used. At least 6 different fruits/genotype and stage were selected for the analyses.

3.11 Pigment analysis

3.11.1 Spectrophotometric analyses of total carotenoid and chlorophyll content in tomato berry

1. Weight approximately 250 mg of pericarp tissue and add 250 ul 2d water;

2. Homogenize with an Ultraturrax homogenizer (IKA, Staufen, Germany) at maximum speed until the sample is completely disrupted;

3. Add of 500 ul of chloroform and vortex extensively and after centrifuge at 15000xg for 3';

4. Dilute 50 ul of the lower phase in 1 ml of 100% acetone. Read at 400-500 nm spectrum against a 95% acetone blank;

5. Express fruit carotenoids as (OD474-OD720)/g of fresh tissue weight (FW) for red-fruited species.

6. For green-fruited species (rich in chlorophylls), the pigment contents were calculated according the following formulas [180]:

Chl a: (12.25 x $A^{663.6\ nm}$-2.55 x $A^{646.6\ nm}$)

Chl b: (20.31 x $A^{646.8\ nm}$- 4.91 x $A^{663.6\ nm}$)

Chl a + Chl b: (17.76 x $A^{646.8\ nm}$ + 7.43 x $A^{663.6\ nm}$)

Total carotenoids= (1000 A_{470} - 1.82 *Chl A* - 85.02 *Chl B*)/198

The results are in ug of pigments/ml extract then divide by g fresh tissue weight (FW) to have ug of pigments/g (FW).

3.11.2 Carotenoid and chlorophyll extraction from tomato berry

Perform all the protocol on ice and in dark conditions:

1. Weigh lyophilized and powdered pericarp (20 mg for red fruits and 50 mg for green/orange fruits) and collects them in a 15 ml polypropylene vial (snap-cap);

2. Add 1 ml of 100% Methanol (HPLC grade) + 2 ml of Chloroform (HPLC grade) and vortex;

3. Add 100 ul of DL-α-tocopherol acetate (1mg/ml) per sample as internal standard and vortex;

4. Prepare samples with 100 ul DL- α-tocopherol acetate (1mg/ml) (SIGMA catalog n° T3366) as standard (the external standard of the HPLC run) in separate eppendorfs: 1 standard each 3 unknown samples.

5. Incubate on ice for 20';

6. Add 1 ml of Tris-HCl 50mM and vortex;

7. Centrifuge 10’ at 3000g (4°C);

8. Recover the lower phase without contaminating the extract with interphase;

9. Extract again the original sample with 1 ml of Chloroform, vortex and centrifuge 10’ at 3000g (4°C);

10. Recover the new lower phase without contaminate the extract with vegetal residues and blend with the previously extract;

11. Evaporate the sample in speedvac concentrator and resuspend the pellet with 100 ul of Chloroform;

12. Transfer the sample in an Eppendorf and evaporate again. Evaporate also the standard previously prepared.

13. Resuspend in 60 ul final volume (Methanol-Chloroform both HPLC grade 1:1 [v/v]);

14. Filter the sample (Syringe Tip Filter, SIGMA catalog n° 54144-U) and fill the chromatography vials (Vials screw top 2 ml amber glass SIGMA catalog n° 29119-U; Inserts for 2 mL large opening vials, SIGMA catalog n° SU860066).

3.11.3 Carotenoid extraction from bacteria

The extraction of carotenoids from *E. coli* was performed as previously described [79] with some modifications.

Aliquots of 5 ml were taken from bacterial suspension cultures after the induction. Cells were harvested by centrifugation at 4000 g for 5 min, washed once with 1.5 ml water and 5 mM of Tris-HCl and finally resuspended in 1.5 ml of acetone (HPLC grade). 100 ul of DL- α-tocopherol acetate (1mg/ml) per sample were added as

internal standard, and the cells were incubated at 65°C for 10 min in the dark.

The samples were then centrifuged again at 13000 g for 5 min and the acetone supernatant containing the pigments was placed in a clean tube. For complete carotenoid extraction, the cells were again resuspended a second time in 1.5 ml of acetone and centrifuged as above. The pigment extract was vacuum-dried and resuspended in 100 µl of Chloroform (HPLC grade). Samples were centrifuged at 13000 g for 3 minutes and supernatant containing the pigments without particles was placed in a clean tube. The extract was again vacuum-dried and finally resuspended in 60 µl of Chloroform and Acetonitrile both HPLC grade 1:1 [v/v].

3.11.4 HPLC analysis

Berry carotenoid analysis was performed using an Accela™ (Thermo Fisher Scientific) system and a C30 reversed-phase column (YMC Europe GmbH, Schermbeck, Germany). The solvent systems were A: MeOH/*tert*-butylmethyl ether (1:1 [v/v]) and B: MeOH/*tert*-butyl-methyl ether/water (60:12:12 [v/v]). The gradient system is described in Tab. 6. For the bacterial carotenoids, HPLC analysis was led with a C18 reversed-phase column (SUPELCO Analytical). An isocratic solvent system (100% Acetonitrile) was used for 45 minutes. Chemicals and solvents were obtained from Sigma-Aldrich (U.S.). Unless stated otherwise HPLC grade chemicals were used. Carotenoids were identified by their absorption spectra, monitored using a photodiode array detector (Accela PDA detector) and by comigration with authentic standards (lycopene, α-carotene, δ-carotene, β-carotene, zeaxanthin, lutein, violaxanthin, neoxanthin).

Carotenoid peak areas were integrated at their individual λmax and DL-α-tocopherol acetate were integrated at 285 nm. For normalization and quantification, all peaks were normalized relative to the internal standard (DL-α-tocopherol acetate) to correct for extraction and injection variability. An external calibration curve of DL-α-tocopherol acetate, run separately, was used to calculate absolute amounts. All carotenoid peaks underwent a second normalization to correct for their individual molar extinction coefficients [181]. Total carotenoid and chlorophyll levels, extracted from berry of different species were normalized using total spectrophotometric measurements.

4. Results and Discussion

4.1 The tomato genome: map position and structure of genes involved in the carotenoid biosynthetic pathway.

Thanks to the availability of the genome sequences and annotation data of tomato, potato and *A. thaliana*, we performed bioinformatic analyses to analyze various the genes belonging to the carotenoid biosynthetic pathway in these genomes.

4.1.1 Scaffold identification

The analysed genes were Phytoene synthase 1 and 2 (*PSY1* and *PSY2*), Phytoene desaturase (*PDS*), Carotene isomerase (*CrtISO*), ζ-carotene desaturase (*ZDS*), Lycopene ε-cyclase (*LCY-e*), 2 Lycopene β-cyclases (*LCY-b* and *CYC-b*, chloroplast and chromoplast-specific respectively), β-carotene hydroxylase 1 and 2 (*CHY1* and *CHY2*) (Fig. 17).

The first step has been to align the amino acid sequence of the proteins with the tomato genome thanks to the tBlastn tool (search translated nucleotide databases using a protein query) (http://solgenomics.net/tools/blast/) in order to identify the scaffolds containing the proteins (Tab. 7). For each protein the first significant aligned sequence was, as expected, the same protein, but in some cases the alignment revealed additional positive sequences. Parts of these were paralogous

proteins; in fact some genes of the pathway are duplicated (*PSY 1* and *2*, *CHY 1* and *2, LCY-b* and *CYC-b*), while other sequences were due to the presence of conserved domains in different enzymes (such as *PDS* and *ZDS*).

Additionally, this analysis has highlighted the presence of three new genes, annotated respectively as putative *PSY*, putative *CrtISO* and putative *LCY-b*; the putative *LCY-b* maps in the same scaffold as the *bona fide CrtISO*.

4.1.2 Fine mapping of genes on the genome

The position of carotenoid genes on the tomato genome was already described in literature [176]; in order to verify these data, the genes have been mapped using the *S. pennellii* ILs (Tab. 8). The results have confirmed the known data with the exception of the chromosome position of *PDS*: in fact this gene maps on BIN I instead of BIN G of Chr. 3.

Thanks to the availability of the whole tomato genome and the accessibility to a set of mapped genetic markers, developed by a Japanese partner of the tomato sequencing project (Shusei Sato, personal communication), it was been possible to map more accurately, respect to IL mapping, the genes on the genetic map. The only gene not finely mapped by this approach was *LCY-b*.

4.1.3 Microsynteny among *A. thaliana, S. lycopersicum* and *S. tuberosum*: the phytoene synthase genes.

With the term "synteny" is described the preservation of the precise order of genes

on a chromosome passed down from a common ancestor. The analysis of synteny has several applications in genomics: a stronger-than-expected synteny can reflect selection for functional relationships between syntenic genes, such as combinations of alleles that are advantageous when inherited together; synteny is one of the most consistent criteria for establishing orthology between genomic regions in different species. Macrosynteny is the preservation of synteny in large portions of a chromosome whereas microsynteny is the preservation of synteny for only a few genes at a time.

A comparison was undertaken to investigate the presence/absence of microsynteny around the *PSY* genes. Annotated proteins, encoded in a range of 100 Kb around these genes have been compared in the genomes of *A. thaliana*, tomato and potato (Fig. 21). *A. thaliana* has a single Phytoene synthase (*PSY*) gene, instead tomato and potato have 3 Phytoene synthase (*PSY1* and *PSY2* chromoplast- and chloroplast-specific, respectively, plus a novel putative *PSY*).

The dendrogram in Fig. 22, built using the protein sequences, shows the evolutionary distance among these enzymes: as an out-group we used the bacterial Phytoene synthase of *Pantoea agglomerans* (CrtB). The phylogenetic tree shows that PSY of *A. thaliana* is the ancestor with respect to tomato and potato and the *PSY1* and *PSY2* are paralogous genes.

Genes flanking *PSY* in *A. thaliana* and *PSY2* in tomato have been compared on the bases of their annotation and deduced amino acid sequences (Tab. 9): 6 of them have the same functions and their relative position is almost conserved. Two genes show differences in number of copies between scaffolds: in tomato the Glutathione

S-transferase genes is present 2 times and Kinase R-like genes is present 4 times. The same analysis has been led on the regions flanking *PSY* of *A. thaliana* and *PSY1* of tomato. Only one gene relates *PSY* of *A. thaliana* with *PSY1* of tomato: the zinc finger-related gene. The comparison of the genes flanking *PSY1* and *PSY2* in tomato (Tab. 10) has revealed instead high levels of similarity: in fact 6 genes are shared and their relative position is almost conserved. Only the 2-succinylbenzoate--CoA ligase genes show differences between the scaffolds both in number than in relative position: in the *PSY2* scaffold are present 3 times and in *PSY1* only 2 times.

Genes flanking *PSY* in *A. thaliana* and the putative *PSY* in tomato have been compared (Tab. 11): only 3 genes encode proteins with the same functions and their relative position and orientation is conserved with respect to *PSY.*

The analysis of the genes flanking *PSY2* and the putative *PSY* reveals the conservation of 4 genes (Tab. 12) instead the comparison of *PSY1* and the putative *PSY* shows that only one gene is conserved: the zinc finger-related gene.

In conclusion the analysis of the genes flanking *PSY* of Arabidopsis, *PSY1*, *PSY2* and the putative *PSY* in tomato has shown that:

- the gene codifying for a deduced amino acid sequences annotate as zinc finger protein, is present in all the scaffolds in the same relative position;
- the genomic regions flanking *PSY* of Arabidopsis show the highest level of microsynteny with the flanking regions of tomato *PSY2* (6 genes), followed by the putative *PSY* (4 genes);
- the flanking regions of *PSY1* of tomato seems to have lost the original microsynteny with Arabidopsis except for one gene.

A further comparison, at the nucleotide level, has been conducted on the genes flanking the *PSY1* and *PSY2* genes of potato and tomato: both genomic regions show a very high level of microsynteny. The Fig. 23 shows clearly the high level of nucleotide conservation of these couples of genes (panels a-d). Since the genomes of potato and tomato are conserved, it has been possible to compare the nucleotide sequences. The scaffolds containing the *PSY1* and *PSY2* genes in tomato and potato were aligned (Fig. 23 panels e and f); it is evident that the scaffolds containing *PSY1* show a large interruption of the synteny (panel e).

4.1.4 Analyses of putative genes

The Blast analysis conducted on the tomato genome has highlighted the presence of three new sequences, annotated respectively as putative *CrtISO*, putative *LCY-b* and putative *PSY*.

4.1.4.a The putative *CrtISO*

This encodes a polypeptide 568 AA long, while CrtISO is 615 AA long. The percentage of identity of the two proteins is only 37% (Tab. 7) (Fig. 24, panel a).

The analysis of conserved domains didn't highlight large differences between the two proteins (Fig. 24, panel b and c).

From the phylogenetic tree, using the bacterial Carotene desaturase (CrtI) of *P. agglomerans* as an outgroup and adding also a putative CrtISO found in potato, it possible to observe that:

- the two putative proteins (tomato and potato) are very closely related;

- the putative CrtISO proteins are very different with respect to the known CrtISO of *A. thaliana*, tomato and potato and more similar to bacterial CrtI.

Preliminary analyses, that are in progress in our lab, on RNAseq data obtained with 454 sequencing showed that this gene is actually expressed in all the analyzed tissue with the exception of root and flower (Tab. 13).

4.1.4.b The putative *LCY-b*

This putative encoded enzyme is 500 AA long, like LCY-b, and its percentage of identity with respect to LCY-b is very high (87% on the whole protein) (Tab. 7) (Fig. 25, panel a). The putative *LCY-b* maps on the same scaffold as *CrtISO*.

The conserved domain analysis didn't show differences between these proteins (Fig. 25, panel b and c). A phylogenetic tree was built, with the protein sequences of all tomato cyclases (putative LCY-b, LCY-b, CYC-b and LCY-e), LCY-b and pututative LCY-b of potato, LCY-b of *A. thaliana* and the bacterial β-cyclase of *P. agglomerans* (CrtY). The results show that the putative *LCY-b* arose from a recent duplication of *LCY-b*, predating the tomato-potato divergence (approx. 10 My before present).

Preliminary transcriptome analyses indicate that this gene is actually expressed in all the tissues with the exception of berry mature green and breaker (Tab. 13).

4.1.4.c The putative *PSY*

The amino acid sequence of the putative PSY is composed of 215 residues (about half with respect to PSY1 and PSY2) and it shows 70% identity with respect to

PSY1 and PSY2 (Tab. 7). The protein sequence alignment shows that, this polypeptide lacks a big part of N-terminal sequence (about 220 amino acids) (Fig. 26, panel a). Despite these observations, RNAseq data obtained with 454 show that putative *PSY* is expressed, even if at low levels, in roots (Tab. 13).

4.2 Gene sequencing

The genes have been sequenced from *cv* Heinz, using sequences deposited in Genbank as a reference, in order to complete the missing parts (such as introns). 5' UTR and promoter regions of all genes have been extended when available, or *de novo* sequenced, searching BACs containing the genes of interest in the tomato sequencing project database (http://solgenomics.net/). Primers used for gene sequencing in Heinz have been employed for sequencing wild species. Often, *ad hoc* primers have been designed to sequence non-conserved regions such as promoters and introns due to the presence of many mutations in the wild genotypes. The genes have been finally completely amplified and sequenced (in all about 50 Kilobases) in *cv* Heinz, in 3 wild ecotypes of *S. lycopersicum* var. *cerasiforme* and in 9 wild species (Tab. 3). The sequences have been checked and assembled, generating a continuous contig for each gene. Next, the contigs have been aligned highlighting the presence of insertions, deletions or point mutations: a summary of these mutations is reported in Tab. 14. As expected, non-conserved genomic regions (promoters, introns, 5' and 3' UTR) present a high mutation rate, while non-synonymous mutations (polymorphisms that cause a change in the amino acid sequence) are less frequent. Promoter regions have been analysed with several bioinformatic tools (see Materials and Methods) however, no evident mutations were found on predicted transcription factor binding sites. All genes nevertheless, have a very large number of mutations in the promoter (including

insertions and deletions of up to 100 bp) and many of those differentiate the red or orange-fruited species from the green-fruited ones. The most interesting mutations found in *PSY1*, and *LCY-e* are reported in Fig. 27, those found in *LCY-b* and *CYC-b* are discussed in the next paragraph (Fig. 29).

4.2.1 The quest for *Beta* mutation

Beta (*B*) is a single dominant allele of *CYC-b* that affects fruit pigmentation in tomato, increasing β-carotene in the fruit. This allele, originated from wild tomato species, was introduced into the cultivated tomato by crossings. The phenotype of *Beta* is due to increased expression of the *CYC-b* (the chromoplast-specific lycopene β–cyclase) during fruit ripening. Ronen at al. [79] hypothesized that the increased transcription of *B* was due to the presence, in the *B* allele, of a short sequence (8 bp long) probably related to ethylene responsiveness, whereas in the *b* allele this sequence is mutated (Fig. 28, panel a). Our sequences do not support this hypothesis: all the species in fact carry the same short sequence without mutations (Fig. 28, panel b). The simplest explanation is probably to a sequencing error on the part of Ronen and coll. Thus, the molecular elucidation of *B* and *b* alleles is still pending. Analyses on *S. cheesmaniae* RILs (data not shown; Fraser P. personal communication) suggest that high β-carotene QTLs map to regions on Chr. 6 and 4. Our hypothesis is that *LCY-b* (Chr. 4) and *CYC-b* (Chr. 6) are the candidate genes for such QTLs. From our sequence data, the likely mutations distinguishing red-fruited species from other species found in β–cyclase promoters are: "-352" in the *LCY-b* (Fig. 29, panel a) and "-784" and "-671" in the *Cyc-b* (Fig.

29, panel b).

4.2.2 Amino acid mutation analysis: comparisons among genotypes

The deduced amino acid sequences of each gene have been aligned using an outgroup the amino acid sequences of *A. thaliana* and *Synechococcus spp.* (unicellular cyanobacterium used as model microorganism).

Amino acid substitutions are reported for each protein from Fig. 30 to Fig. 39 and these have been weighted on the basis of:

- their position in the protein sequence: for example the chloroplast transit peptide *(cTP)* is known to be more tolerant to mutations or deletions [182] [183];

- their position in the alignment: mutations in regions highly conserved with respect to *A. thaliana* and *Synechococcus* could alter functional domains (hyper or hypomorphic alleles);

- the chemical and physical properties of the mutated amino acids; this analysis is performed using SIFT, a bioinformatics tool that predicts whether a mutation affects protein function. SIFT bases the prediction also on sequence homology using the UniProt database (Embl-Ebi) (see Materials and Methods)

From these analyses the most interesting mutations seem to be:

- in PSY1 (Fig. 30, panel a) substitution n° 3 in *S. cheesmaniae* from arginine (R, basic) to a tryptophan (W, hydrophobic) and it is predicted by SIFT to affect protein function. Moreover, this mutation is on the predicted cTP cleavage site (according to ChloroP prediction, see Materials and Methods). In *S. cheesmaniae*,

the predicted mature protein is 5 amino acids longer (Fig. 30, panel a and panel b). Thus in *S. cheesmaniae* has lower carotenoid levels than *S. lycopersicum*, this is predicted to be a hypomorphic allele.

- in LCY-e (Fig. 35) substitution n°5 from Histidine (H, polar) to Asparagine (N, polar) is the only mutation in the amino acid sequence that differentiates the green-fruited species from red and orange-fruited ones. Since green-fruited species have a higher level of ε-ionone rings, this is predicted to be a hypermorphic allele. This prediction has been verified through expression in *E. coli* (see below).

- in LCY-b (Fig. 36), substitution n°12 in *S. pimpinellifolium* from Aspartic Acid (D, hydrophilic) to Tyrosine (Y, hydrophobic). This mutation is in a conserved region with respect to *A. thaliana* and is predicted by SIFT to affect protein function. Since *S. pimpinellifolium* has different level of β-ionone rings with respect to *S. lycopersicum*, this is predicted to be a hypermorphic allele.

4.2.3 Phylogenetic tree building

Phylogenetic trees have been built using both nucleotide (data not shown) and amino acid sequences. All the dendrograms, according to the literature [50], [51], shown a maintained topology in which the coloured-berry species are more closely related with the modern tomato and the green-berry species in a more ancestral position. These data suggest that events of lateral gene transfer did not happen. In Fig. 40 a dendrogram, built on the basis of the deduced amino acid sequence sum of the analysed genes, is shown.

4.2.4 Diversifying and purifying selection on carotenoid biosynthetic pathway enzymes

Nonsynonymous mutations are missense mutations that cause an amino acid change, whereas synonymous mutations do not (silent mutations). In order to evaluate the selective pressure on these gene products, the frequency of nonsynonymous (dN) and synonymous mutations (dS) have been calculated. In fact the comparison between dN and dS can suggest whether the natural selection is acting, at the molecular level, to promote the fixation of advantageous mutations (diversifying selection) or to remove deleterious mutations (purifying selection). A low value of dN can indicate two different situations: the gene function is conserved or the gene region has itself a low rate of mutation. To eliminate this second possibility, the ratio dN on dS is analysed.

In general, when diversifying selection dominates, the dN/dS ratio is greater than 1; in this case, diversity at the amino acid level is favoured, likely due to the fitness advantage provided by the mutations. Conversely, when purifying selection dominates, the dN/dS ratio is less than 1; in this case, most amino acid changes are deleterious and, therefore, are selected against. When the positive and negative selection forces balance each other, the dN/dS ratio is close to 1.

The analysis results are reported in Tab. 15. Other than the carotenoid genes, we have sequenced a molecular clock gene, the gene for the large subunit of the ribulose-bisphosphate carboxylase (*rbcL*) [184].

All proteins showed a low dN/dS ratio, indicating the purifying selection dominates; it is possible however to observe some differences between proteins.

The products of the unique genes are in general more highly conserved. This is not surprising, since the *A. thaliana* mutants for *PDS* and *ZDS* which are unique genes have lethal phenotype [185], [186], while mutants for *LCY-e* are more sensitive to photooxidative stresses [187], *CrtISO*, among the unique genes, showed the highest dN/dS ratio value: in fact mutants for this gene do not show a lethal phenotype, even if their growth is not optimal [141]. Among the duplicate genes, *PSY1* show a low dN/dS value. The conservation of *PSY1*, the chromoplast-specific paralog of an enzyme pair, suggests either a yet undescribed role for this enzyme, or an important role of fruit pigmentation for reproductive fitness.

4.3 Functional analyses

Functional analyses (ethylene production, gene expression, carotenoid and chlorophyll composition experiments) were conducted on *S. lycopersicum* (*cv* Heinz), *S. pimpinellifolium*, *S. cheesmaniae*, *S. arcanum* and S*. neorickii*, representing the main phenotypes of interest: respectively 2 species have red, 1 orange and 2 green berries. In addition to the species listed above, *var*. *cerasiforme* (the wild ecotypes of *S. lycopersicum*) was included for ethylene measurements. For each line, three plants were grown in the greenhouse. At specific time points, berries were collected and stored at -80° (see Materials and Methods). Some species (*S. habrochaites*, *S. neorickii* and *S. chilense*) did not fruit; others did not produce a quantity of berries sufficient for all experiments.

4.3.1 Ethylene production and fruit softening

Ethylene is the main hormone controlling tomato fruit ripening. This is a very complex process that involves numerous genes and transcription factors, some of them without a completely clarified role [17]. Ripening leads to a large-scale alteration of several fruit traits, among which carotenoid accumulation and loss of firmness. Berries of wild tomatoes have different biochemical composition [188]. The ripening profile and the firmness of these species were investigated. The experiments were performed staging berries system on the basis of days post-anthesis (DPA). The selected time points (44, 51 58 and 65 DPA) correspond, in cultivated tomato, to the following developmental stages: mature green, breaker,

red ripe and over ripe (see Materials and Methods). The data showed that all the wild species were able to produce ethylene (Fig. 41, panel a), in agreement with literature data [189]. The fruits that change colour, i.e. *S. lycopersicum var. cerasiforme*, *S. pimpinellifolium* and *S. cheesmaniae* exhibited climacteric ethylene production, similar to the cultivated tomato. There were differences, however, in the magnitude of the climacteric response. Regarding the green-fruited species, *S. neorickii* exhibited a small climacteric peak, while *S. arcanum* showed an ethylene peak shortly before berry abscission that occurred around 70 DPA. Fruits from all species showed loss of firmness (Fig. 41, panel b) with *S. cheesmaniae* and *S. arcanum* remaining the most firm, even at 65 DPA [17]. Berries of *S. pimpinellifolium* and *S. neorickii* at 65 DPA were too soft to be assayed.

These data suggest that, between tomato species, differences exist in climacteric ethylene production. In the green-fruited species, like *S. neorickii*, the lack of fruit colour change cannot be simply attributed to a lack of ethylene production.

4.3.2 Fruit carotenoid composition

The fruit carotenoid composition from *S. lycopersicum* (*cv* Heinz), and its wild relatives, *S. pimpinellifolium*, *S. cheesmaniae*, *S. arcanum* and *S. neorickii,* were analyzed through diode array HPLC (LC-PDA, Fig. 42 and Fig. 43). Total carotenoid content analysis was performed in 58 and 65 DPA berries, showing significant differences among the genotypes (Fig. 44, Fig. 45 and Tab. 16). Besides chlorophylls, it is noteworthy that the major difference between the red fruited and the other species (*S. cheesmaniae* and the two green-fruited species) is in the fruit xanthophyll contents. Two compounds are inferred to be xanthophylls (not corresponding to those of the all-*trans*-xanthophyll standards and, thus, classified as "Unknown xanthophyll") on the basis of their spectra, wavelength range and retention time (Fig. 42 and Fig. 43). The precise identification of these compounds can not be confirmed by through the technological platform (LC-PDA) but further analysis based on liquid chromatography-mass spectrometry (LC-APCI-MS) will be performed in order to clarify the identity of these carotenoids. We hypothesize that these peaks are attributable to 9-*cis*-xanthophylls, found in berries of wild *Solanum* species. [191].

The red-fruited species, *S. lycopersicum* (Heinz) and *S. pimpinellifolium*, generally display similar profiles with some notable exceptions: *S. pimpinellifolium,* in fact, lacks ζ-carotene and β, β-xanthophylls, and contains approx. 5-fold lower levels of lutein with respect to Heinz at 58 DPA. The highest content of total carotenoids has been found in these two red-fruited species at 65 DPA. On the other side *S. cheesmaniae*, the orange-fruited species, shows the lowest content of total

carotenoids (about 5-fold less then Heinz at 58 DPA) and shares a similar pigment profile, although at lower levels, with the green-fruited species, consisting essentially in xanthophylls and chlorophylls (Fig. 46 and Fig. 47). Moreover, these species lack metabolites produced in early steps of the carotenoid pathway, such as phytoene and phytofluene (Fig. 46). This observation is more evident grouping carotenoid metabolites in three classes: "early" (from Geranylgeranyl pyrophosphate to lycopene), "β-branch" (from β-carotene to neoxanthin) and "α-branch" (from δ-carotene to lutein) reactions (Fig. 17).

Analyses of wild-specie HPLC chromatograms, compared to Heinz data, can be summarized as follows: (Fig. 44 and Fig. 45):

- All the samples have similar amount of β-carotene with the exception of *S. cheesmaniae* at 65 DPA, which shows the minimum amount (Tab. 16).
- A dramatic accumulation of β, β-xanthophylls (neoxanthin and violaxanthin) is evident in all orange and green-fruited species at 58 and 65 DPA, with a maximum level of 20-fold in *S. neorickii*.
- An increased accumulation of lutein (α-branch) has been also observed, especially in *S. neorickii*, although at lower (about 3-fold) than β-branch xanthophylls levels.
- *S. neorickii,* a green-fruited species, unexpectedly shows lycopene accumulation, although at 60-fold lower content with respect to Heinz.

According to these data, it appears that green-fruited species mainly accumulate carotenoids as xanthophylls and β-carotene, like leaves, and the orange-fruited *S. cheesmaniae* specie actually displays a is more similar to green relatives

carotenoid profile.

S. cheesmaniae is an endemic species from Galapagos Islands. These islands are still now a living laboratory of evolution. Their ecosystem is severe, and they are populated by several endemic species that have independently evolved from human domestication. Endemism, in fact, frequently occurs in islands because of their geographical barriers, which block genetic dispersion. In Galapagos Islands both endemic animal species (the well known birds, iguanas and penguins) and above all endemic plants are present, with around 180 species in the latter, among which endemic species of coffee, cotton and chili pepper (http://www.darwinfoundation.org) have been reported. It is therefore possible hypothesizing that some physiological characters in *S. cheesmaniae,* such as berry metabolite composition, have been subjected to independent evolution phenomenon with respect to the other wild tomatoes which originated from mainland South America.

4.3.3 Expression of carotenoid genes

The expression of carotenoid gene transcripts in fruits has been analyzed via Real-Time PCR. The analyses have been conducted on berry at two different ripening stages (58 and 65 DPA). Obtained data have been normalized for expression of the housekeeping *β-tubulin* (data not shown) and *Actin* genes and then for the expression level in the modern tomato (*S. lycopersicum, cv.* Heinz). The analysis of the expression profiles in berries of wild species compared to Heinz (Fig. 48 and Fig. 49) indicated that:

- The chromoplast-specific genes *PSY1* and *CHY2* were significantly expressed only in red-fruited species (*S. lycopersicum* and *S. pimpinellifolium*). The other species (orange and green-fruited) had a practically undetectable (*PSY1*) or very low (C*HY2*) expression (Fig. 50) at both ripening stages.

- At 58 DPA, the chloroplast-specific *PSY2* showed a higher expression level in green and orange-fruited tomato (up to 4-fold). Since *PSY1* in green and orange berries is not expressed, it is likely that *PSY2* is the sole phytoene synthase responsible for the biosynthetic carotenoid pathway in non red-fruited species.

- The expression level of *PDS, CrtISO, ZDS* and *CHY1* was similar among all species at both ripening stages.

- *LCY-e* shows much higher expression in non red-fruited species, especially at 65 DPA. In particular, *S. neorickii* showed the highest expression (about 3000-fold). HPLC data confirmed the higher levels of α-branch carotenoids, essentially lutein, in these species (Fig. 45).

- Both *LCY-b* and *CYC-b* show a much higher expression level in non red-fruited

species. Thus, our data suggest that, contrary to what suggested by Ronen et al [79], the emergence of red-fruited species involved hypomorphic mutations in both *CYC-b* and *LCY-b,* decreasing their expression levels in the fruit.

4.3.4 Expression in *E. coli* of carotenoid gene

The analyses on deduced amino acid sequences have revealed the presence of some mutations that could be hypo or hypermorphic alleles of the enzyme. In order to study the effect on enzyme activity of these substitutions, a project to express in *E. coli* the mature proteins has been undertaken. In particular, in LCY-e, substitution n° 5 from Histidine to Asparagine was the only mutation, in the deduced amino acid sequence that differentiated the green fruit species from all the others. Since green-fruited species have a higher level of α-branch carotenoids, this has been predicted to be a hypermorphic allele (Fig. 35).

cDNA was prepared from total RNA and amplified with primers tailed with recognition sites for restriction enzymes (*BamHI* and *KpnI*) to ensure in-frame cloning in pQE-50 expression vector. An intermediate cloning was conducted in the pBlueScriptSK+ vector. The intermediate constructs were sequenced to verify the absence of Taq-induced mutations, then digested with *BamHI* and *KpnI*, and the insert was isolated and cloned into the pQE-50 expression vector [190].

The constructs were then transformed into an *E. coli* strain carrying the plasmid pFabeR, which catalyzes lycopene production [177]. After induction with IPTG, the cells were collected and the carotenoids were extracted (see Materials and Methods). Cells transformed with the pQE50/LCY-e plasmids shown an evident colour change from red (lycopene) to orange (Fig. 51). HPLC analysis confirmed that the cells have expressing LCY-e (Fig. 52) contain δ-carotene; additionally, the enzyme from *S. arcanum* converts lycopene into δ-carotene more efficiently than that of *S. lycopersicum* (Fig. 53). Thus, *S. arcanum* indeed seems to carry a

hypermorphic allele of *LCY-e.* These preliminary data must be repeated and normalized for the levels of expression of the two proteins in *E. coli*. However, since the deduced amino acid sequences of *S. arcanum* and *S. lycopersicum LCY-e* are very similar, we believe that their expression level in *E. coli* will be comparable.

5. Conclusions

We took a candidate gene approach to identify genetic differences responsible for the variability of the different colour of ripe berries in *S. lycopersicum* and it closely related wild tomato species. We focused our attention on the carotenoid biosynthetic pathway. For this reason, some genes of this pathway have been analyzed in the wild tomato species with different berry colour. The carotenoid genes have been studied thanks to the availability of the genome sequences and annotation data of tomato, potato and *A. thaliana.* Bioinformatic analyses are performed to compare the genes among these organisms. The study has highlighted the presence of three new genes, annotated respectively as putative *LCY-b*, putative *CrtISO* and putative *PSY*. The putative *LCY-b* gene seems to encode a functional protein, in fact, preliminary analyses on RNAseq data obtained with 454 sequencing show that this gene is actually expressed. The phylogenetic tree analysis moreover, shows that putative *LCY-b* arose from a recent duplication of *LCY-b,* predating the tomato-potato divergence (approx. 10 My before present). The putative *CrtISO* seems also to encode a functional protein, which is more similar to bacterial CrtI than the canonical one. Although the putative *PSY* gene is expressed, it seems to encode a non-functional, truncated protein protein.

The sequencing of carotenoid genes from *PSY* down to *LCY-e* (α-branch) and *CHY1-2* (β-branch) has been completed for all the wild specie object of this thesis. The sequence analysis has highlighted the presence of numerous mutations that

differentiate the colour-fruited species from the green-fruited ones. Some non-synonymous substitutions are candidate to be hypo or hypermorphic alleles. Preliminary analysis on *in vitro* protein expression seems to confirm that *S. arcanum* carries a hypermorphic allele of *LCY-e.*

Green-fruited species did not differ systematically from coloured-fruited ones in climacteric ethylene production or fruit softening, with the exception of *S. arcanum*, which shows an ethylene peak just before fruit abscission.

Not many data were available before this thesis on the carotenoid profiles of the berries of the wild species. The analysis of carotenoid composition and the expression profile of the genes involved in their synthesis, have been performed in berries from 5 wild tomato species representing the main phenotypes of interest: 2 red, 1 orange and 2 green-fruited species. We observed that the red-fruited species, *S. lycopersicum* and *S. pimpinellifolium*, are very similar both for biochemical and transcriptional profiles. This is in agreement with their phylogenetic closeness. In contrast with what we expected, the orange and the green-fruited species show similar biochemical and transcriptional profiles. They accumulate essentially xanthophylls and chlorophylls, although *S. cheesmaniae* to a lesser extent respect to the green-fruited species. This species has 5-fold less carotenoids than *S. lycopersicum*.

Furthermore, in orange and green-fruited species, the expression of *LCY-e*, *LCY-b* and *CYC-b* is much higher than that of red-fruited species. Moreover, *PSY1* expression in orange and green fruits is practically undetectable. On the contrary, *PSY2* is more expressed, suggesting that this gene, and not *PSY1*, is responsible for

the synthesis of carotenoids in these species. Thus, the emergence of red fruits in the tomato clade entailed the coordinated down-regulation of *PSY2, LCY-e*, *LCY-b* and *CYC-b,* the up-regulation of *PSY1 and CHY2* and the appearance of a hypomorphic LCY-e protein. The mechanism through which all these coordinated changes occurred deserves further investigation.

6. Abbreviations

ABA, abscisic acid

ACC, 1-aminocyclopropane-1-carboxylic acid

AFLP, amplified fragment length polymorphism

AIDS, acquired Immune Deficiency Sindrome

AO, antioxidant content

AOA, overall antioxidant activity

ARP, anti-radical power

BAC, bacterial artificial chromosome

BIN, discrete marker with unique composition partitioned by the Ils

cDNA, complementary DNA

CDS, coding sequence

CMV, cucumber mosaic virus

cTP, chloroplast transit peptide

DDB, UV-DAMAGED DNA-BINDING PROTEIN

DEPC, diethylenepyrocarbonate

dN, numbers of non-synonymous substitutions per non-synonymous site

DNA, deoxyribonucleic acid

dNTPs, deoxynucleotide triphosphates

DOXP, 1-deoxyxylulose 5-phosphate

DPA, days post-anthesis

DPME, 4-diphosphocytidyl-2C-methyl-D-erythritol

dS, numbers of synonymous substitutions per synonymous site

DW, dry tissue weight

EDTA, Ethylenediaminetetraacetic acid

EST, expressed sequence tag

FAD, flavin adenine dinucleotide

FW, fresh tissue weight

Gb, giga bases

GC-MS, Gas chromatography-mass spectrometry

HPLC, High performance liquid chromatography

IL, introgression line

ILP, inhibition of lipid peroxidation

IPP, isopentenyl diphosphate

IPTG, isopropylthio-β-galactoside

LC-APCI-MS, Liquid Chromatography/Atmospheric Pressure Chemical Ionization Mass Spectrometric

LHCs, light-harvesting complexes

My*:* million Years

Mb: mega bases

NAD, nicotinamide adenine dinucleotide

NJ, Neighbor Joining method

NPQ, nonphotochemical quenching

OD, optical density

PCR, polymerase chain reaction

QTL, quantitative trait locus

rDNA, ribosomal DNA

RIL, recombinant inbred lines

RNA, ribonucleic acid

RT-PCR, real time PCR

SDS, sodium dodecyl sulfate

SI, self-incompatible

SOL, Solanaceae Genome Project

Taq, DNA polymerase from *Thermus acquaticus*

TGRC, Tomato Genetics Resource Centre

TYLCV, tomato yellow leaf curl virus

USDA, United States Department of Agriculture

UTR, untranslated region

WGS, whole genome shotgun sequencing

Genes

ACO, ACC oxidase

ACS, ACC synthase

AtpB, ATP Synthase B

CCS, capsanthin-capsorubin synthase

CHY, β-carotene hydroxylase

CrtB, bacterial phytoene synthase

CrtI, bacterial phytoene desaturase

CrtISO, carotene isomerase

CrtY, bacterial lycopene β-cyclase

CrtZ, bacterial carotene hydroxylase

CYC-b, chromoplast-specific lycopene β-cyclase

DXE, DOXP reductoisomerase

DXS, 1-deoxyxylulose 5-phosphate synthase

EIN, ETHYLENE-INSENSITIVE

EIL, EIN-like

EREBP, Ethylene Response Element Binding Protein

GBSSI, Granule-bound starch synthase

GGPS, geranylgeranyl diphosphate synthase

IPI, IPP isomerase

ISP-e, DPME kinase

ISP-f, 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase

IspD, DPME synthase

LCY-b, lycopene β-cyclase

LCY-e, lycopene ε–cyclase

LUT, LUTEIN DEFICIENT

LytB, (E)-4-hydroxy-3-methylbut-2-enyl diphosphate reductase

NXS, neoxanthin synthase

PDS, phytoene desaturase

PSY, phytoene synthase

RbcL, ribulose-bisphosphate carboxylase large subunit

VDE, violaxanthin deepoxidase

ZDS, ζ-carotene desaturase

ZEP, zeaxanthin epoxidase

Mutants

at, apricot

B, Beta

B72, mutant deficient in *PHYTOCHROME B2*

c, crimson allele

chs, chilling-sensitive

Cnr, colourless non-ripening

Del, Delta

DET, Deetiolated

dg, dark green

dps, diospyros

fri, far red insensitive

gh, ghost

gf, green flesh

Gr, green-ripe

hp, high pigment

nor, nonripening

Nr, never ripe

og , Beta allele 2

r, yellow flesh

rin, ripening-inhibitor

sh, sherry

t, tangerine

tri, temporarily red light insensitive

7. Tables

Lycopersicon Name	*Solanum* Classification
L. esculentum	*S. lycopersicum*
L. pimpinellifolium	*S. pimpinellifolium*
L. cheesmanii	*S. cheesmaniae*
	S. galapagense
L. chilense	*S. chilense*
L. chmielewskii	*S. chmielewskii*
L. hirsutum	*S. habrochaites*
L. parviflorum	*S. neorickii*
L. pennellii	*S. pennellii*
L. peruvianum	*S. arcanum*
	S. corneliomuelleri
	S. huaylasense
	S. peruvianum

Tab. 1: Summary of the changes in taxonomic treatment of the section *Lycopersicon.*

Species	Daylength Preference	Sowing Date (Davis)*	Mating System	Pollination Method	# Plants / Gener.	# Plants per gal. pot	Notes
L. cheesmanii	short day	Nov - wk 4	autogamous (SC)	self	10	2	Seed produced under low light conditions is of poor quality.
L. chilense	short day	July - wk 2	allogamous (SI)	mass sib	50	5	
L. chmielewskii	day neutral	May - wk 2	facultative (SC)	mass sib	50	5	
L. esculentum var. *cerasiforme*	day neutral	April - wk 2	autogamous (SC)	self	6	(field)	
L. hirsutum	short day	July - wk 4				3	Forms edema on leaves under high humidity.
f. *glabratum*			facultative (SC)	self	15		
f. *typicum*			allogamous (SI) or facultative	mass sib	50		
L. parviflorum	day neutral	May - wk 2	autogamous (SC)	self	15	3	
L. pennellii	day neutral	June - wk 1	allogamous (SI) or facultative (SC)	mass sib	50	5	Use well-drained soil and water sparingly.
L. peruvianum	mostly day neutral	June - wk 4	allogamous (SI) or facultative (SC)	mass sib	50	5	f. *glandulosum* and mountain races are short day.
L. pimpinellifolium							
selfing pops:	day neutral	April - wk 2	autogamous (SC)	self	6	(field)	
---	---	---	---	---	---	---	---
outcrossing pops:	mostly short day	Feb - wk 2	facultative (SC)	mass sib	50	5	Regenerate in greenhouse to limit outcrossing.
S. juglandifolium	short day	Aug - wk 2	allogamous (SI)	mass sib	50	3 (2 gal pot)	Graft onto LA4135; water heavily; 8 mo. seed maturation.
S. lycopersicoides	short day	Aug - wk 2	allogamous (SI)	mass sib	50	3	Pollinate in spring months; 6 mo. seed maturation.
S. ochranthum	short day	Aug - wk 2	allogamous (SI)	mass sib	50	2 (2 gal pot)	Graft onto LA4135; water heavily; 10 mo. seed maturation.
S. sitiens	short day (almost neutral)	Aug - wk 2	allogamous (SI)	mass sib	50	3	Graft onto LA4135; 6 mo. seed maturation.

Tab. 2: TGRC recommendations for flowering and reproducing wild tomato species (the old taxonomic nomenclature is still shown) (http://tgrc.ucdavis.edu/spprecommed.html).

Taxon	Accession n°	Origin	Fruit Color	Mating System	Daylength preference	Altitude (m)	Latitude (degree)
Modern tomato cultivar							
S. lycopersicum (cv Heinz)	LA1706	USA	red	Autogamus-SC	day neutral	-	-
Wild ecotypes							
S. lycopersicum var. cerasiforme	LA1542	Costa Rica	red	Autogamus-SC	day neutral	800	9
Wild *Solanum* spp.							
S. pimpinellifolium	LA0722	Peru	red	Autogamus-SC	day neutral	-	-8
S. cheesmaniae	LA0428	Galapagos I.	orange	Autogamus-SC	short day	700	0
S. neorickii	LA1319	Peru	green	Autogamus-SC	day neutral	1800	-13.38
S. chmielewskii	LA1028	Peru	green	Facultative-SC	day neutral	3000	-13.55
S. chilense	LA2930	Chile	green	Allogamous-SI	day neutral	500	-25
S. arcanum	LA2172	Peru	green	Allogamous-SI	day neutral	-	-6
S. habrochaites	LA1353	Peru	green	Allogamous-SI	short day	2650	-7
S. pennellii	LA0716	Peru	green	Facultative-SC	day neutral	50	-16

Tab. 3: Analysed genotypes and their main characteristics.

Gene		Expression in *S. lycopersicum*	Protein (aa)	Initial available data			Obtained data: promoter + 5'UTR (bp)
				GenBank accession n°	Features	Promoter availability	
PSY1	Phytoene synthase 1	Chromoplast specific	412	EF157835	Gene	Yes	1400
PSY2	Phytoene synthase 2	Chloroplast specific	438	X60440	Gene	No	2000
PDS	Phytoene desaturase	Chloroplast+Chromoplast	583	X78271	Gene	Yes	1500
ZDS	ζ-carotene desaturase	Chloroplast+Chromoplast	588	DQ412572	Complete CDS	No	700
CrtISO	Carotene isomerase	Chloroplast+Chromoplast	615	AF416727	Gene	Yes	1900
LCY-b	Lycopene β-cyclase	Chloroplast specific	500	X86452	Complete CDS	No	600
CYC-b	Lycopene β-cyclase	Chromoplast specific	498	AF254793	Complete CDS	No	1800
LCY-e	Lycopene ε-cyclase	Chloroplast specific	527	Y14387	Complete CDS	No	1560
CHY1	β-carotene hydroxylase 1	Chloroplast specific	314	CQ793406	Complete CDS	No	1500
CHY2	β-carotene hydroxylase 2	Chromoplast specific	314	DQ650804	Gene	Yes	1200

Tab. 4: Genes belonging to carotenoid pathway analysed in this project.

Primer name	Sequence	Use
Actin Up	AGGTATTGTGTTGGACTCTGGTGAT	Real-time assay
Actin Dw	ACGGAGAATGGCATGTGGAA	Real-time assay
Tubul Up	ATATGCCCCCGTCATTTCTG	Real-time assay
Tubul Dw	GGCACTGTTAGTGATTTCTGCAA	Real-time assay
Psy1 Up	GGGCGGCCATTTGACAT	Real-time assay
Psy1 Dw	AATGGCTGAATATCAACTGGAAAGT	Real-time assay
Psy2 Up	CGGTTGATATTCAGCCATTCAG	Real-time assay
Psy2 Dw	GCGACATAGTAACAATAGAGATATAGCTCAT	Real-time assay
Pds up	TGCTGGAGGCAAGGGATG	Real-time assay
Pds dw	TGCAATCGATCGTTAATCCCTA	Real-time assay
Zds up	TTGCCATGTCAAAGGCCA	Real-time assay
Zds dw	ACAGGCACTCCGACCAATTT	Real-time assay
CrtISO up	AATGCTGGTAGCATCGCTC	Real-time assay
CrtISO dw	ATTCCGCCAAAATGTCTGTCAC	Real-time assay
Lcy-b Up	TCGTTGGAATCGGTGGTACAG	Real-time assay
Lcy-b Dw	AGCTAGTGTCCTTGCCACCATATAA	Real-time assay
Cyc-b up	TCCCTATGGGAGGACCACTTC	Real-time assay
Cyc-b dw	CCAAGCCCCTCGACGAT	Real-time assay
Chy1 up	CCAGCAATAGCCCTCCTCAA	Real-time assay
Chy1 dw	CAGGTCCAACTGGGAATCTCTT	Real-time assay
Chy2 up	TCTTTATGGCACATGCACGAGT	Real-time assay
Chy2 dw	CGAAACAGAGGCCAGGGAC	Real-time assay
Lcy-e up	GCCACAGGTTATTCAGTCGTCA	Real-time assay
Lcy-e dw	GCCAAAGAGTGTTCCAAGCTTG	Real-time assay
LCY-eBamHI_up	CGGATCCTGTAATAGCAGTAGTGGT	LCY-e cloning
LCY-eKpnI_dw	CGGATCCCTAAAATGTAAGATAAGTTC	LCY-e cloning

Tab. 5: Sequences of the primers used for cloning of the genes and for real-time quantification of transcript levels.

	Time (min)	A%	B%	ul/min
0	0	0	100%	1000
1	45	57	43	1000
2	65	82	18	1000
3	69	100	0	1000
4	85	100	0	1000
5	87	0	100	1000
6	107	0	100	1000

Tab. 6: HPLC method used for analyzing tomato fruit carotenoids: Solvent A: MeOH/*tert*-butylmethyl ether (1:1 [v/v]) and solvent B: MeOH/*tert*-butyl-methyl ether/water (60:12:12 [v/v]).

Amino acid sequence	Ver 1.0 scaffold (ITAG)	Amino acid sequences with signficant aligment	Residue n°	Scaffold region (bp)	Identities	E value (cutoff 1e-10)	Functional description by ITAG
PSY1	SL1.00sc06741_45.1.1	**PSY1**	412	356079-359380	412/412 (100%)	0	Phytoene synthase PSY / geranylgeranyl-diphosphate geranylgeranyl transferase; contains Interpro domain(s) IPR002060 Squalene/phytoene synthase
	SL1.00sc00226_682.1.1	PSY2	438	6523749-6526780	340/433 (78%)	0	Phytoene synthase PSY / geranylgeranyl-diphosphate geranylgeranyl transferase; contains Interpro domain(s) IPR019845 Squalene/phytoene synthase, conserved site IPR002060 Squalene/phytoene synthase
	SL1.00sc02227_88.1.1	PutativePSY	215	613095-614330	148/211 (70%)	3E-84	Phytoene synthase, chloroplastic; contains Interpro domain(s) IPR008949 Terpenoid synthase
PSY2	SL1.00sc00226_682.1.1	**PSY2**	438	6523749-6526780	438/438 (100%)	0	Phytoene synthase PSY / geranylgeranyl-diphosphate geranylgeranyl transferase; contains Interpro domain(s) IPR019845 Squalene/phytoene synthase, conserved site IPR002060 Squalene/phytoene synthase
	SL1.00sc06741_45.1.1	PSY1	412	356079-359380	340/433 (78%)	0	Phytoene synthase PSY / geranylgeranyl-diphosphate geranylgeranyl transferase; contains Interpro domain(s) IPR002060 Squalene/phytoene synthase
	SL1.00sc02227_88.1.1	PutativePSY	215	613095-614330	149/210 (70%)	8E-84	Phytoene synthase, chloroplastic; contains Interpro domain(s) IPR008949 Terpenoid synthase
PDS	SL1.00sc04007_627.1.1	**PDS**	583	4695410-4702265	583/583 (100%)	0	Phytoene desaturase; contains Interpro domain(s) IPR014102 Phytoene desaturase
	SL1.00sc05858_507.1.1	ZDS	588	4274433-4281957	171/533 (32%)	1E-66	Zeta-carotene desaturase; contains Interpro domain(s) IPR014103 Carotene 7,8-desaturase
ZDS	SL1.00sc05858_507.1.1	**ZDS**	588	4274433-4281957	533/533(100%)	0	Zeta-carotene desaturase; contains Interpro domain(s) IPR014103 Carotene 7,8-desaturase
	SL1.00sc04007_627.1.1	PDS	583	4695410-4702265	171/533 (32%)	1E-66	Phytoene desaturase; contains Interpro domain(s) IPR014102 Phytoene desaturase
CrtISO	SL1.00sc05992_456.1.1	**CrtISO**	615	6311561-6316109	615/615 (100%)	0	Carotenoid isomerase; contains Interpro domain(s) IPR014101 Carotene isomerase
	SL1.00sc00395_74.1.1	Putative CrtISO	568	734404-738943	166/516 (32%)	7E-57	Carotenoid isomerase, chloroplastic; contains Interpro domain(s) IPR006076 FAD dependent oxidoreductase
LCY-b	SL1.00sc06399_26.1.1	**LCY-b**	500	544580-546082	500/500 (100%)	0	Lycopene beta-cyclase; contains Interpro domain(s) IPR010108 Lycopene cyclase, beta and epsilon
	SL1.00sc05992_247.1.1	Putative LCY-b	500	4675967-4677469	436/500 (87%)	0	Beta-lycopene cyclase; contains Interpro domain(s) IPR010108 Lycopene cyclase, beta and epsilon
	SL1.00sc06019_130.1.1	CYC-b	498	942434-943930	252/455 (55%)	1E-148	Chromoplast-specific lycopene beta-cyclase; contains Interpro domain(s) IPR010108 Lycopene cyclase, beta and epsilon
	SL1.00sc07122_65.1.1	LCY-e	527	555727-560682	166/442 (37%)	6E-78	Lycopene epsilon cyclase; contains Interpro domain(s) IPR010108 Lycopene cyclase, beta and epsilon
CYC-b	SL1.00sc06019_130.1.1	**CYC-b**	498	942434-943930	494/498 (99%)	0	Chromoplast-specific lycopene beta-cyclase; contains Interpro domain(s) IPR010108 Lycopene cyclase, beta and epsilon
	SL1.00sc06399_26.1.1	LCY-b	500	544580-546082	251/455 (55%)	1E-147	Lycopene beta-cyclase; contains Interpro domain(s) IPR010108 Lycopene cyclase, beta and epsilon
	SL1.00sc05992_247.1.1	Putative LCY-b	500	4675967-4677469	248/441 (56%)	1E-147	Beta-lycopene cyclase; contains Interpro domain(s) IPR010108 Lycopene cyclase, beta and epsilon
	SL1.00sc07122_65.1.1	LCY-e	527	555727-560682	166/414 (40%)	7E-74	Lycopene epsilon cyclase; contains Interpro domain(s) IPR010108 Lycopene cyclase, beta and epsilon
LCY-e	SL1.00sc07122_65.1.1	**LCY-e**	527	555727-560682	527/527 (100%)	0	Lycopene epsilon cyclase; contains Interpro domain(s) IPR010108 Lycopene cyclase, beta and epsilon
	SL1.00sc06399_26.1.1	LCY-b	500	544580-546082	166/442 (37%)	6E-78	Lycopene beta-cyclase; contains Interpro domain(s) IPR010108 Lycopene cyclase, beta and epsilon
	SL1.00sc05992_247.1.1	Putative LCY-b	500	4675967-4677469	165/442 (37%)	3E-76	Beta-lycopene cyclase; contains Interpro domain(s) IPR010108 Lycopene cyclase, beta and epsilon
	SL1.00sc06019_130.1.1	CYC-b	498	942434-943930	165/442 (37%)	7E-74	Chromoplast-specific lycopene beta-cyclase; contains Interpro domain(s) IPR010108 Lycopene cyclase, beta and epsilon
CHY1	SL1.00sc06217_29.1.1	**CHY1**	314	539172-540726	309/309 (100%)	1E-180	Beta-carotene hydroxylase 1; contains Interpro domain(s) IPR006694 Fatty acid hydroxylase
	SL1.00sc06183_135.1.1	CHY2	314	1308061-1310125	232/311 (74%)	1E-135	Beta-carotene hydroxylase; contains Interpro domain(s) IPR006694 Fatty acid hydroxylase
CHY2	SL1.00sc06183_135.1.1	**CHY2**	314	1308061-1310125	313/314 (99%)	0	Beta-carotene hydroxylase; contains Interpro domain(s) IPR006694 Fatty acid hydroxylase
	SL1.00sc06217_29.1.1	CHY1	314	539172-540726	232/311 (74%)	1E-135	Beta-carotene hydroxylase 1; contains Interpro domain(s) IPR006694 Fatty acid hydroxylase

Tab. 7: Summary of data generated from the tBlastn alignment using as database the 1.0 tomato genome release annotated by ITAG.

Gene	Ver 1.0 scaffold	Ver 1.03 scaffold	IL map position		Loci	Position (cM)	Chromatin	Strand
			Chr.	Bin				
PSY1	scaffold06741	scaffold00066	3	C	TES1084	53.443	Hetero	+
					TES1317	53.41	Hetero	
PSY2	scaffold00226	scaffold00215	2	H	C2_At5g64670	85.534	Eu	+
					T702	85.903	Eu	
PSY (putative)	scaffold02227	scaffold01157	1	D	TGS3437	35.84	Hetero	-
					TGS0972	34.801	Hetero	
PDS	scaffold04007	scaffold01712	3	I	tme0062	131.964	Eu	-
					C2_At5g49830	133.514	Eu	
ZDS	scaffold05858	scaffold00106	1	G	cLPT-5-M7	100.214	Eu	-
					TES0114	94.765	Eu	
CrtISO	scaffold05992	scaffold01076	10	E	TEI0272	57.423	Eu	-
					TES0075	57.315	Eu	
CrtISO (putative)	scaffold00395	scaffold00208	5	C	CT93	61.444	Eu	-
					TES0151	47.911	Eu	
LCY-b	scaffold06399	scaffold00762	4	E	n. a.	46*	n. a.	-
					n. a.	62*	n. a.	
LCY-b (putative)	scaffold05992	scaffold01076	10	E	TES1177	52.616	Eu	+
					cLEX-11-C19	49.776	Eu	
CYC-b	scaffold06019	scaffold00823	6	E	TGS1973	80.069	Eu	+
					TES0771	74.948	Eu	
LCY-e	scaffold07122	scaffold00061	12	C	TEI0737	35.647	Eu	+
					TM26	33.003	Eu	
CHY1	scaffold06217	scaffold00934	6	A	TGS0330	5.769	Hetero	+
					T1500	11.067	Hetero	
CHY2	scaffold06183	scaffold01216	3	C	TES1139	5.635	Eu	+
					TGS0827	4.421	Eu	

Tab. 8: Mapping data: the position of the genes on the genome is shown. For each genes is indicated the IL map position and 2 markers (loci) that flanking the start and the stop of the CDS (n.a.= not available). *The cM position of *LCY-b* is calculated on the base of the BIN position in *S. pennellii* ILs.

PSY A. thaliana				*PYS2 S. lycopersicum*			
Scaffold	**Region (bp)**	**Strand**	**Protein function**	**Scaffold**	**Region (bp)**	**Strand**	**Protein function**
	5225804.. 5226850		receptor protein kinase-like protein		6418884..6419198	-	Light-induced protein, chloroplastic
	5579143 - 5580780	-	NADP-dependent oxidoreductase, putative		6426353..6426911	+	zinc finger protein-related #
	5584801 - 5587063	-	NADP-dependent oxidoreductase, putative		6447356..6447840	+	Expansin-A13
	5587351 - 5592446	-	sugar transporter		6455708..6456653	+	ATGSTU25 GLUTATHIONE S-TRANSFERASE
	5594736 - 5602955	+	exportin 1A		6467687..6467773	+	Ran GTPase binding
	5603133 - 5604738	-	UDP-glucosyltransferase		6488020..6494024	+	NAC domain protein IPR003441
	5605284 - 5606970	-	UDP-glucosyltransferase		6500692..6502827	+	DEAD-box ATP-dependent RNA helicase 21
	5607788 - 5609492	-	anthocyanidin 3-O-glucosyltransferase		6504542..6505112	-	Sucrose synthase
	5610806 - 5613065	+	ARF GTPase family		6511435..6511698	-	50S ribosomal protein L15
	5619594 - 5620949	+	Peptidase C1A		6517334..6517338	-	SET DOMAIN GROUP 40
	5636972 - 5637534	+	Cysteine-type peptidase		**6523749..6524232**	+	***PSY2***
	5637830 - 5638273	+	cysteine proteinase-related		6530501..6531625	-	2-succinylbenzoate--CoA ligase
	5649193 - 5651158	+	enhancer of sos3-1 (ENH1)		6529006..6530098	+	Glutathione S-transferase
	5653524 - 5655336	-	polygalacturonase (pectinase) family protein		6533133..6533815	-	Acyl coa ligase acetate-coa synthetase-like protein
	5658460 - 5659386	+	Glutathione S-transferase		6543338..6543375	+	2-succinylbenzoate--CoA ligase
	5659574-5662865	-	***PSY***	00226	6553545..6553580	+	2-succinylbenzoate--CoA ligase
Chr 5	5666852 - 5668907	+	SET DOMAIN GROUP 40		6559060..6559160	-	GTP binding protein
	5668900 - 5673516	-	catalytic/ transferase		6564270..6564587	-	Polyamine oxidase
	5675323 - 5677912	-	NAC domain containing protein 86		6574332..6574520	-	Intracellular protease 1
	5679991 - 5685597	+	tetratricopeptide repeat (TPR)		6586466..6586682	-	tRNA-splicing endonuclease positive effector-related
	5686945 – 5689900	-	Autophagy protein ATG5		6594494..6594606	+	Microsomal glutathione S-transferase 3
	5690227 - 5692902	-	Myb-like transcription factor		6602520..6602686	-	Gamma-tubulin complex component 2
	5696645 - 5700922	-	UTP-glucose-1-phosphate uridylyltransferase		6615882..6616259	-	Receptor-like protein kinase
	5703380 - 5707637	-	HDG9, HOMEODOMAIN GLABROUS9		6616464..6617755	-	Stress-induced receptor-like kinase
	5711012 - 5715075	+	GLUTAMATE DECARBOXYLASE		6621763..6622183	+	Kinase R-like protein
	5721801 - 5724719	-	WD-40 repeat family protein		6628506..6628581	+	serine/threonine protein kinase family protein
	5724759 - 5726746	-	pyruvate decarboxylase family protein		6637312..6637318	-	Nibrin
	5727640 - 5728688	+	universal stress protein (USP) family protein		6631331..6632783	+	Receptor-like kinase
	5728789 - 5730376	-	Endoplasmatic reticulim nucleotide transp.		6647811..6649034	+	ubiquitin thiolesterase
	5730610 - 5736695	+	tubulin family protein		6653363..6653943	+	Cell division protease ftsH homolog 6, chloroplastic
	5736630 - 5741465	-	cellulose synthase catalytic subunit		6657464..6661409	-	Oligopeptide transporter OPT family
	7758310 - 7775674	+	zinc finger C3HC4-type RING finger family protein				

Tab. 9: Comparison between the genes flanking *A. thaliana PSY* and tomato *PSY2* (highlighted in yellow). Proteins with the same functions are highlighted with the same colour. There are 6-shared genes and their relative position and orientation is conserved, with the exception of the gene encoding the tubulin-related protein that is inverted with respect to *PSY.* In *A. thaliana*, the gene encoding the zinc finger-related protein is distant 2 Mb from *PSY.* The black row indicates not included genes.

PSY1 S. lycopersicum			
Scaffold	**Region (bp)**	**Strand**	**Protein function**
06741	228267..228555	-	Glucose-6-phosphate/phosphate translocator
	306317..307339	-	Magnesium transporter protein 1
	312213..312393	+	Xyloglucan endotransglucosylase/hydrolase 3
	316915..316959	-	Protein translation factor Sui1 homolog
	313747..314193	+	Xyloglucan endotransglycosylase/hydrolase XTH-25
	336340..337077	+	zinc finger C3HC4-type RING finger family protein
	340488..340981	+	Expansin-A13
	356079..356490	+	***PSY1***
	364728..364925	+	2-succinylbenzoate--CoA ligase
	372508..372816	-	Polyamine oxidase
	386511..387602	-	Proteinase inhibitor II
	393892..394036	-	tRNA-splicing endonuclease positive effector-related
	406142..407260	-	Oligopeptide transporter OPT family
	432896..433344	-	2-succinylbenzoate--CoA ligase
	485219..485261	+	Auxin response factor 8
	493683..494482	-	Lectin protein kinase family protein

PSY2 S. lycopersicum			
Scaffold	**Region (bp)**	**Strand**	**Protein function**
00226	6418884..6419198	-	Light-induced protein, chloroplastic
	6426353..6426911	+	zinc finger protein-related #
	6447356..6447840	+	Expansin-A13
	6455708..6456653	+	ATGSTU25 GLUTATHIONE S-TRANSFERASE
	6467687..6467773	+	Ran GTPase binding
	6488020..6494024	+	NAC domain protein IPR003441
	6500692..6502827	+	DEAD-box ATP-dependent RNA helicase 21
	6504542..6505112	-	Sucrose synthase
	6511435..6511698	-	50S ribosomal protein L15
	6517334..6517338	-	SET DOMAIN GROUP 40
	6523749..6524232	+	***PSY2***
	6530501..6531625	-	2-succinylbenzoate--CoA ligase
	6529006..6530098	+	Glutathione S-transferase
	6533133..6533815	-	Acyl coa ligase acetate-coa synthetase-like protein
	6543338..6543375	+	2-succinylbenzoate--CoA ligase
	6553545..6553580	+	2-succinylbenzoate--CoA ligase
	6559060..6559160	-	GTP binding protein
	6564270..6564587	-	Polyamine oxidase
	6574332..6574520	-	Intracellular protease 1
	6586466..6586682	-	tRNA-splicing endonuclease positive effector-related
	6594494..6594606	+	Microsomal glutathione S-transferase 3
	6602520..6602686	-	Gamma-tubulin complex component 2
	6615882..6616259	-	Receptor-like protein kinase
	6616464..6617755	-	Stress-induced receptor-like kinase
	6621763..6622183	+	Kinase R-like protein
	6628506..6628581	+	serine/threonine protein kinase family protein
	6637312..6637318	-	Nibrin
	6631331..6632783	+	Receptor-like kinase
	6647811..6649034	+	ubiquitin thiolesterase
	6653363..6653943	+	Cell division protease ftsH homolog 6, chloroplastic
	6657464..6661409	-	Oligopeptide transporter OPT family

Tab. 10: Comparison between the genes flanking tomato *PSY1* and *PSY2* (highlighted in yellow). Genes with the same functions are highlighted with the same colours. In *PSY2* table, the pale grey-highlighted boxes are syntenic genes with *A. thaliana PSY*. There are 6 syntenic genes and their relative position and orientation is almost conserved.

PSY A. thaliana				Putative *PSY S. lycopersicum*			
Scaffold	**Region (bp)**	**Strand**	**Protein function**	**Scaffold**	**Region (bp)**	**Strand**	**Protein function**
	5225804.. 5226850		receptor protein kinase-like protein		13349-15273		GAD2 GLUTAMATE DECARBOXYLASE 2
	5579143 - 5580780	-	NADP-dependent oxidoreductase, putative				
	5584801 - 5587063	-	NADP-dependent oxidoreductase, putative		452294..453165	-	zinc finger protein-related #
	5587351 - 5592446	-	sugar transporter		469732..469827	+	protein homodimerization
	5594736 - 5602955	+	exportin 1A		475246..476190	-	SUB1 calcium ion binding
	5603133 - 5604738	-	UDP-glucosyltransferase		484446..485904	+	AtRLP2 Receptor Like Protein
	5605284 - 5606970	-	UDP-glucosyltransferase		487865..488252	+	Receptor-like protein kinase 2
	5607788 - 5609492	-	anthocyanidin 3-O-glucosyltransferase		493651..496212	+	Receptor protein kinase-like protein
	5610806 - 5613065	+	ARF GTPase family		498945..501251	+	Receptor-like protein kinase 2
	5619594 - 5620949	+	Peptidase C1A		502719..504095	-	pentatricopeptide PPR
	5636972 - 5637534	+	Cysteine-type peptidase		570053..570126	+	Ubiquitin carrier protein
	5637830 - 5638273	+	cysteine proteinase-related		576660..577571	+	Extensin-like protein Dif10
	5649193 - 5651158	+	enhancer of sos3-1 (ENH1)		579046..581610	-	Receptor-like protein kinase 2
	5653524 - 5655336	-	polygalacturonase (pectinase) family protein		582954..583997	+	Extensin-like protein Dif10
	5658460 - 5659386	+	Glutathione S-transferase	02227	585567..585713	+	Extensin-like protein Dif10
	5659574-5662865	-	***PSY***		587599..587861	-	Taxadien-5-alpha-ol O-acetyltransferase
Chr 5	5666852 - 5668907	+	SET DOMAIN GROUP 40		598051..599478	-	Neutral amino acid transport protein
	5668900 - 5673516	-	catalytic/ transferase		**613095..613148**	+	***Putative PSY***
	5675323 - 5677912	-	NAC domain containing protein 86		623766..624473	+	DEAD-box ATP-dependent RNA helicase
	5679991 - 5685597	+	tetratricopeptide repeat (TPR)		638967..639029	+	kelch repeat-containing F-box family protein
	5686945 – 5689900	-	Autophagy protein ATG5		669661..669945	+	Protein quaking-A
	5690227 - 5692902	-	Myb-like transcription factor		675917..676111	-	At3g32930-like protein
	5696645 - 5700922	-	UTP-glucose-1-phosphate uridylyltransferase		709975..710106	-	HUB2 HISTONE MONO-UBIQUITINATION
	5703380 - 5707637	-	HDG9, HOMEODOMAIN GLABROUS9		734111..734689	+	disease resistance response
	5711012 - 5715075	+	GLUTAMATE DECARBOXYLASE		784965..785022	-	CDT1A
	5721801 - 5724719	-	WD-40 repeat family protein		792741..792920	-	Metal tolerance protein
	5724759 - 5726746	-	pyruvate decarboxylase family protein		804379..805269	-	rRNA processing protein-related
	5727640 - 5728688	+	universal stress protein (USP) family protein				
	5728789 - 5730376	-	Endoplasmatic reticulum nucleotide transp.		2105540-2108653		ADP-ribosylation factor
	5730610 - 5736695	+	tubulin family protein				
	5736630 - 5741465	-	cellulose synthase catalytic subunit				
	7758310 - 7775674	+	zinc finger C3HC4-type RING finger family protein				

Tab. 11: Comparison between the genes flanking *A. thaliana PSY* and putative *PSY* (highlighted in yellow). Genes with the same functions are highlighted with the same colours. The pale grey-highlighted boxes are syntenic genes between *PSY2* and *A. thaliana PSY*. In *A. thaliana*, the gene encoding the zinc finger-related protein is distant 2 Mb with respect to *PSY* and in *S. lycopersicum,* the genes encoding the glutamate decarboxylase and the ADP-ribosylation factor proteins are distant respectively 0,5 MB and 1,5 Mb from putative *PSY*. Black rows indicate not included genes.

Putative *PSY S. lycopersicum*			
Scaffold	**Region (bp)**	**Strand**	**Protein function**
	13349-15273		GAD2 GLUTAMATE DECARBOXYLASE 2
	452294..453165	-	zinc finger protein-related #
	469732..469827	+	protein homodimerization
	475246..476190	-	SUB1 calcium ion binding
	484446..485904	+	AtRLP2 Receptor Like Protein
	487865..488252	+	Receptor-like protein kinase 2
	493651..496212	+	Receptor protein kinase-like protein
	498945..501251	+	Receptor-like protein kinase 2
	502719..504095	-	pentatricopeptide PPR
	570053..570126	+	Ubiquitin carrier protein
	576660..577571	+	Extensin-like protein Dif10
	579046..581610	-	Receptor-like protein kinase 2
	582954..583997	+	Extensin-like protein Dif10
02227	585567..585713	+	Extensin-like protein Dif10
	587599..587861	-	Taxadien-5-alpha-ol O-acetyltransferase
	598051..599478	-	Neutral amino acid transport protein
	613095..613148	**+**	***Putative PSY***
	623766..624473	+	DEAD-box ATP-dependent RNA helicase
	638967..639029	+	kelch repeat-containing F-box family protein
	669661..669945	+	Protein quaking-A
	675917..676111	-	At3g32930-like protein
	709975..710106	-	HUB2 HISTONE MONO-UBIQUITINATION
	734111..734689	+	disease resistance response
	784965..785022	-	CDT1A
	792741..792920	-	Metal tolerance protein
	804379..805269	-	rRNA processing protein-related
	2105540-2108653		ADP-ribosylation factor

PSY2 S. lycopersicum			
Scaffold	**Region (bp)**	**Strand**	**Protein function**
	6418884..6419198	-	Light-induced protein, chloroplastic
	6426353..6426911	+	zinc finger protein-related #
	6447356..6447840	+	Expansin-A13
	6455708..6456653	+	ATGSTU25 GLUTATHIONE S-TRANSFERASE
	6467687..6467773	+	Ran GTPase binding
	6488020..6494024	+	NAC domain protein IPR003441
	6500692..6502827	+	DEAD-box ATP-dependent RNA helicase 21
	6504542..6505112	-	Sucrose synthase
	6511435..6511698	-	50S ribosomal protein L15
	6517334..6517338	-	SET DOMAIN GROUP 40
	6523749..6524232	**+**	***PSY2***
	6530501..6531625	-	2-succinylbenzoate--CoA ligase
	6529006..6530098	+	Glutathione S-transferase
	6533133..6533815	-	Acyl coa ligase acetate-coa synthetase-like protein
	6543338..6543375	+	2-succinylbenzoate--CoA ligase
00226	6553545..6553580	+	2-succinylbenzoate--CoA ligase
	6559060..6559160	-	GTP binding protein
	6564270..6564587	-	Polyamine oxidase
	6574332..6574520	-	Intracellular protease 1
	6586466..6586682	-	tRNA-splicing endonuclease positive effector-related
	6594494..6594606	+	Microsomal glutathione S-transferase 3
	6602520..6602686	-	Gamma-tubulin complex component 2
	6615882..6616259	-	Receptor-like protein kinase
	6616464..6617755	-	Stress-induced receptor-like kinase
	6621763..6622183	+	Kinase R-like protein
	6628506..6628581	+	serine/threonine protein kinase family protein
	6637312..6637318	-	Nibrin
	6631331..6632783	+	Receptor-like kinase
	6647811..6649034	+	ubiquitin thiolesterase
	6653363..6653943	+	Cell division protease ftsH homolog 6, chloroplastic
	6657464..6661409	-	Oligopeptide transporter OPT family

Tab. 12: Comparison between the genes flanking tomato *PSY2* and putative *PSY* (highlighted in yellow). Genes with the same functions are highlighted with the same colours. Black rows indicate not included genes.

	reads per million						
	root	stem	leaf	flower	berry		
					MG	B	10 DPB
PSY1	0	38.62	162.69	519.68	133.2	8114.02	8532.95
PSY2	19.67	43.16	114.48	6.11	12.89	12.14	10.2
Putative PSY	4.37	0	0	0	0	0	0
PDS	13.11	38.62	58.25	186.47	79.49	72.85	112.21
ZDS	65.57	43.16	156.66	42.8	64.45	39.46	137.71
CrtISO	17.49	9.09	10.04	12.23	75.19	106.24	86.71
Putative CrtISO	0	2.27	12.05	0	12.89	3.04	7.65
LCY-b	24.04	97.68	263.11	0	40.82	3.04	0
Putative LCY-b	8.74	6.82	24.1	9.17	2.15	0	0
CYC-b	6.56	4.54	2.01	0	0	3.04	0
LCY-e	0	4.54	38.16	0	0	0	0
CHY1	0	29.53	88.37	3.06	27.93	27.32	48.45
CHY2	2.19	0	0	30.57	2.15	3.04	5.1

Tab. 13: RNAseq data obtained with 454 sequencing. Values are expressed as reads per million. The putative *PSY*, putative *CrtISO*, putative *LCY-b* are actually expressed. Threshold: identity 95%. MG: mature green, B: breaker.

Promoter, 5' and 3' UTR	Intron	CDS	
		Synonymous	Non synonymous
3047	4552	526	274

Tab. 14: Summary of mutations found in all the sequenced genes in the different genomic regions. Nonsynonymous mutations are the missense mutations the synonymous mutations are the silent mutations.

CDS	dN	dS	dN/dS	
			Avg.	St. Dev.
PSY1	2.51	25.25	0.1	0.003
PSY2	4.09	22.92	0.18	0.004
ZDS	0.8	22.21	0.04	0.004
CrtISO	3.53	21.55	0.16	0.003
PDS	1.57	17.3	0.09	0.006
LCY-e	2.05	19.1	0.11	0.004
LCY-b	4.81	12.08	0.4	0.019
CYC-b	5.62	25.48	0.22	0.003
CHY1	6.14	15.27	0.4	0.014
CHY2	4.2	29.27	0.14	0.004
RbcL	0.93	3.33	0.3	0.018

Tab. 15: Value of the frequency of nonsynonymous (dN) and synonymous mutations (dS) and their ratio calculated with MEGA4.

a

Genothypes	Phytoene	Phytofluene	ζ-carotene	Poly-*cis*-lycopene	*Cis*-lycopene (1+2)	All-*trans*-lycopene	δ-carotene	α-carotene	Lutein	β-carotene	Unknown Xanthophylls	Neoxanthin + Violaxanthin	Total carotenoids	Chlorophyll a	Chlorophyll b
Heinz	1.289±0.211	0.622±0.129	0.022±0.004	0.037±0.007	0.2±0.022	0.583±0.068	0.006±0.001	0	0.196±0.047	0.252±0.032	0.005±0.001	0.007±0.002	3.217±0.525	0	0
S. pimpinellifolium	0.772±0.128	0.359±0.06	0	0.018±0.006	0.173±0.016	0.413±0.027	0.006±0.002	0.008±0.002	0.041±0.007	0.216±0.044	0.003±0.001	0	2.001±0.291	0	0
Fold variation	**0.599**	**0.578**	**0**	**0.483**	**0.865**	**0.709**	**0.99**	-	**0.212**	**0.856**	**0.651**	**0**	**0.622**	-	-
S. cheesmaniae	0	0	0	0	0	0	0	0.015±0.001	0.367±0.016	0.099±0.018	0.09±0.006	0.062±0.007	0.618±0.046	0.069±0.02	0.192±0.04
Fold variation	**0**	**0**	**0**	**0**	**0**	**0**	**0**	-	**1.874**	**0.391**	**19.518**	**9.052**	**0.192**	-	-
S. arcanum	0	0	0	0	0	0	0	0.044±0.005	0.639±0.007	0.309±0.003	0.117±0.004	0.085±0.005	1.15±0.019	0.236±0.015	0.486±0.018
Fold variation	**0**	**0**	**0**	**0**	**0**	**0**	**0**	-	**3.261**	**1.226**	**25.434**	**12.375**	**0.357**	-	-
S. neorickii	0	0	0	0	0	0.009±0.001	0	0.032±0.003	0.75±0.022	0.458±0.01	0.136±0.019	0.144±0.015	1.497±0.067	0.117±0.02	0.778±0.053
Fold variation	**0**	**0**	**0**	**0**	**0**	**0.016**	**0**	-	**3.826**	**1.817**	**29.758**	**20.992**	**0.465**	-	-

b

Genothypes	Phytoene	Phytofluene	ζ-carotene	Poly-*cis*-lycopene	*Cis*-lycopene (1+2)	All-*trans*-lycopene	δ-carotene	α-carotene	Lutein	β-carotene	Unknown Xanthophylls	Neoxanthin + Violaxanthin	Total carotenoids	Chlorophyll a	Chlorophyll b
Heinz	2.062±0.155	1.016±0.159	0	0.069±0.012	0.372±0.079	0.682±0.135	0.008±0.001	0	0.155±0.006	0.245±0.004	0.014±0.003	0	4.625±0.554	0	0
S. pimpinellifolium	1.415±0.077	0.596±0.058	0	0.022±0.002	0.18±0.035	0.546±0.081	0	0	0.071±0.018	0.19±0.031	0	0	3.019±0.301	0	0
Fold variation	**0.686**	**0.586**	-	**0.313**	**0.484**	**0.8**	**0**	-	**0.458**	**0.773**	**0**	-	**0.652**	-	-
S.cheesmaniae	0	0	0	0	0	0	0	0	0.277±0.022	0.028±0.003	0.075±0.003	0.075±0.009	0.456±0.037	0.079±0.005	0.075±0.004
Fold variation	**0**	**0**	-	**0**	**0**	**0**	**0**	-	**1.783**	**0.114**	**5.485**	-	**0.098**	-	-
S. arcanum	0	0	0	0	0	0	0	0.023±0.002	0.401±0.003	0.215±0.007	0.065±0.004	0.065±0.007	0.746±0.021	0.08±0.005	0.449±0.006
Fold variation	**0**	**0**	-	**0**	**0**	**0**	**0**	-	**2.58**	**0.878**	**4.715**	-	**0.161**	-	-
S. neorickii	0	0	0	0	0	0	0	0.026±0.001	0.674±0.008	0.4±0.009	0.13±0.009	0.094±0.005	1.299±0.032	0.188±0.019	0.608±0.020
Fold variation	**0**	**0**	-	**0**	**0**	**0**	**0**	-	**4.338**	**1.632**	**9.485**	-	**0.28**	-	-

Tab. 16: Carotenoid composition (ug/mg dry weight) measured via diode array HPLC in berry at 58 (panel a) and 65 DPA (panel b). Fold variation with respect to cultivated tomato (*S. lycopersicum* cv Heinz) is reported for each carotenoid compound and for each wild genotype.

8. Figures

Fig. 1: Some examples of *Solanaceae,* from top left: *A. belladonna*, *C. annuum*, *P. hybrida*, *C. annuum, Lycium andersonii, S. lycopersicum, S. tuberosum* and *S. melongena.*

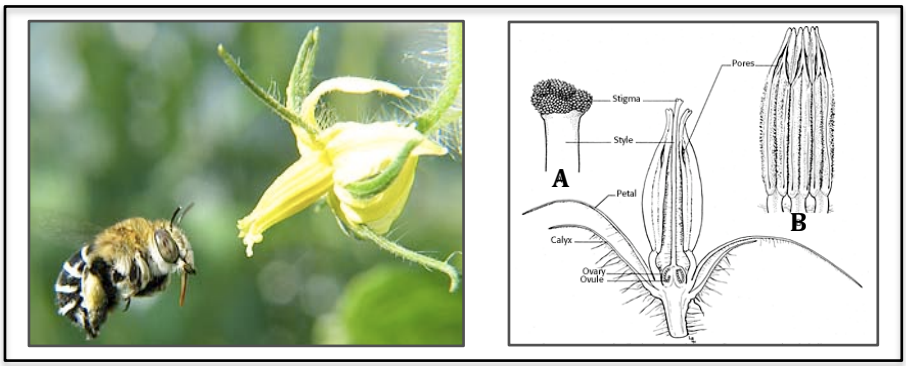


Fig. 2: Tomato flower and its longitudinal section: A, tip of pistil; B, three anthers greatly enlarged.

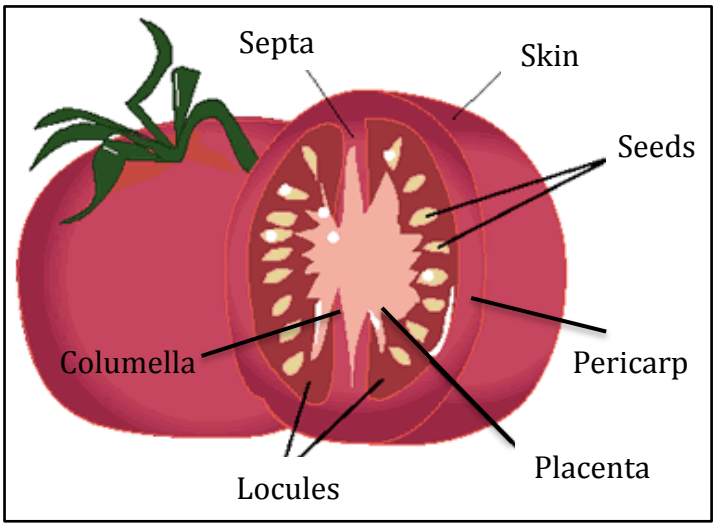


Fig. 3: Longitudinal section of tomato berry.

Fig. 4: Examples of the wide variability in tomato berry colour.

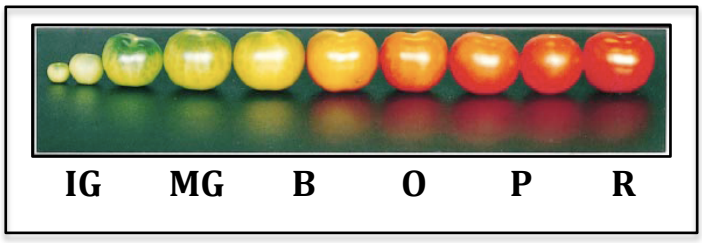


Fig. 5: Size and colour modification during tomato fruit ripening (IG, immature green; MG, mature green; B, breaker; O, orange; P, pink; R, ripe).

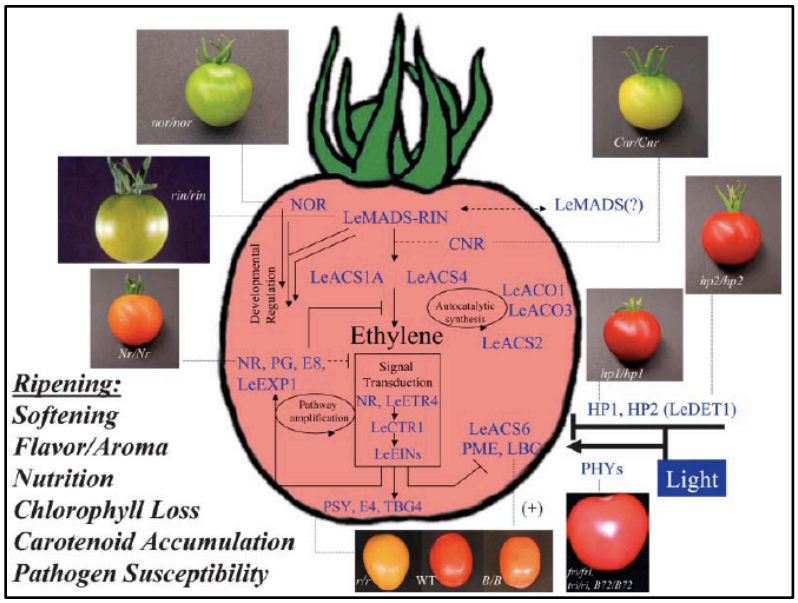


Fig. 6: Current model for the molecular regulation of tomato fruit ripening. Fruits harboring homozygous mutations for the indicated genes or loci are shown (*nor*, *rin*, *Nr*, *Cnr*, *hp1*, and *hp2*, *r* and *B, fri*, *tri* and *B72* mutants) [17].

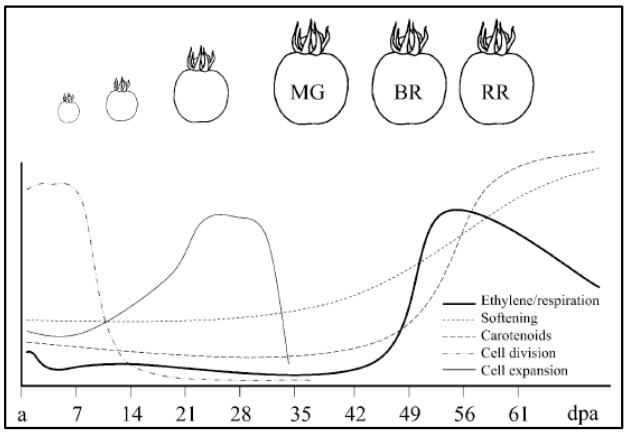


Fig. 7: Main developmental changes during tomato fruit development and ripening. Day post-anthesis (a) to mature green (MG; fully expanded unripe fruit with mature seed), breaker (BR; first visible carotenoid accumulation), and red ripe (RR) can vary substantially among cultivars [17].

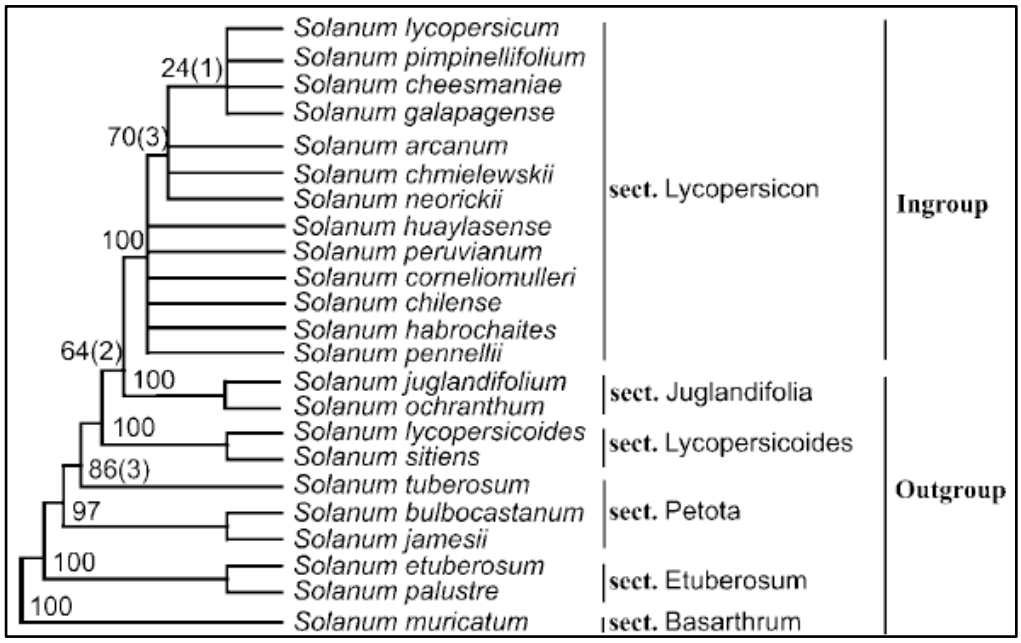


Fig. 8: Taxonomic relationships of the 13 tomato species (ingroup sect. *Lycopersicon*) and ten species in five outgroups (sect. *Juglandifolia*, sect. *Lycopersicoides*, sect. *Petota*, sect. *Etuberosum* and sect. *Basarthrum*). Numbers indicate bootstrap values [50].

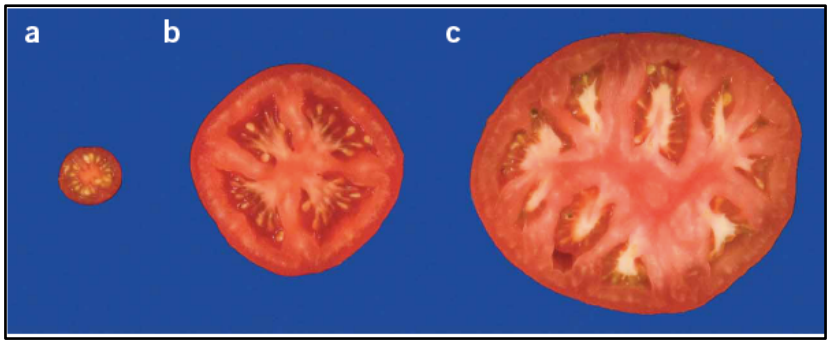


Fig. 9: Fruit size increase during tomato domestication: a, *S. pimpinellifolium*; b, an intermediate-sized processing variety (E6203); c, a large-sized fresh market variety (Jumbo Red) [55].

Fig. 10: Geographic distribution of *Solanum* sections *Lycopersicoides*, *Juglandifolia* and *Lycopersicon* [56].

Fig. 11: Examples of the natural habitats of wild tomatoes: A, B arid coastal zone or in a narrow canyon located in the Andes; C, D arid slope in the Cordillera de Domeyko (Chile) or in a shallow depression near a dry salt flat; E abandoned agricultural field; F, G a dry riverbed or on exposed slopes at high elevation in the Andes [57].

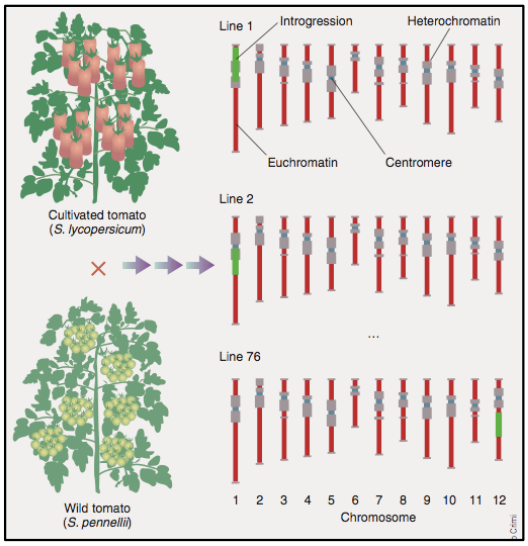


Fig. 12: Development of introgression lines to define and map QTLs for crop improvement [192].

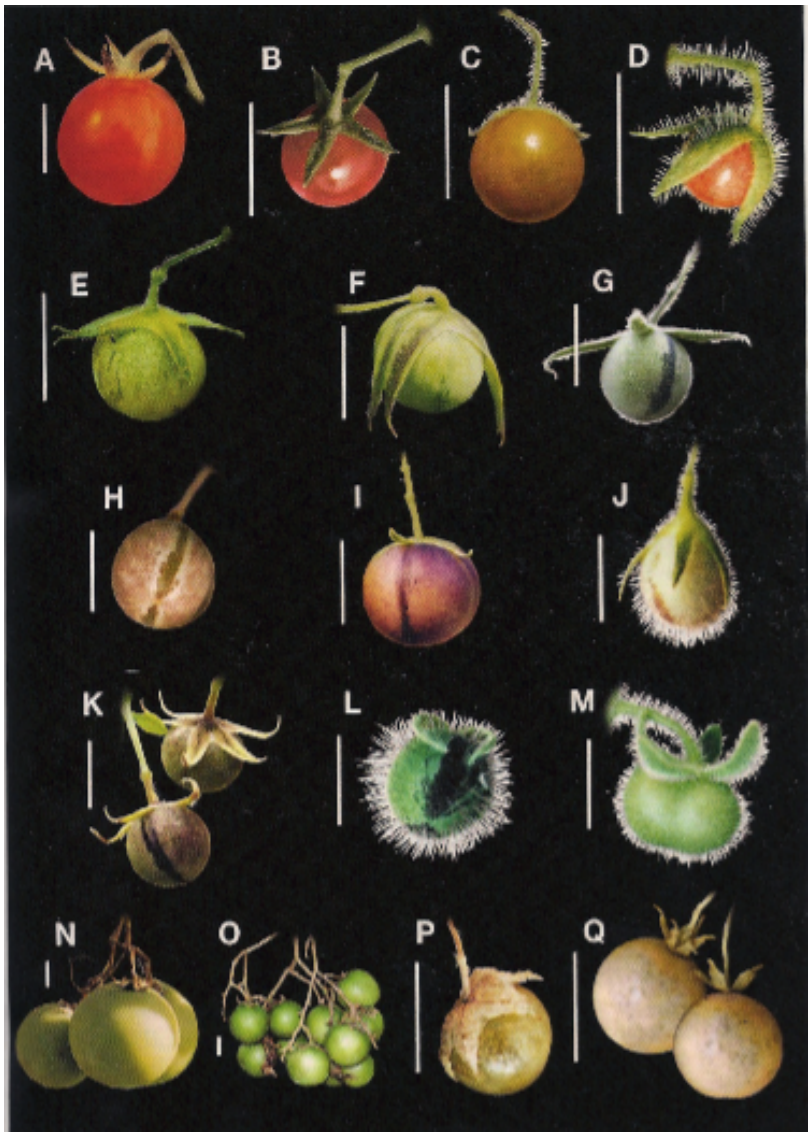


Fig. 13: Fruits of *Solanum* sections *Lycopersicoides*, *Juglandifolia* and *Lycopersicon*: A. *S. lycopersicum*; B. *S. pimpinellifolium*; C. S*. cheesmaniae*; D. *S. galapagense*; E. *S. neorickii*; F. *S. chmielewskii*; G. *S. arcanum*; H. *S. huaylasense*; I. S*. peruvianum*; J. *S. corneliomulleri*; K. *S. chilense*; L. *S. habrochaites*; M. *S. pennellii*; N. *S. ochrantum*; O. *S. juglandifolium*; P. S*. lycopersicoides*; Q. *S. sitiens*. Scale bars=1 cm [56].

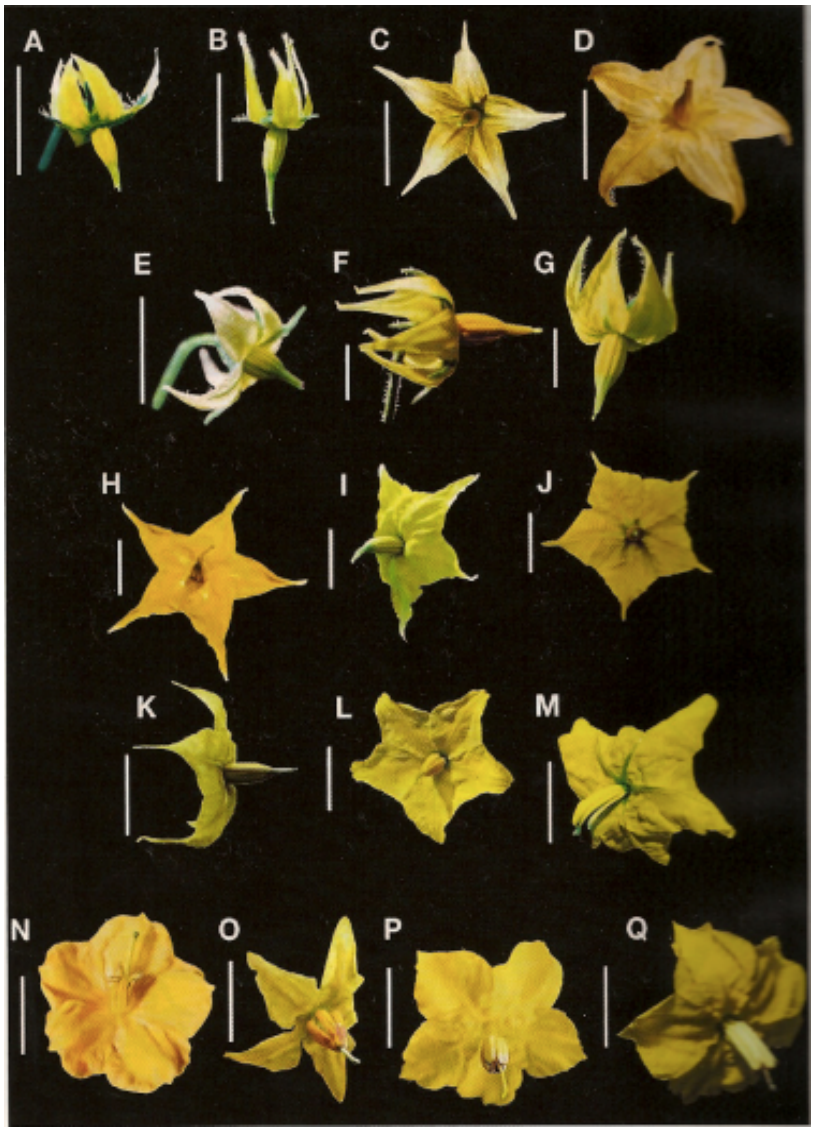


Fig. 14: Flowers of *Solanum* sections *Lycopersicoides*, *Juglandifolia* and *Lycopersicon*: A. *S. lycopersicum*; B. *S. pimpinellifolium*; C. *S. cheesmaniae*; D. *S. galapagense*; E. *S. neorickii*; F. *S. chmielewskii*; G. *S. arcanum*; H. *S. huaylasense*; I. S. *peruvianum*; J. *S. corneliomulleri*; K. *S. chilense*; L. *S. habrochaites*; M. *S. pennellii*; N. *S. ochrantum*; O. *S. juglandifolium*; P. S. *lycopersicoides*; Q. *S. sitiens*. Scale bars: A-D, H-Q=1 cm; E-G=0,5 cm [56].

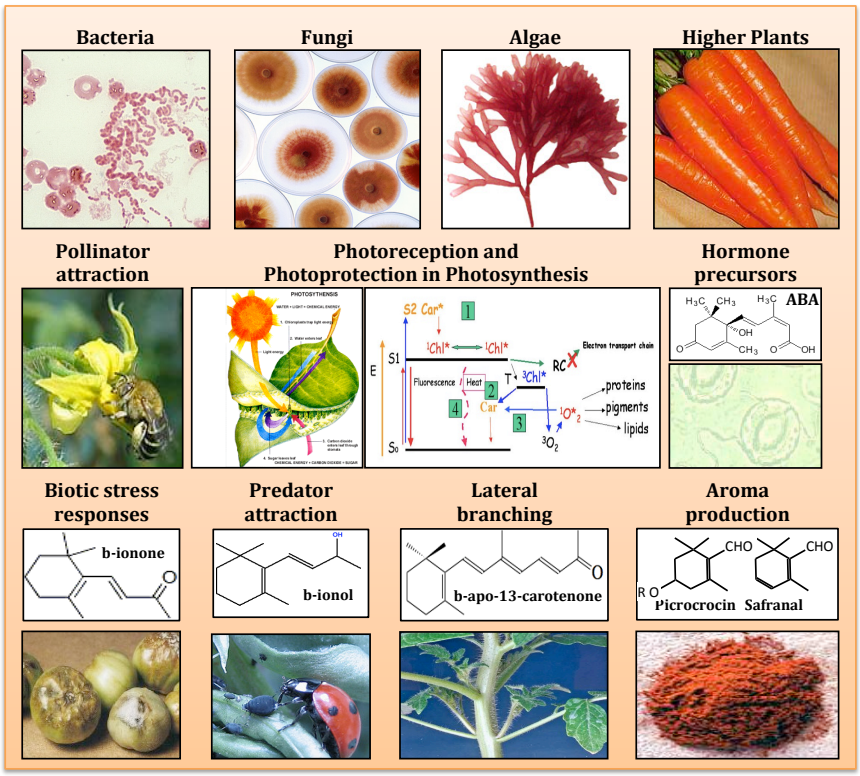


Fig. 15: Example of carotenoid and apocarotenoid distribution and their functions in plants.

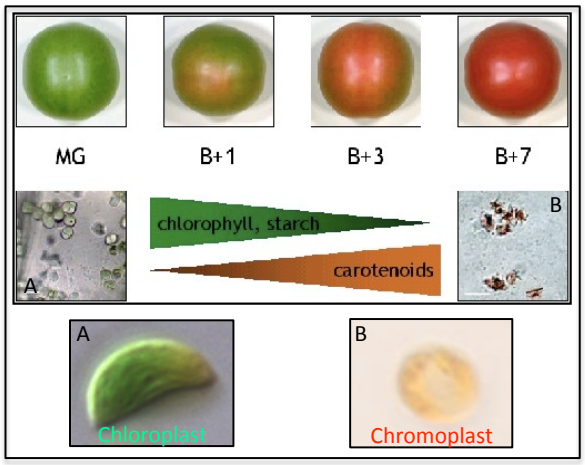


Fig. 16: Plastid development during ripening: chloroplasts, that contain carotenoids and chlorophyll, are converted into red chromoplasts; chlorophylls and starches are replaced with red carotenoids.

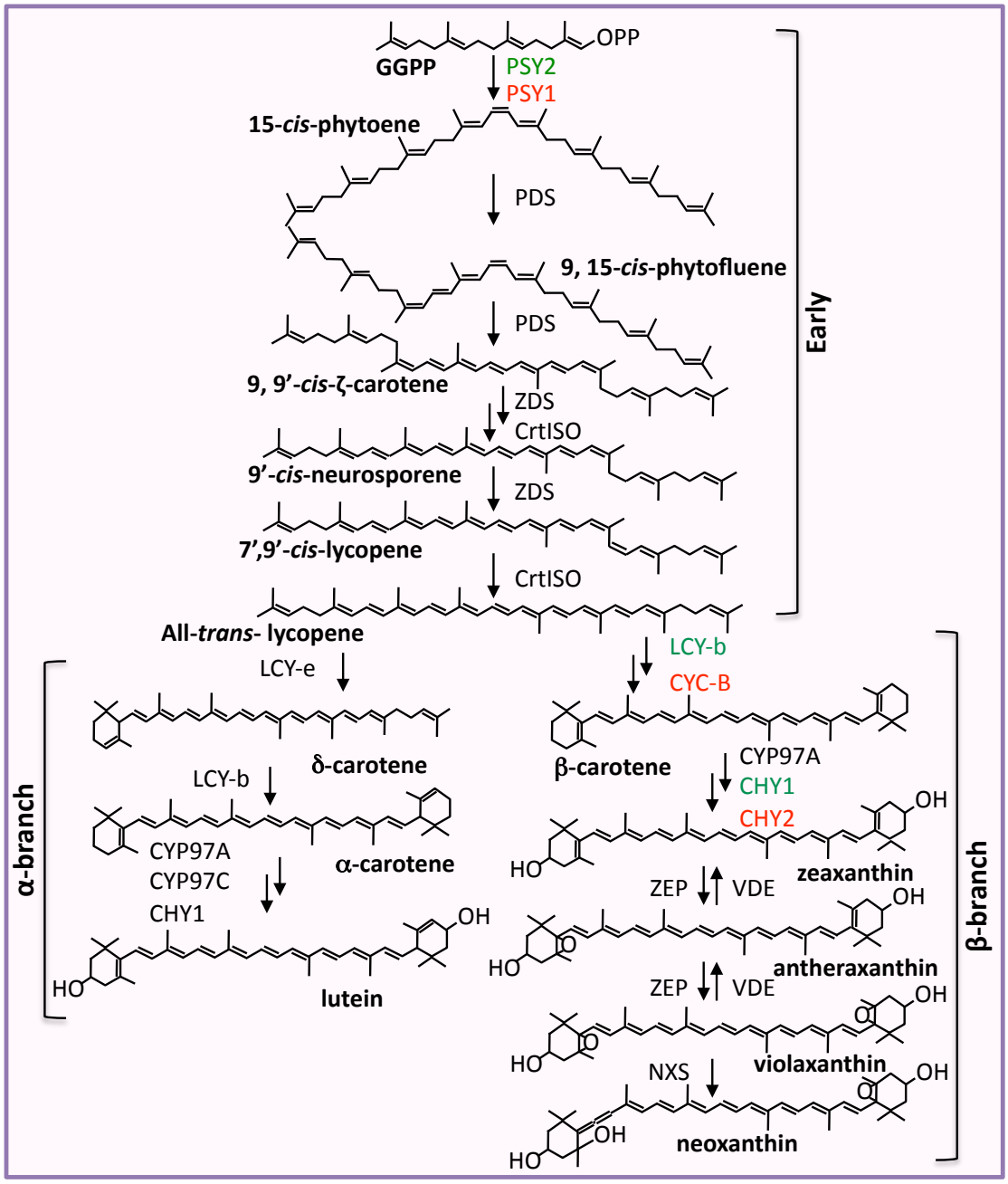


Fig. 17: Representative scheme of carotenoid biosynthesis in tomato. Some steps are mediated by multiple proteins, hypothesized to be chloroplast-specific (green) or chromoplast-specific (red).

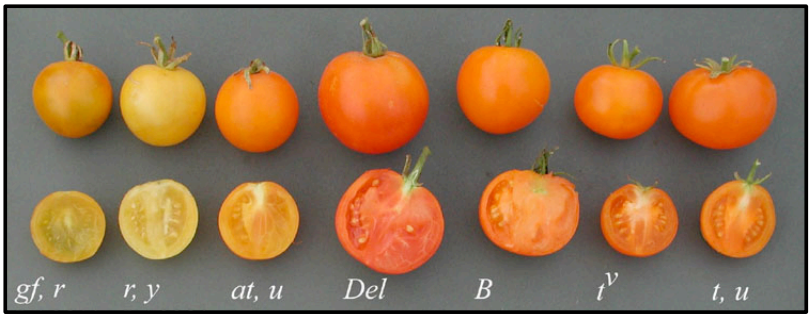


Fig. 18: Example of some fruit colour tomato mutants: *green flesh* (*gf*), *yellow flesh* (*r*), *apricot/uniform ripening* (*at*), *Delta* (*Del*), *Beta* (*B*) *tangerine/ virescent* (*t^v*), *tangerine/uniform ripening* (*t*).

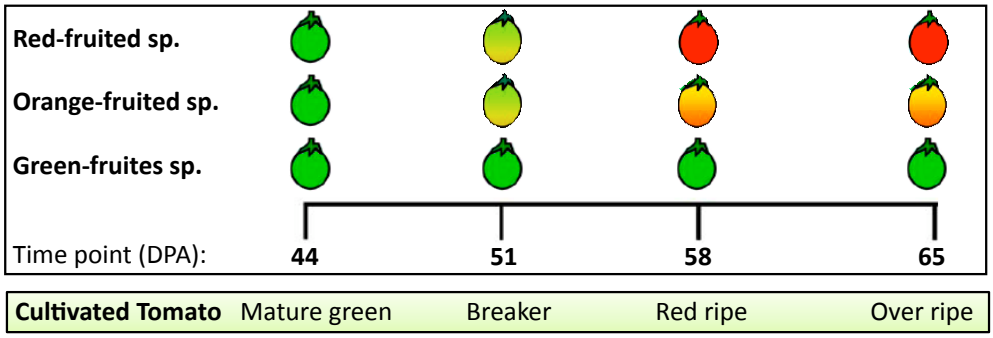


Fig. 19: Berries were collected for experiments measuring days post-anthesis (DPA). The time points 44, 51, 58 and 65 DPA correspond, in cultivated tomato, to the following developmental stages: mature green, breaker, red ripe and over ripe.

Fig. 20: Example of ethylene analysis: a berry of *S. neorickii* in a chromatography vial (11 ml).

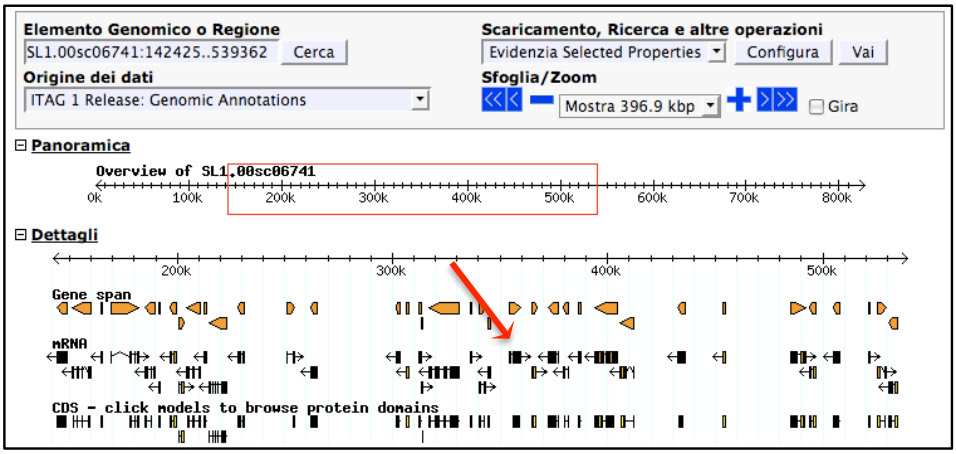


Fig. 21: *PSY1* and its flanking regions. Example of output generated from tomato genome browser: the central ruler represents the scaffold, the red box indicates the magnified region. The arrow points to the gene of interest, in this case the *PSY1*.

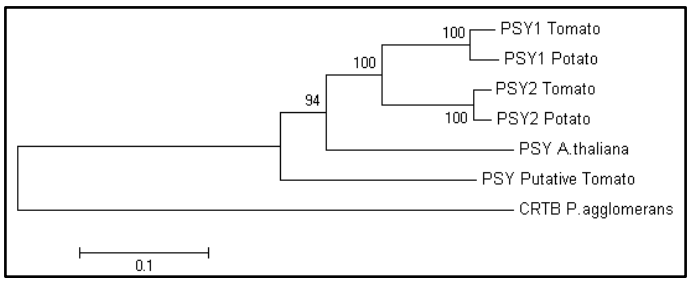


Fig. 22: The dendrogram, built on the base of the protein sequences, shows the evolutionary distance among the Phytoene synthase enzymes of different organisms. As out-group has been used the bacterial Phytoene synthase of *Pantoea agglomerans* (CrtB). The tree is generated using ClustalW and the neighbor-joining method; branch lengths are proportional to the amount of inferred evolutionary change (www.ebi.ac.uk/clustalw).

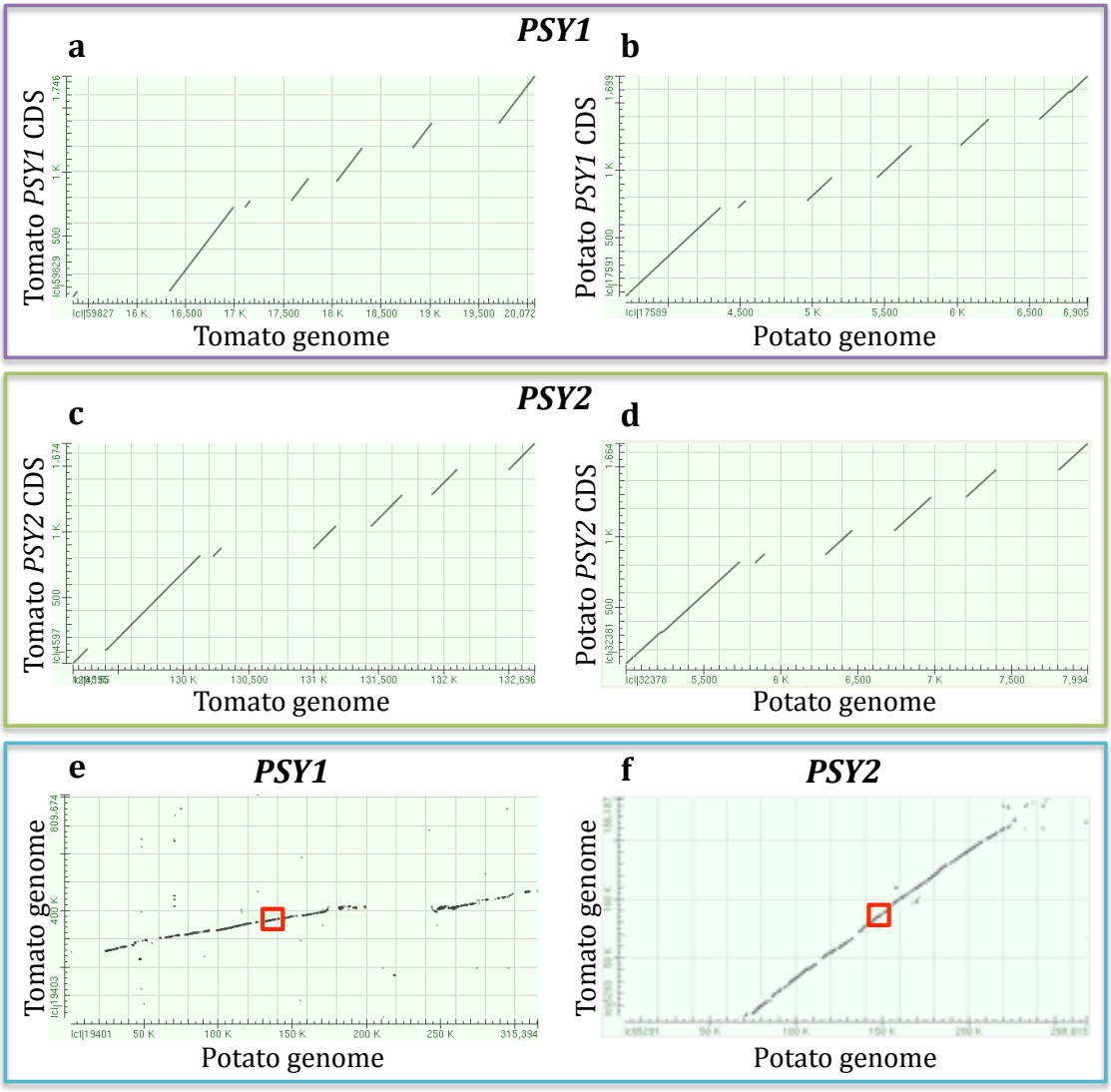


Fig. 23: Alignments based on nucleotide sequences (Blast2seq algorithm) demonstrate that the gene structure of *PSY1* (panel a and b) and of *PSY2* (panel c and d) is almost conserved in the tomato and potato genomes. Moreover, the genomic regions flanking these genes show a high level of microsynteny between these two organisms (panel e and f). The red boxes in panels e and f indicate the position of *PSY1* and *PSY2* genes.

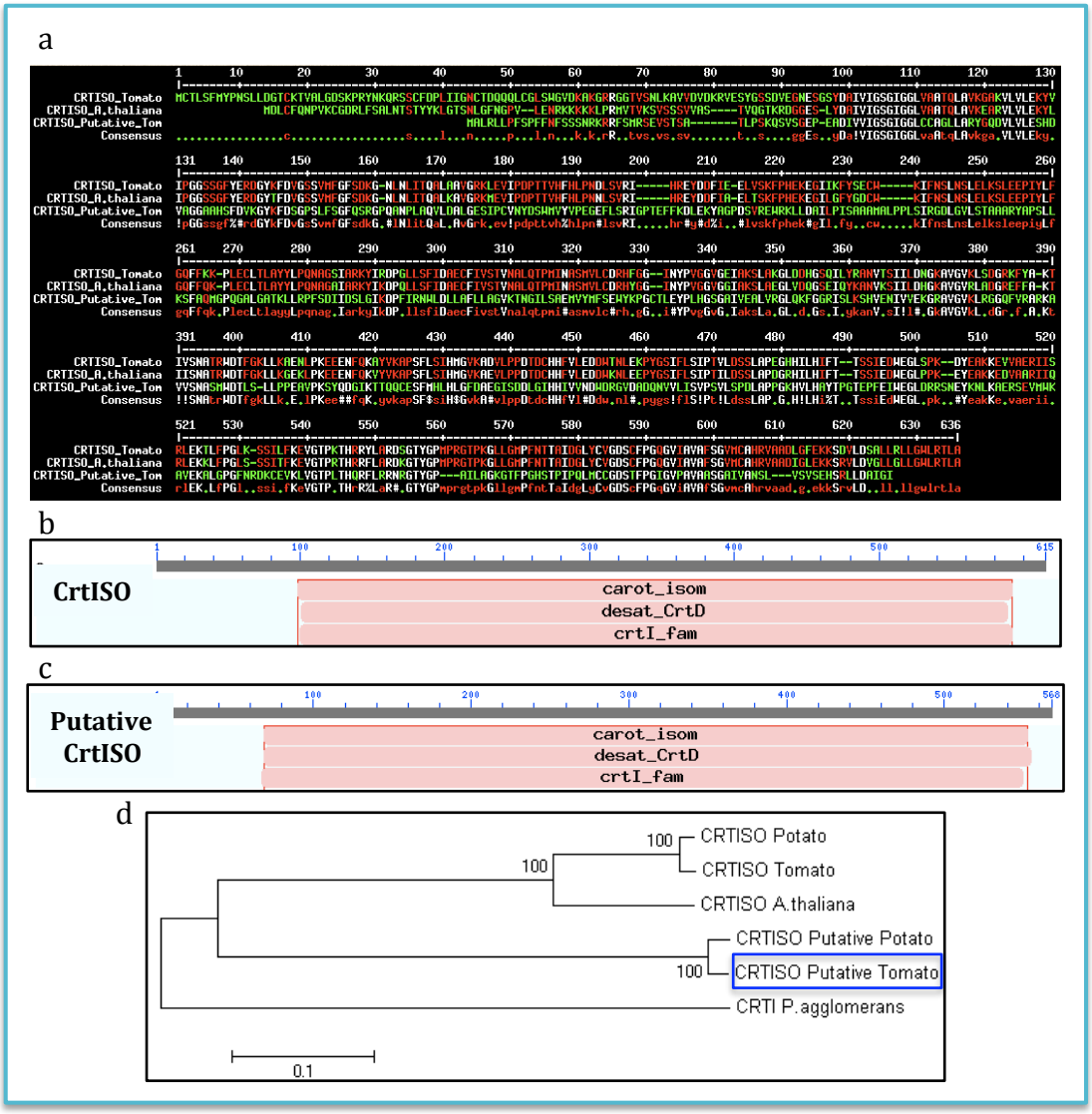


Fig. 24: Analysis of putative *CrtISO*. Panel a. The alignment, based on protein sequences (ClustalW algorithm), shows a very high divergence in the putative CrtISO amino acid sequence with respect to the other proteins. Panel b. Functional domains in CrtISO. Panel c. Functional domains in the putative CrtISO. Panel d. ClustalW dendrogram (Neighbor-joining method); the putative CrtISO is more closely related to the bacterial Carotene desaturase (CrtI) with respect to the other CrtISO.

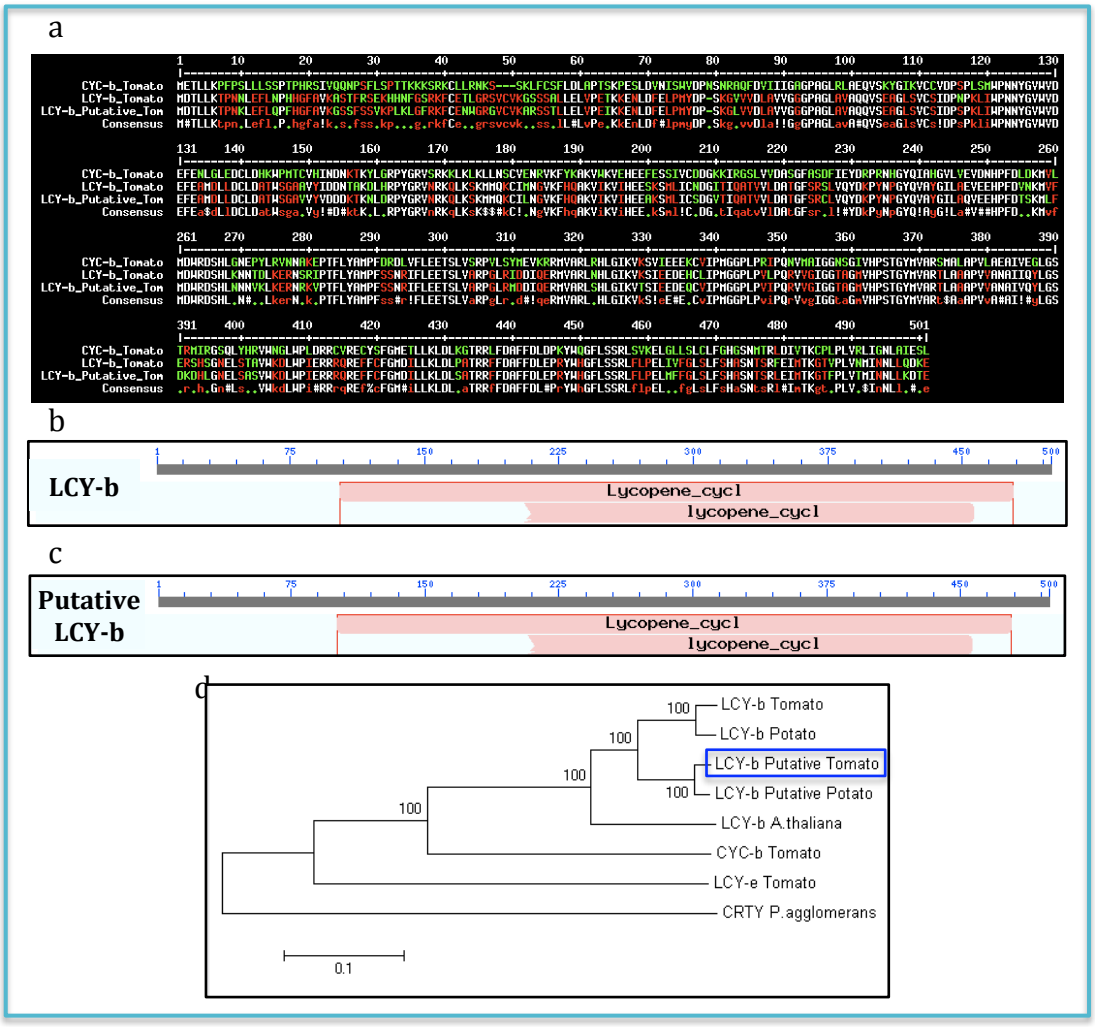


Fig. 25: Analysis of putative *LCY-b*. Panel a. The alignment, based on protein sequences (ClusatlW algorithm), shows a very high similarity among the proteins. Panel b. Functional domains in tomato LCY-b. Panel c. Functional domains in the putative LCY-b. Panel d. Dendogram (Neighbor-joining method): the putative tomato LCY-b is closely related to tomato LCY-b and their ortholog proteins in potato show the same trend.

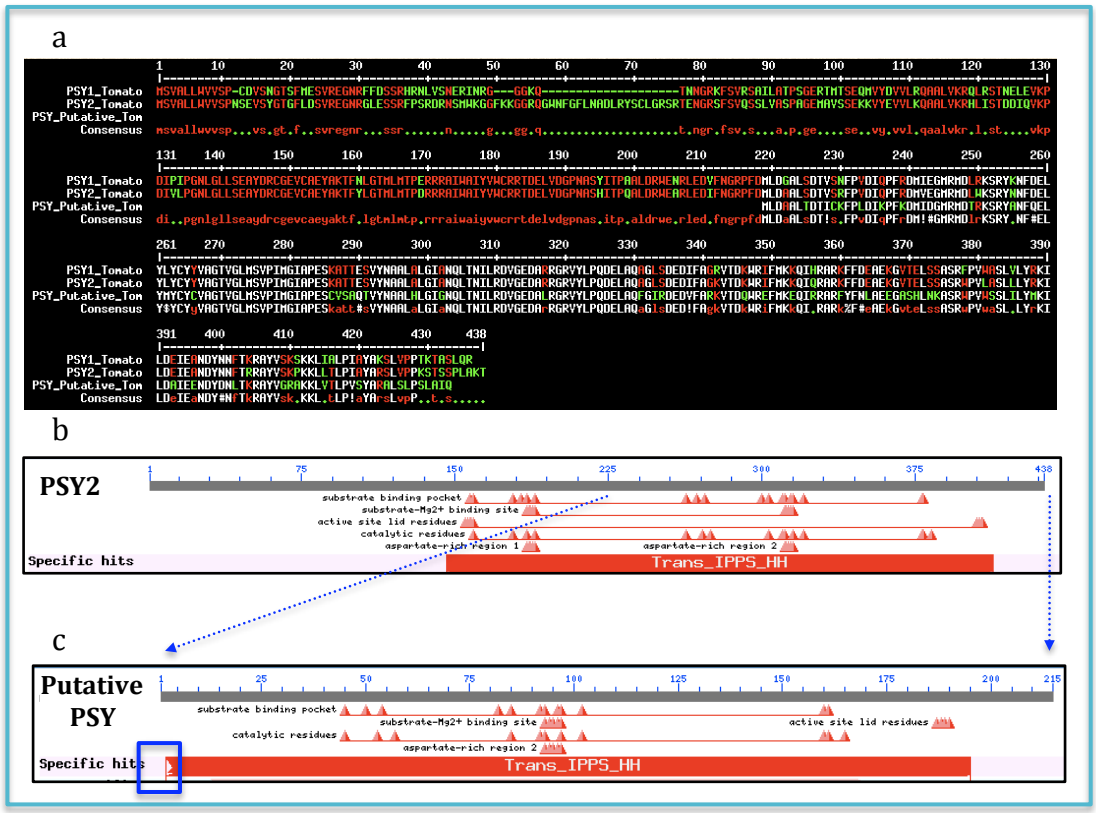


Fig. 26: Analysis of putative *PSY*. Panel a. The alignment, based on amino acid sequences (ClusatlW algorithm), shows that the putative PSY is lacking a N-terminal part with respect to the other PSY proteins. Panel b. Functional domain scheme of the PSY2 (http://www.ncbi.nlm.nih.gov/Structure). Panel c. Functional domain scheme of the putative PSY: it is lacking half of the conserved domains.

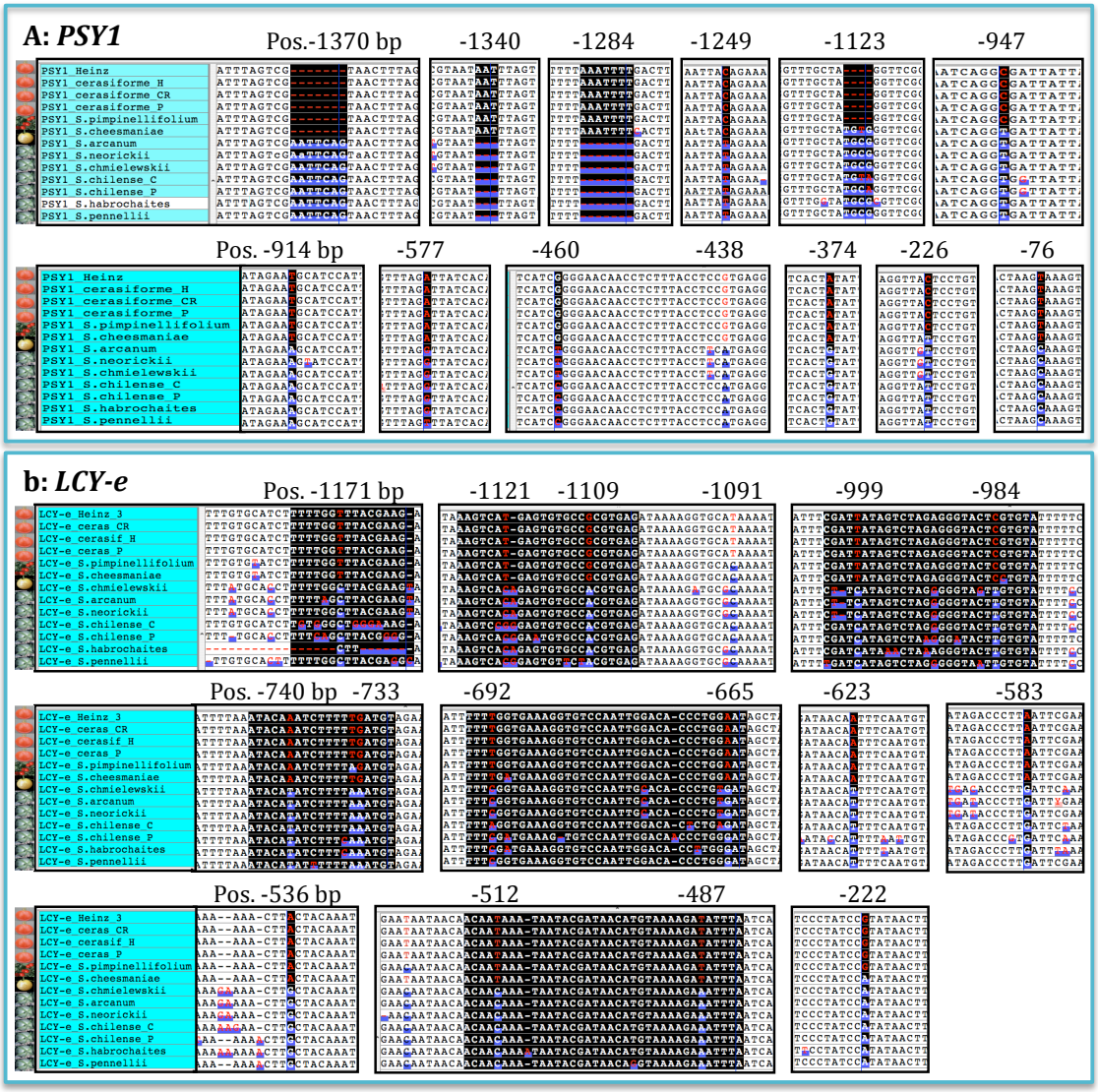


Fig. 27: Examples of promoter mutations. Panel a. Mutations found in *PSY1.* *mutations that differentiate the green berry from the others: pos. 5' to ATG: -1370, -1340, -1284, -1249, -914, -577, -460, -438, -374 and – 76; *mutations *S. lycopersicum*-specific: -1123, -947 and -226. Panel b. Mutations found in *LCY-e:* *mutations that differentiate the green berry from the others: pos. 5' to ATG: -1171, -1121, -1109, -999, -984, -740, -733, -692, -665, -623, -583, -536, -512 and -487; *red berry-specific mutations: -222; **S. lycopersicum*-specific mutation: -1091.

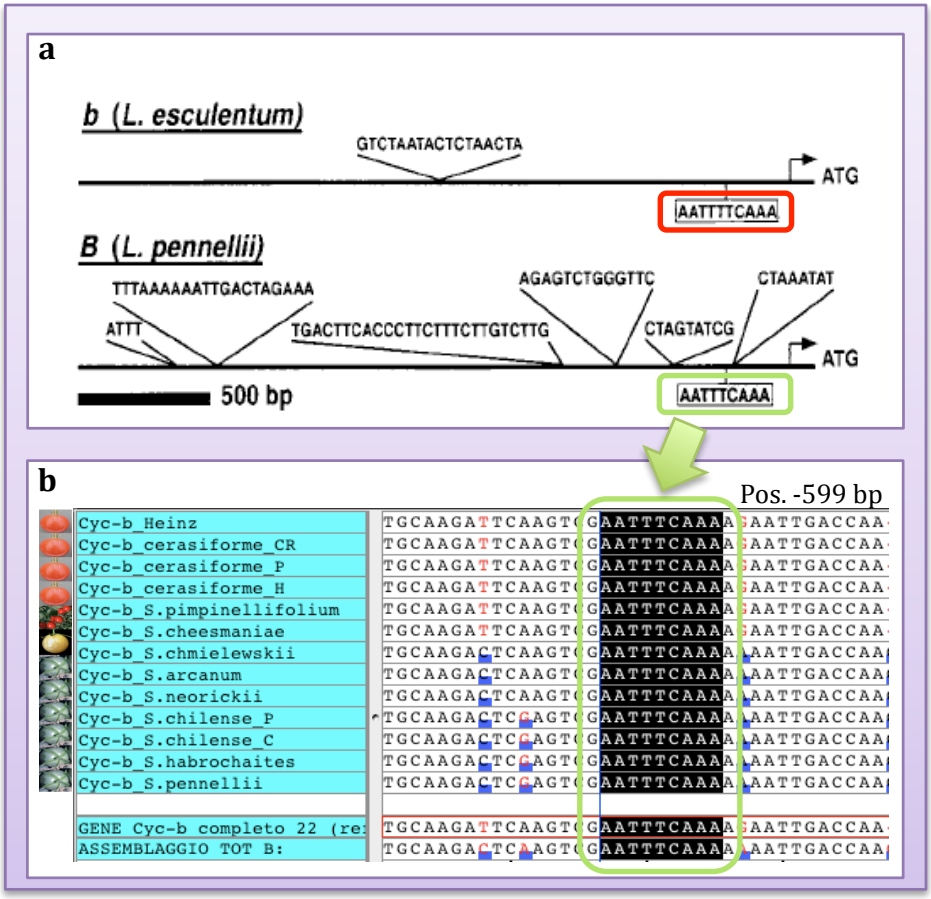


Fig. 28: Panel a. Organization of the genomic sequences upstream to the promoter of *B* and *b* according to Ronen at al 2006. In green box the putative sequence, (*B* allele from *S. pennellii*) is highlighted; in red box the mutated sequence (T insertion) is the *b* allele of *S. lycopersicum* [79]. Panel b. In our sequences the T insertion is absent. All the species have in fact the same sequence of *S. pennellii*.

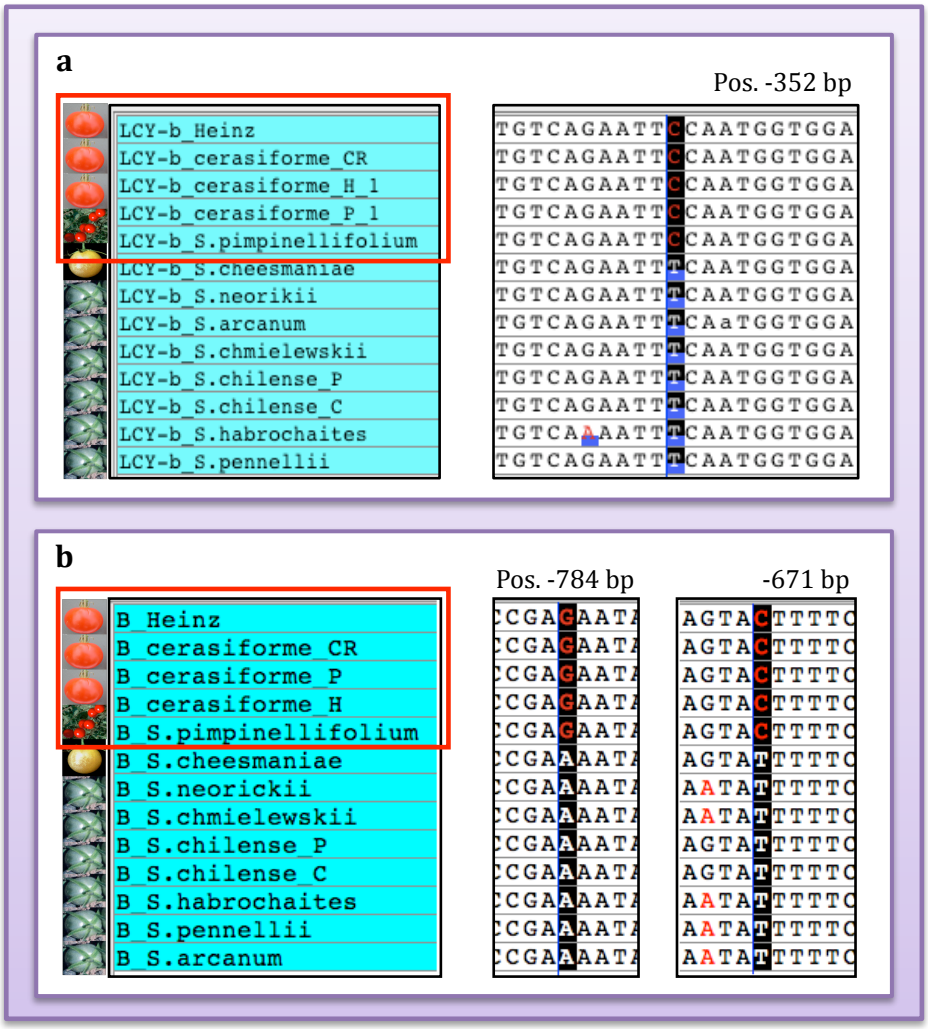


Fig. 29: Red fruit-specific mutations found in promoter of β–cyclases (position with regard to ATG): panel a. Mutation -352 in the *LCY-b*; panel b. Mutations -784 and -671 in the *CYC-b.*

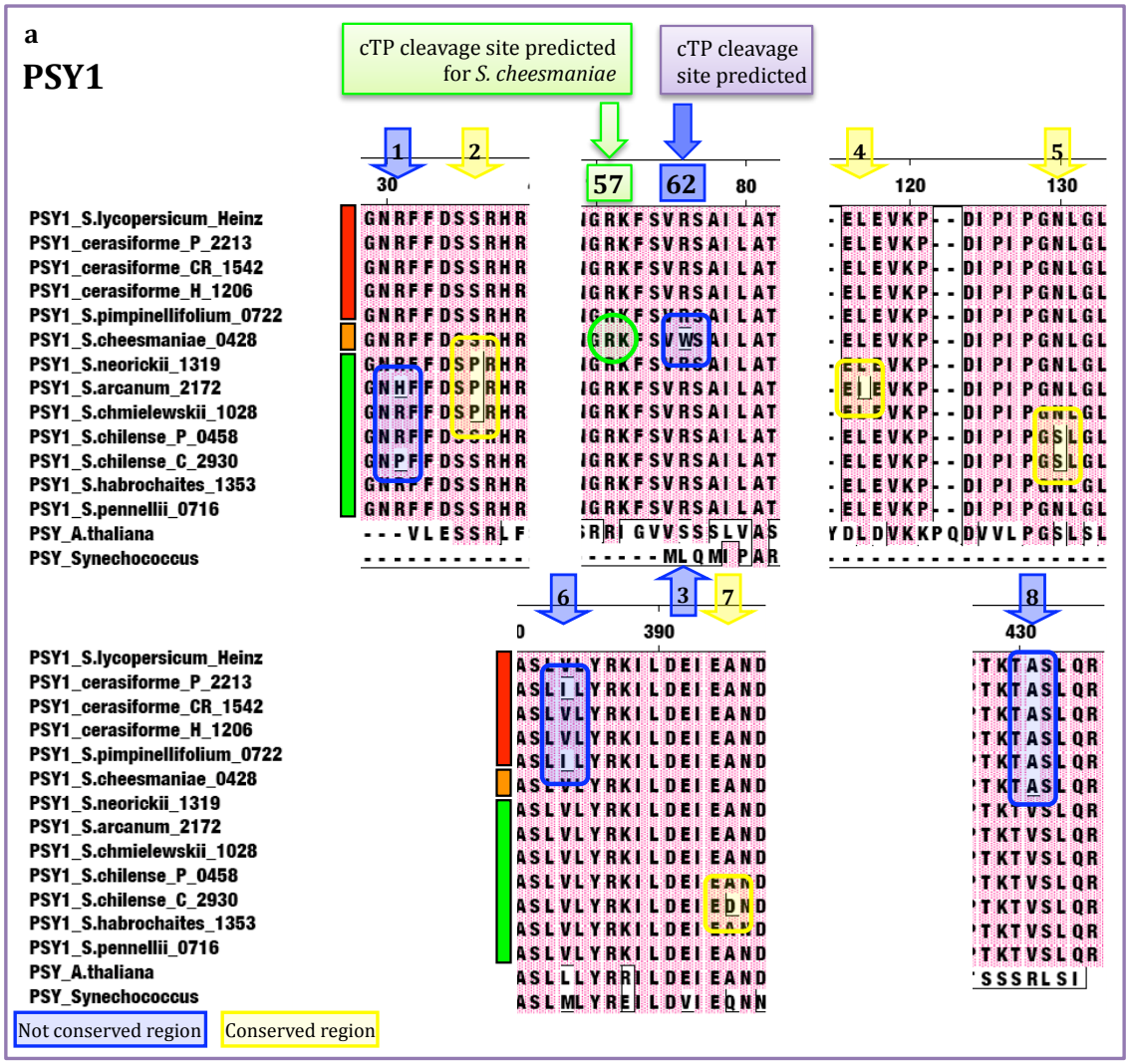


Fig. 30: panel a. Summary of amino acid substitutions found in PSY1 and its chloroplast transit peptide (cTP) cleavage site; panel. b: cTP length prediction according to the ChloroP software.

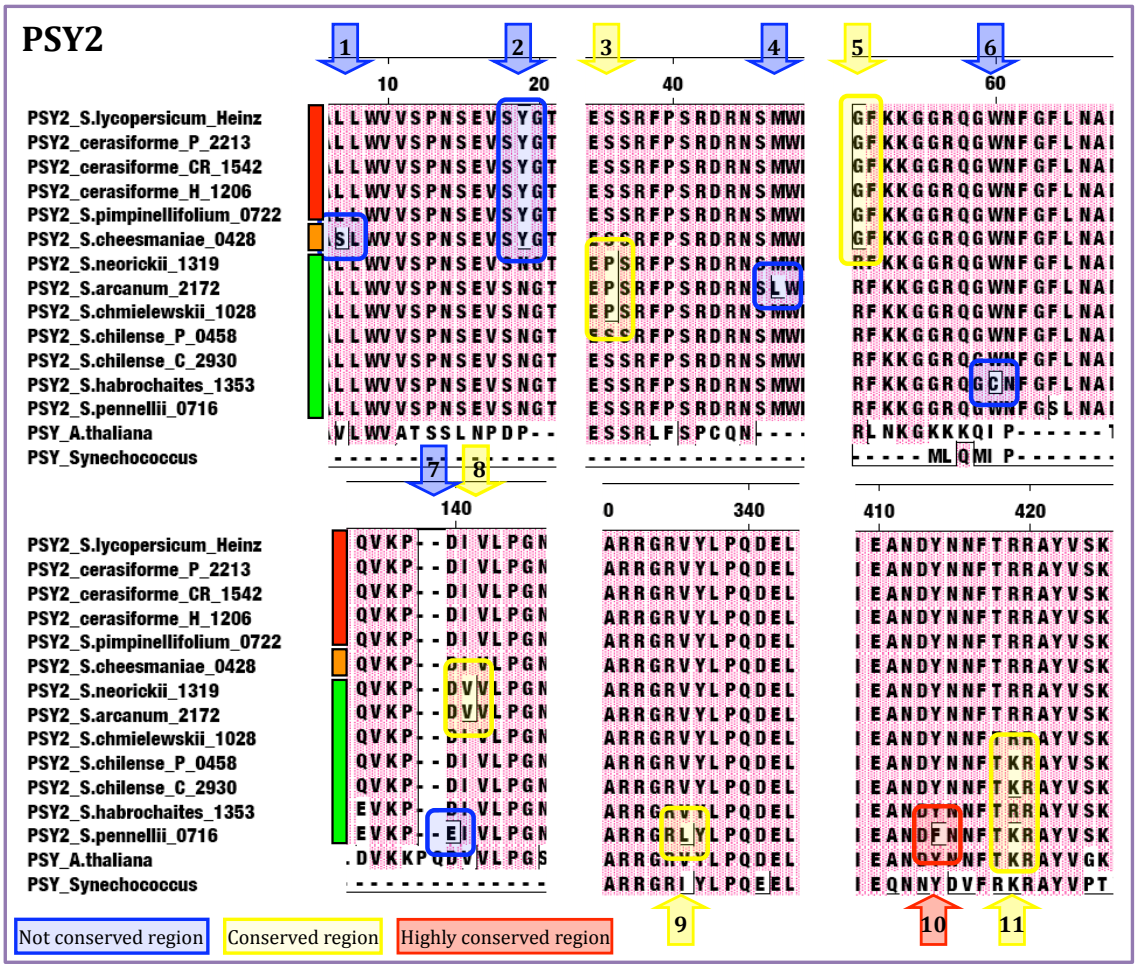


Fig. 31: Summary of amino acid substitutions found in PSY2.

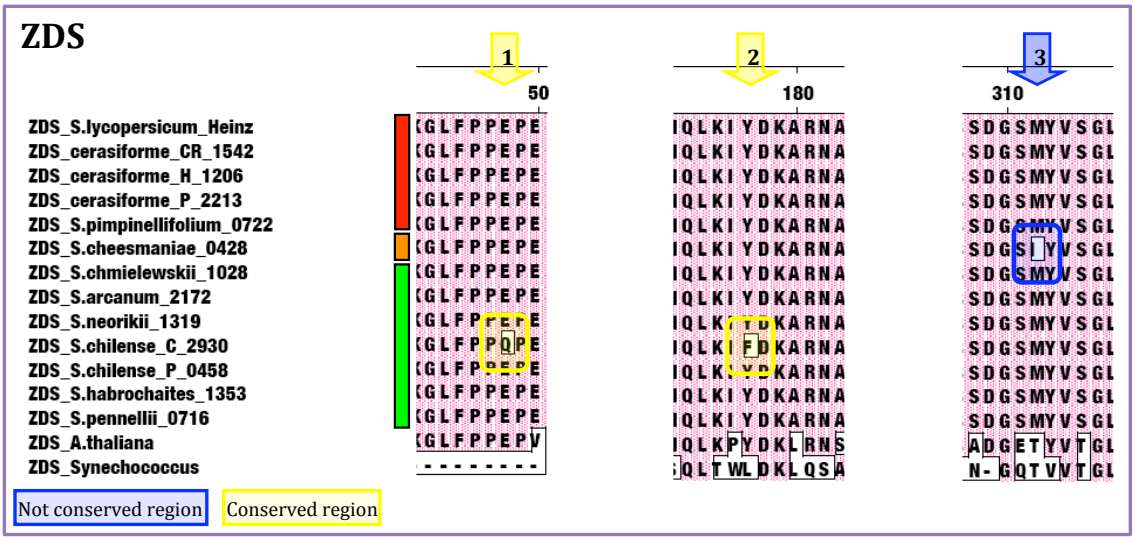


Fig. 32: Summary of amino acid substitutions found in ZDS.

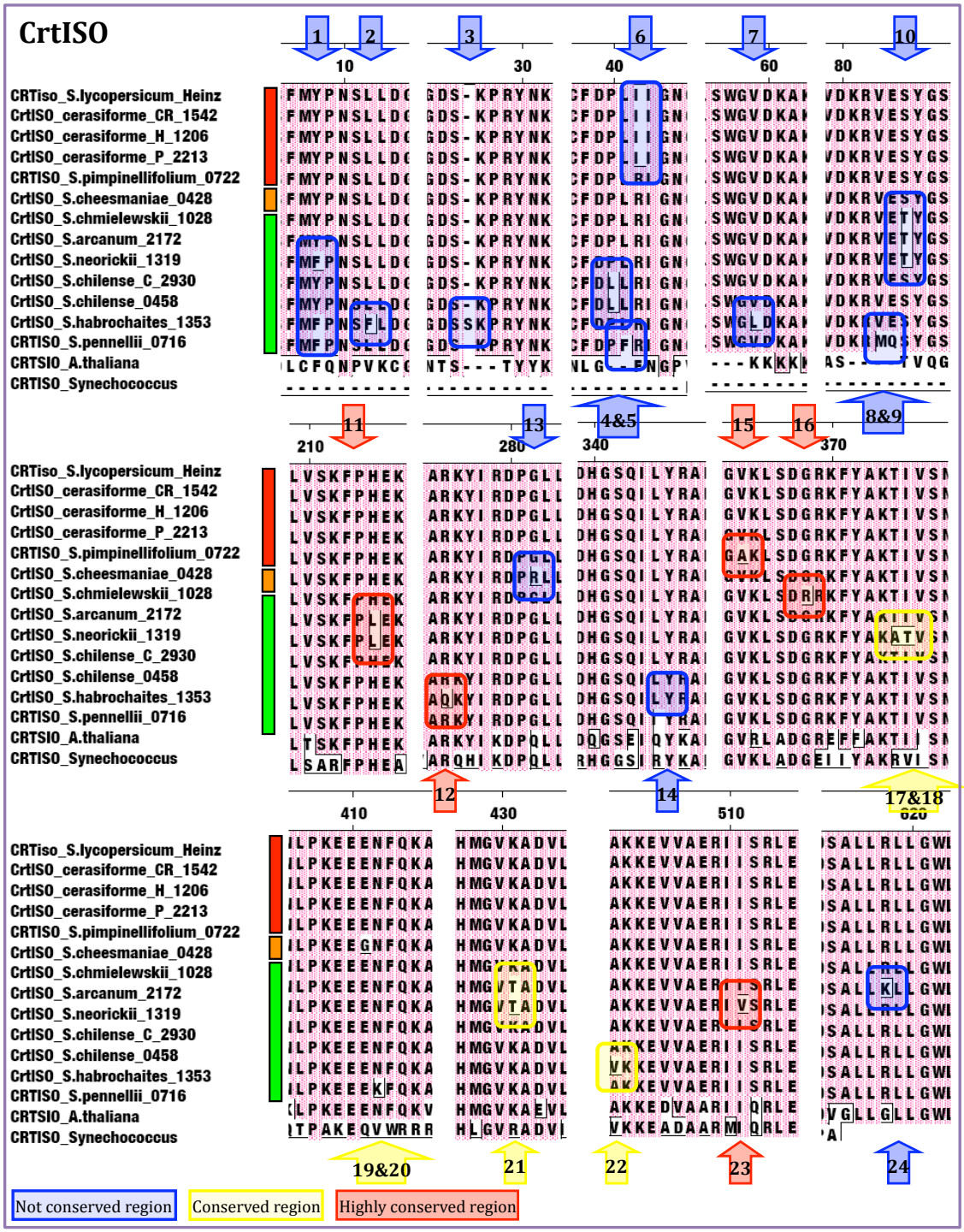


Fig. 33: Summary of amino acid substitutions found in CrtISO.

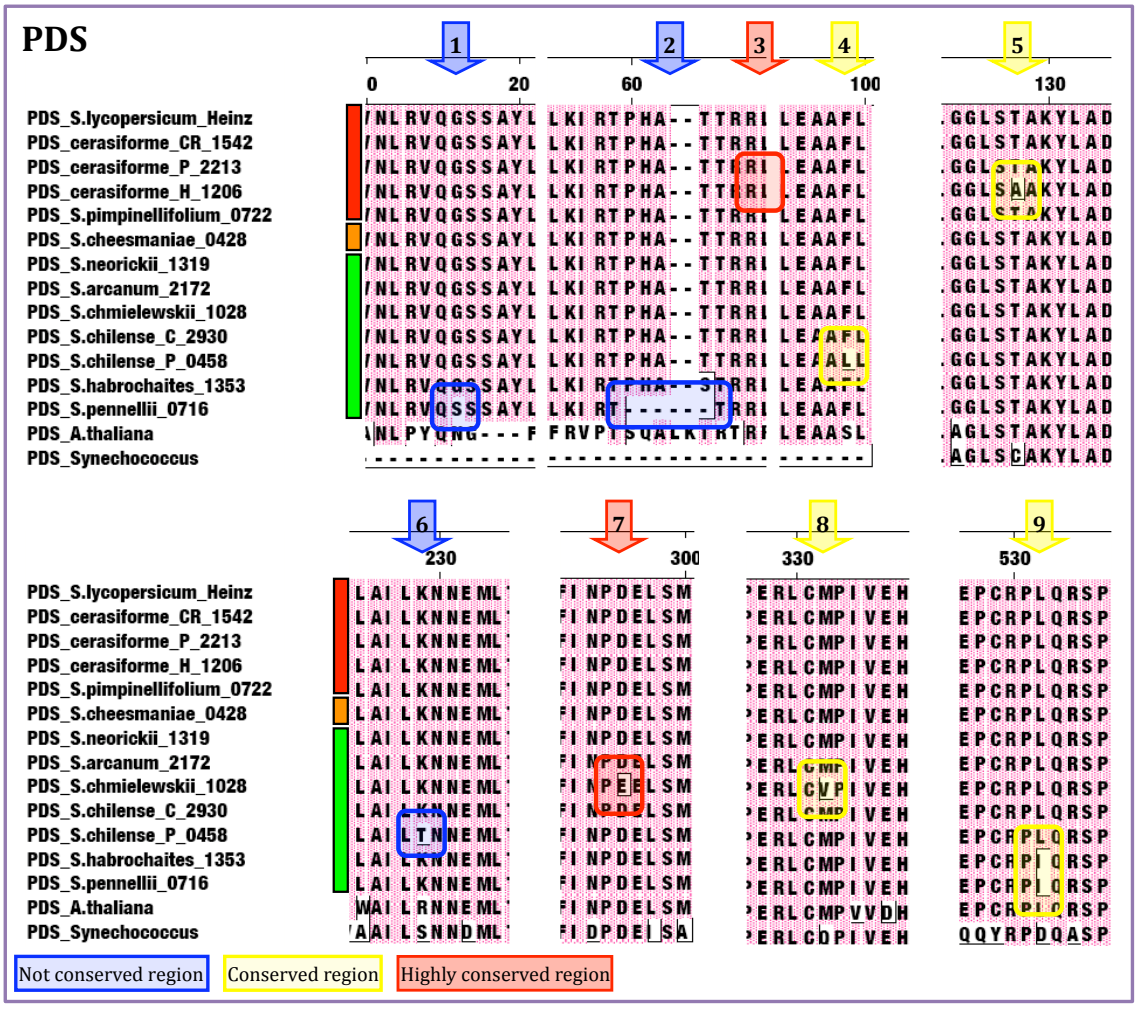


Fig. 34: Summary of amino acid substitutions found in PDS.

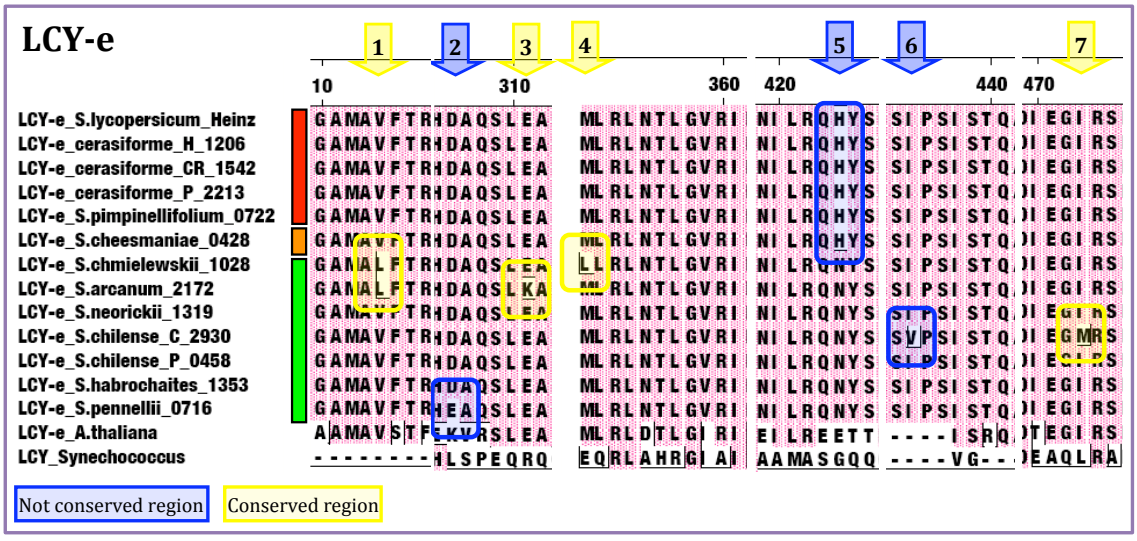


Fig. 35: Summary of amino acid substitutions found in LCY-e.

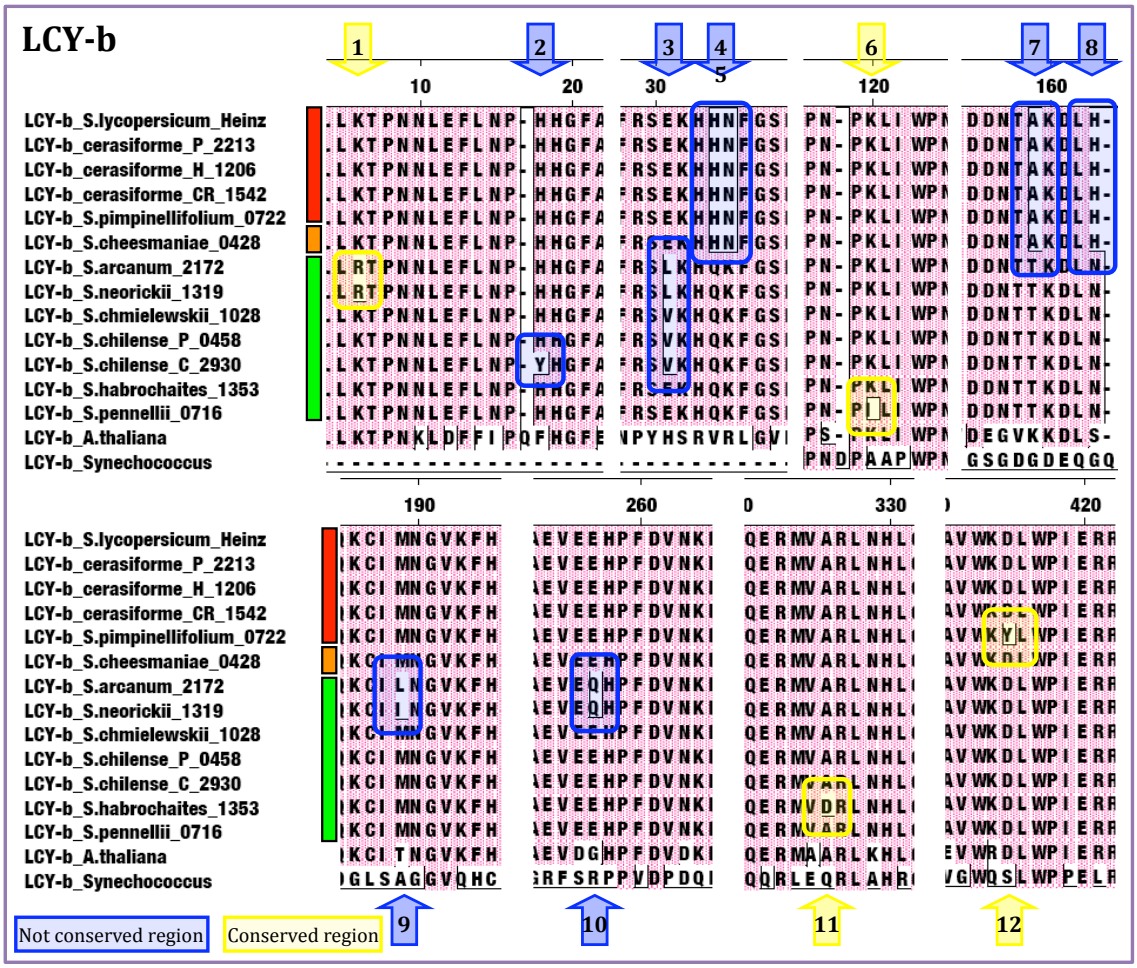


Fig. 36: Summary of amino acid substitutions found in LCY-b.

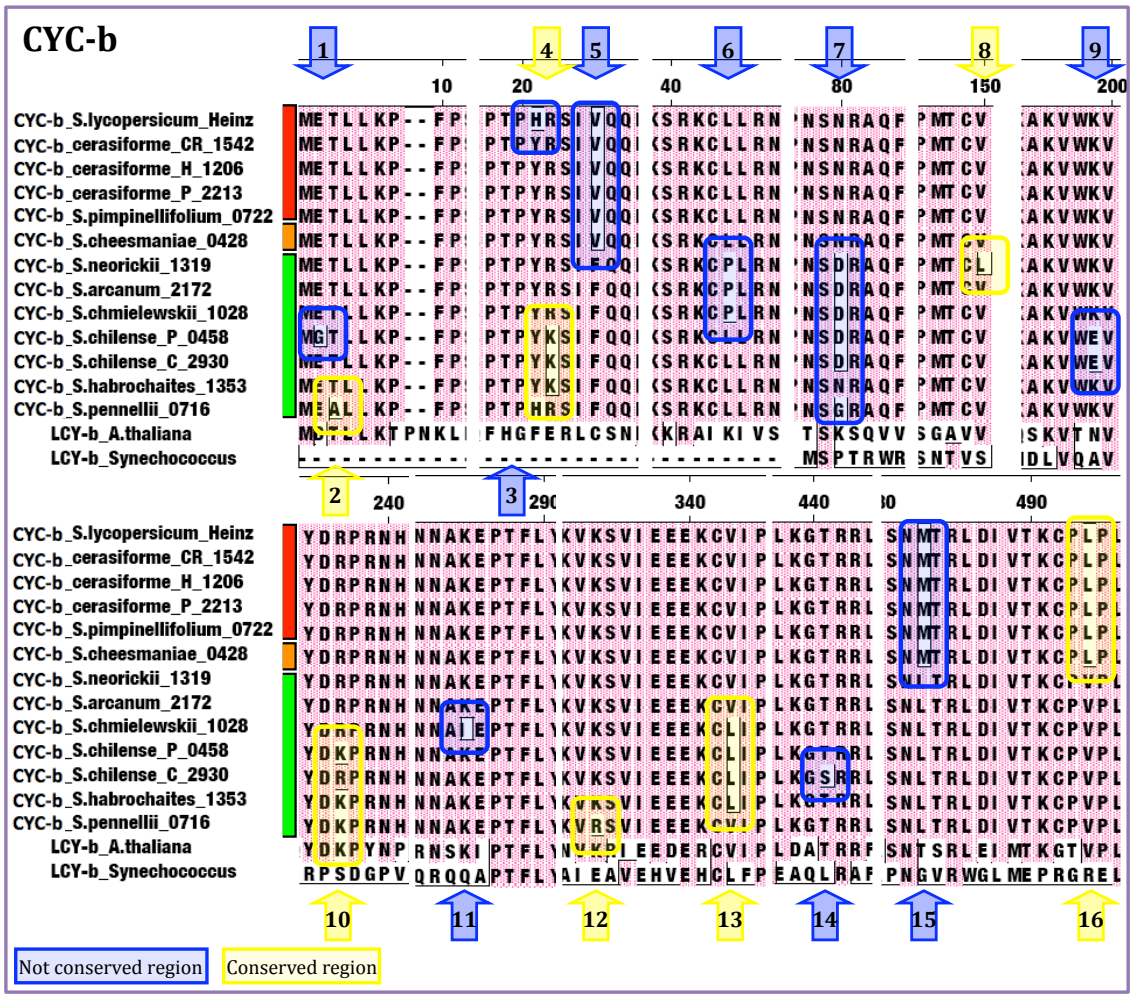


Fig. 37: Summary of amino acid substitutions found in CYC-b.

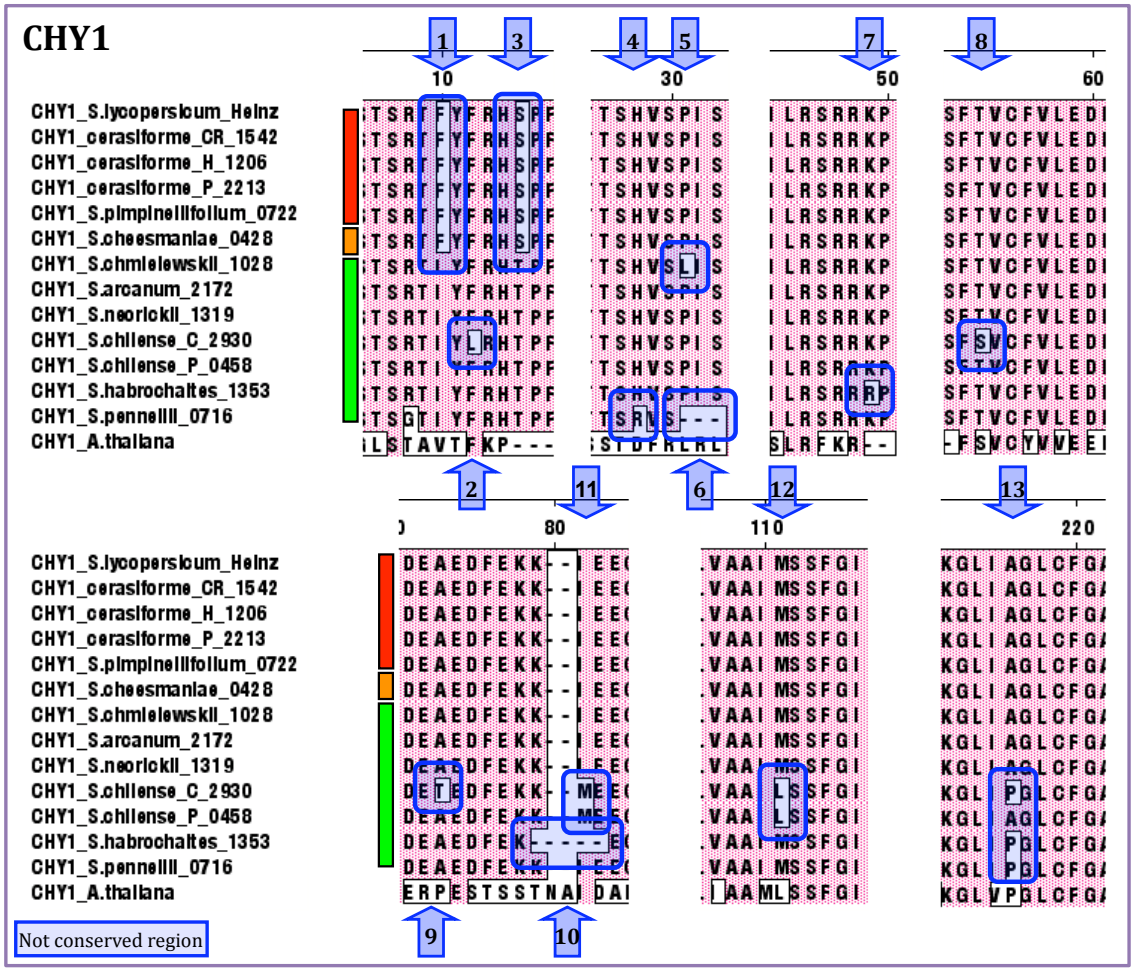


Fig. 38: Summary of amino acid substitutions found in CHY1.

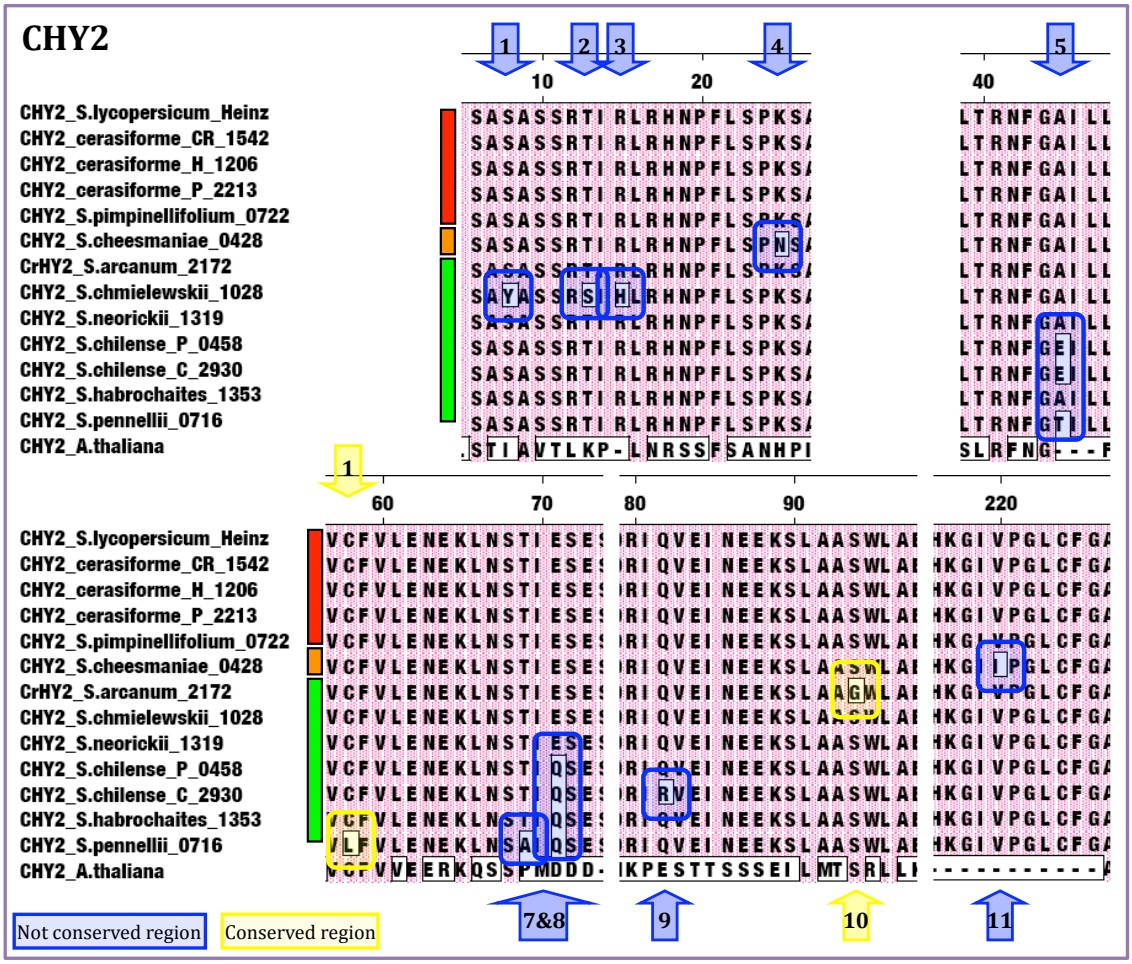


Fig. 39: Summary of amino acid substitutions found in CHY2.

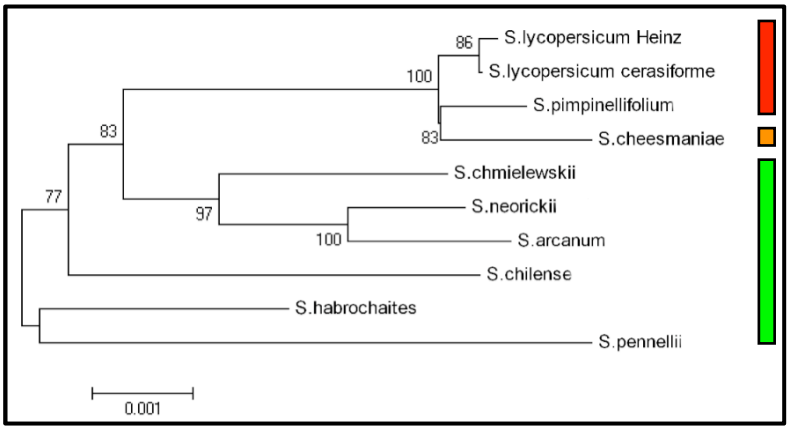


Fig. 40: Phylogenetic tree topology based on amino acid sequences: The dendogram is built on the basis of the sum of the deduced amino acid sequences of 10 genes belonging to the carotenoid biosynthetic pathway (*PSY1, PSY2, PDS, ZDS, CrtISO, LCY-b, CYC-b, LCY-e, CHY1* and *CHY2).* Software: MEGA4. Method Neighbor-Joining, model: Amino: p-distance. Numbers are bootstrap values for each branch (500 replications).

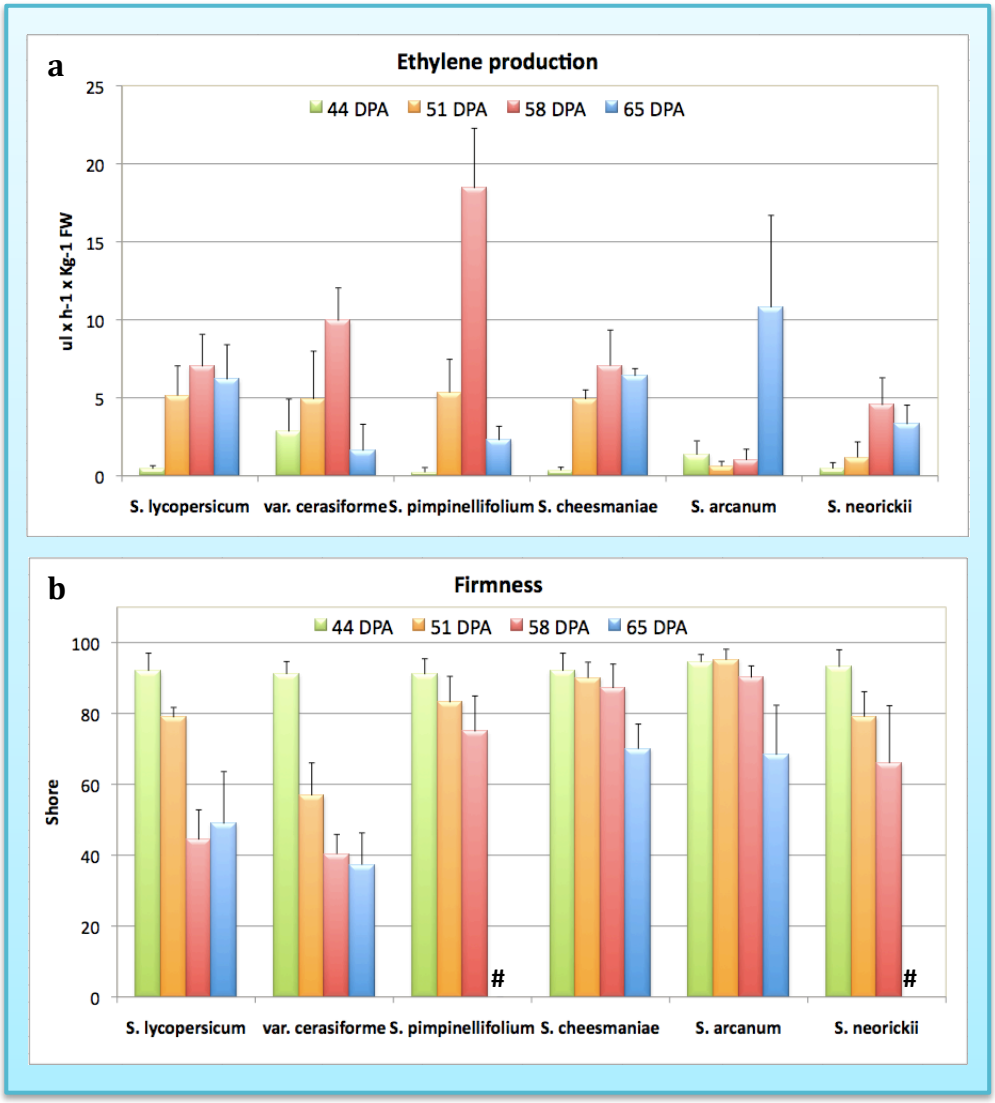


Fig. 41: Ethylene production in berry (panel a) and fruit firmness measured with durometer (panel b). DPA: days post-anthesis; # Berries of *S. pimpinellifolium* and *S. neorickii* at 65 dpa were too soft to be assayed.

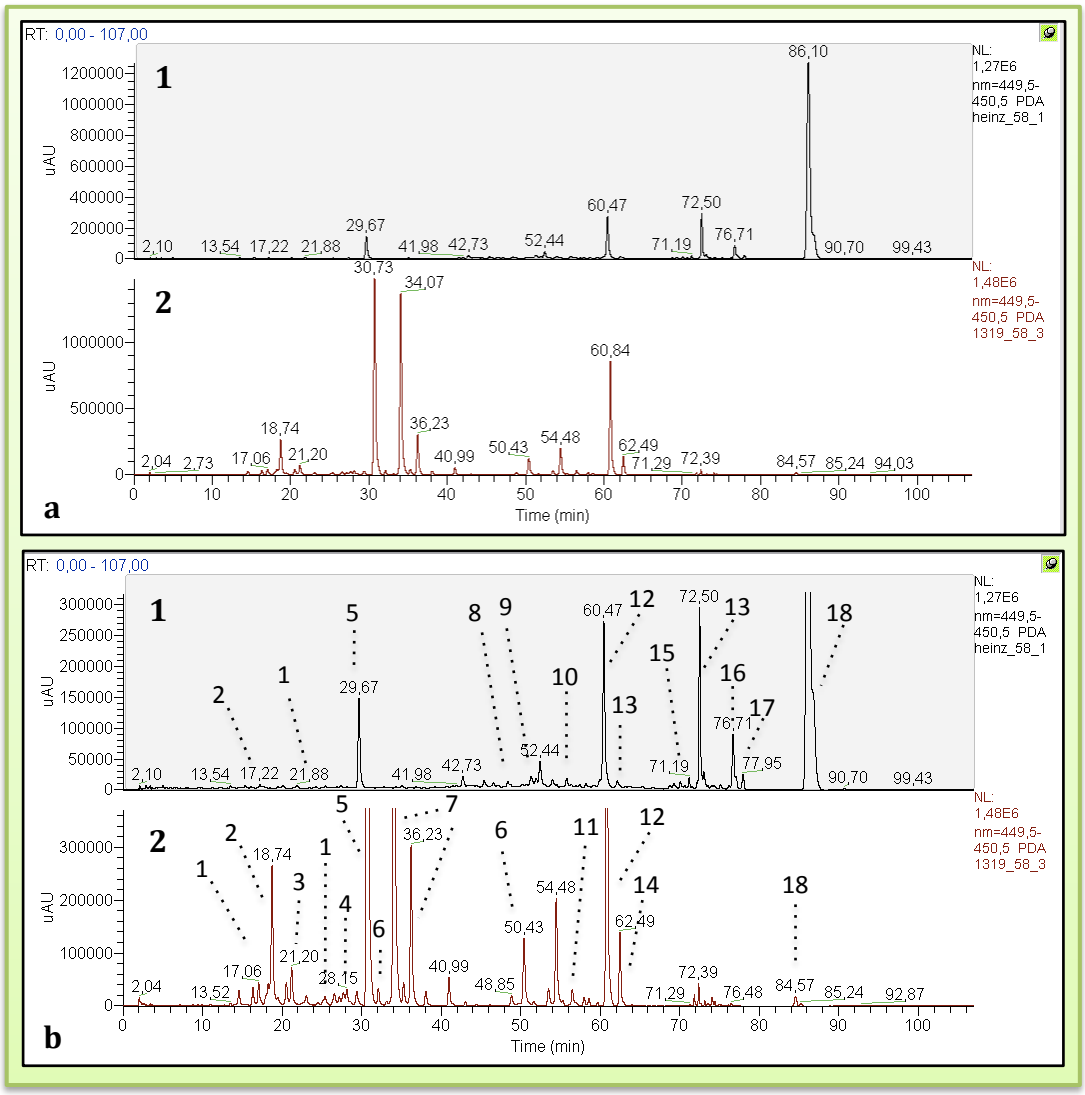


Fig. 42: Panel a. Profiles of carotenoids recorded at 450 nm present in a red fruited-species (1. *S. lycopersicum*) and in a green fruited-species (2. *S. neorickii*) at 58 DPA; panel b. enlargement of panel a.

1. Unknown xanthophyll (1), 2. Neoxanthin, 3. Unknown xanthophyll (2), 4. Violaxanthin, 5. Lutein, 6. Chlorophyll a, 7. Chlorophyll b, 8. Phytoene, 9. Phytofluene, 10. ζ-carotene, 11. α-carotene, 12. β-carotene, 13. Poli-*cis*-lycopene, 14. *Cis*-β-carotene, 15. δ-carotene, 16. *Cis*-lycopene-1, 17. *Cis*-lycopene-2, 18. All-*trans*-lycopene.

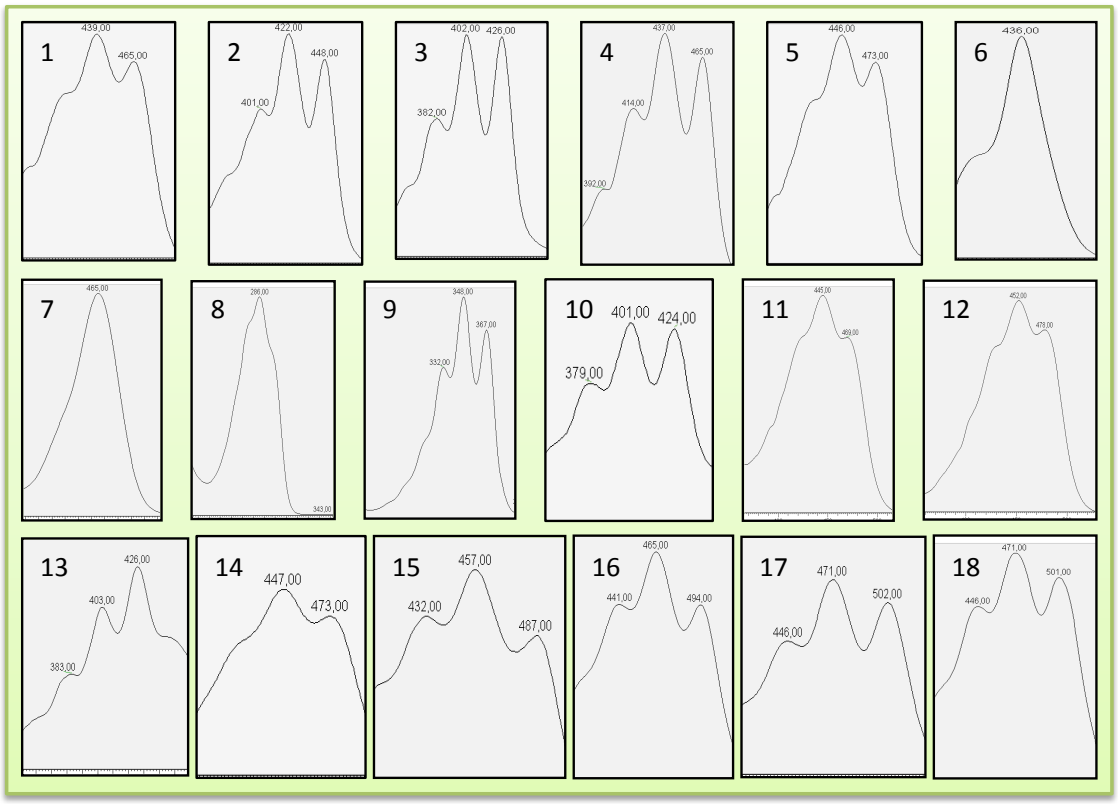


Fig. 43: Absorption spectra of carotenoids present in the samples: 1. Neoxanthin, 2. Unknown xanthophyll (1), 3. Unknown xanthophyll (2), 4. Violaxanthin, 5. Lutein, 6. Chlorophyll a, 7. Chlorophyll b, 8. Phytoene, 9. Phytofluene, 10. ζ-carotene, 11. α-carotene, 12. β-carotene, 13. Poli-*cis*-lycopene, 14. *Cis*-β-carotene, 15. δ-carotene, 16. *Cis*-lycopene-1, 17. *Cis*-lycopene-2, 18. All-*trans*-lycopene.

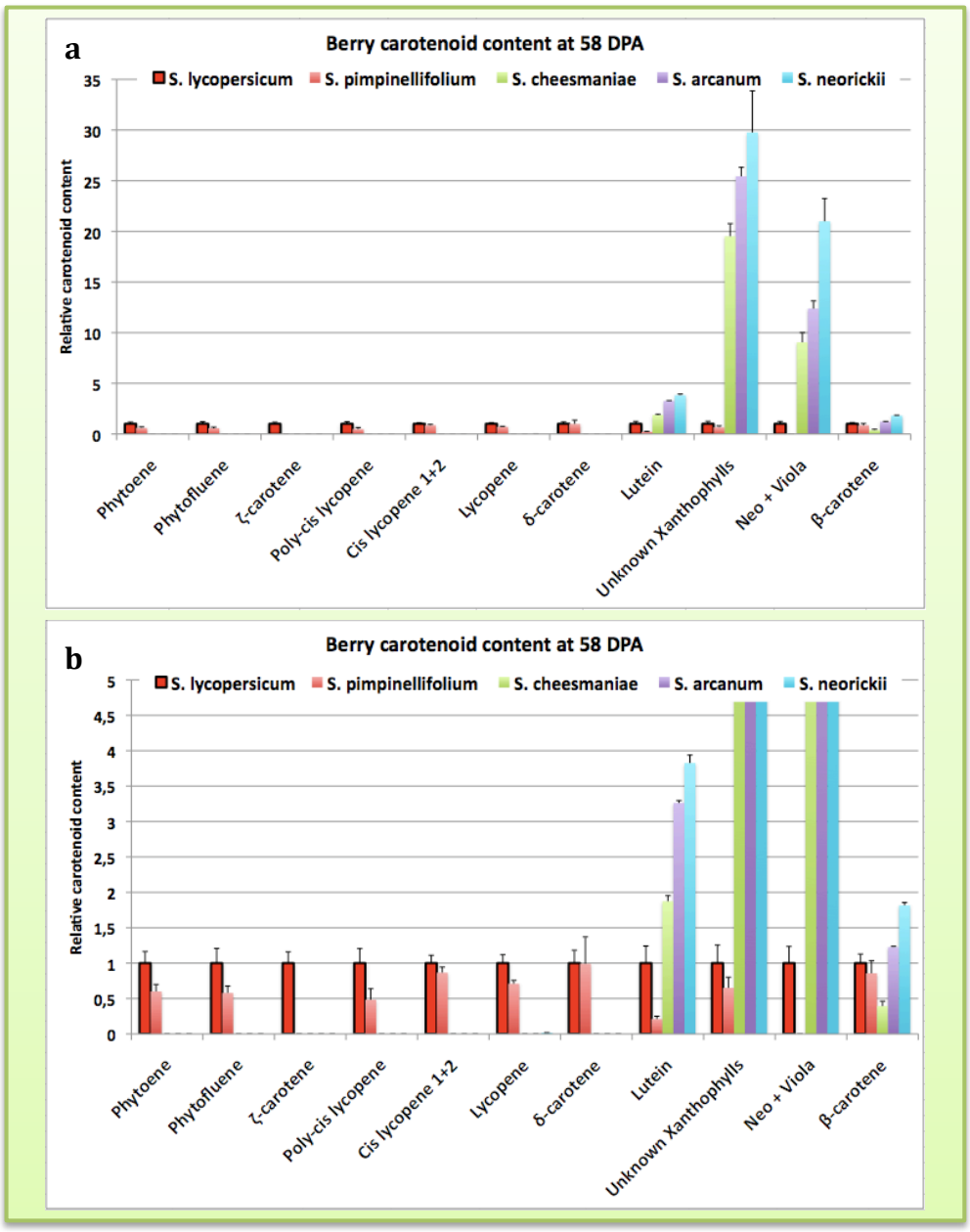


Fig. 44: HPLC analysis: panel a. Carotenoid composition in *S. lycopersicum* and its wild relatives in berry at 58 DPA. Fold variation with respect to cultivated tomato (*S. lycopersicum cv* Heinz) at 58 DPA is reported; b. enlargement of panel a.

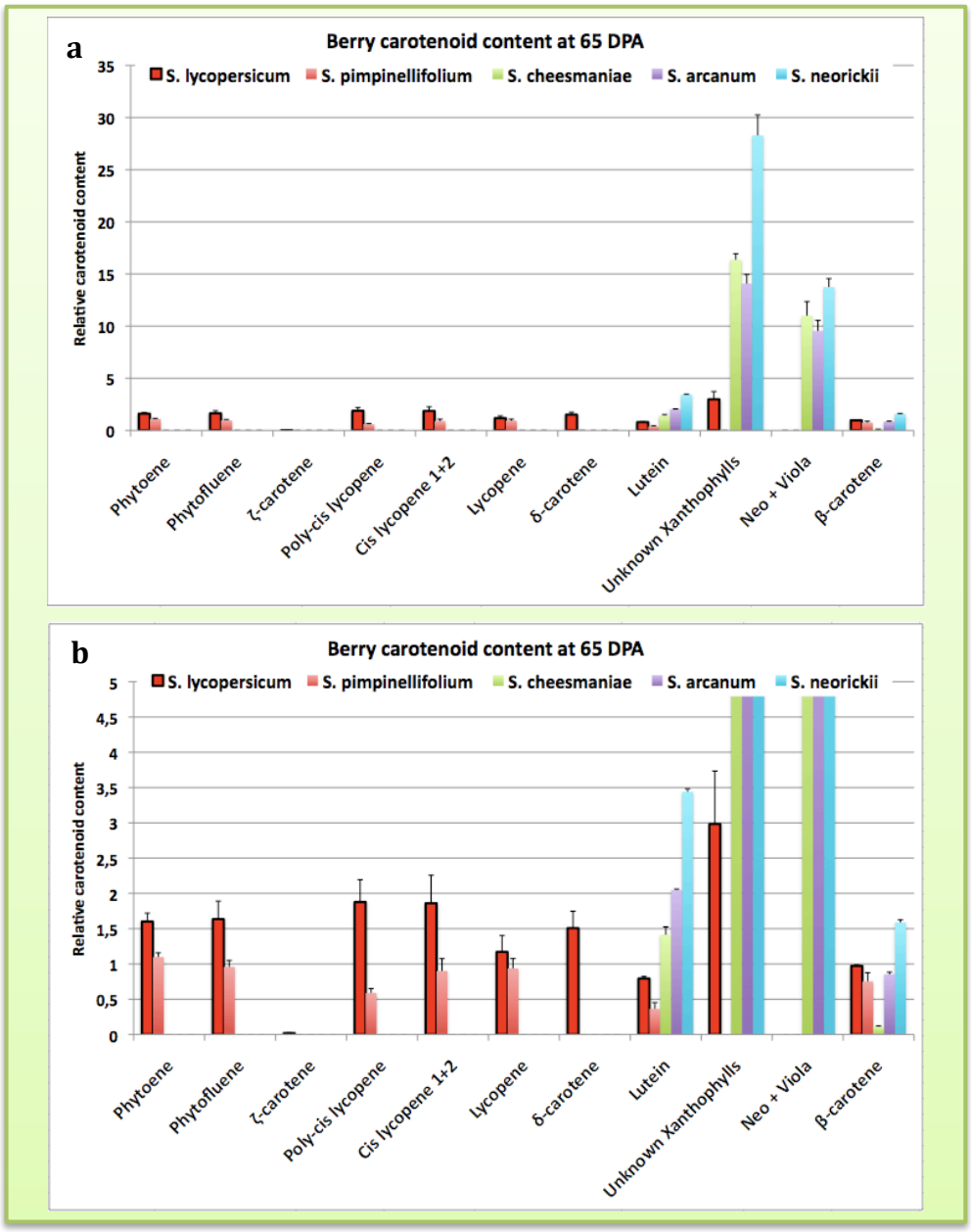


Fig. 45: HPLC analysis: panel a. Carotenoid composition in *S. lycopersicum* and its wild relatives in berry at 65 DPA. Fold variation with respect to cultivated tomato (*S. lycopersicum cv* Heinz) at 58 DPA is reported; panel b. enlargement of panel a.

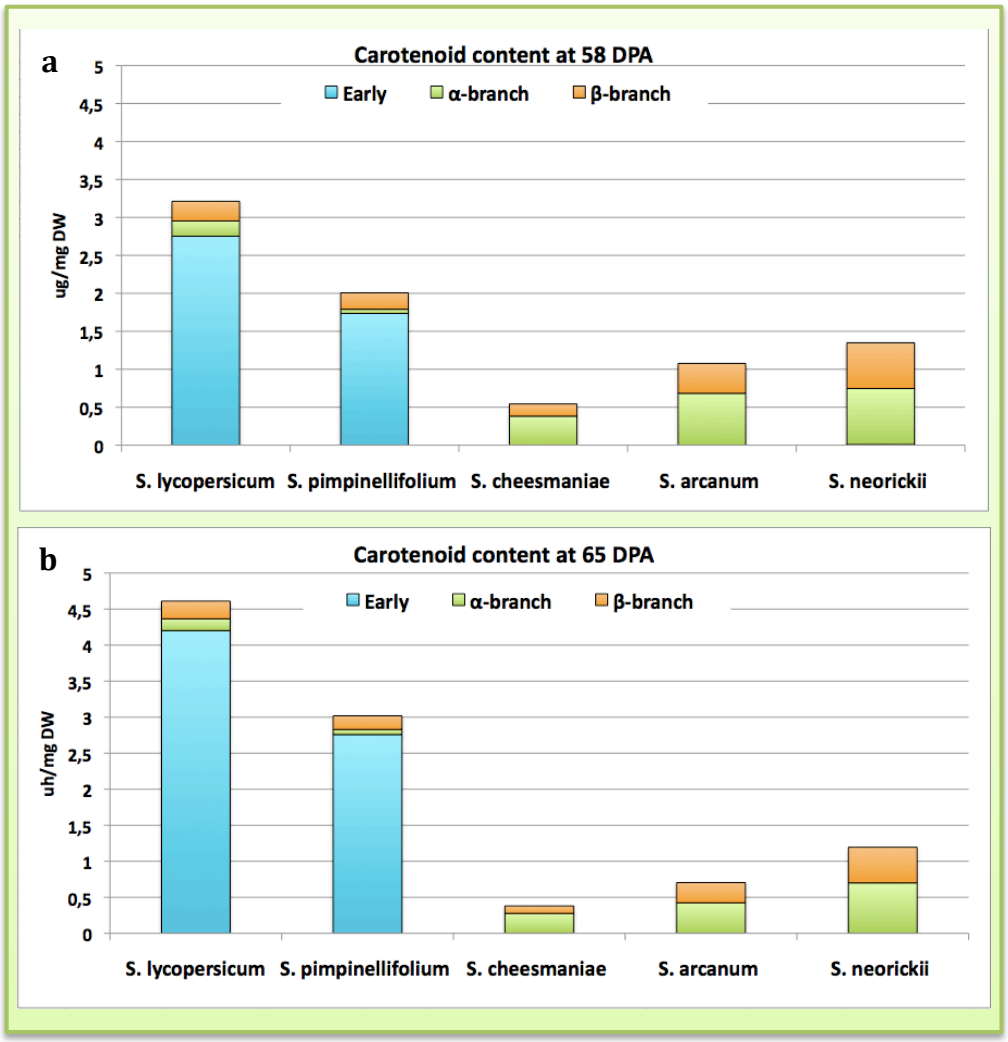


Fig. 46: HPLC analysis: comparison among specific groups of carotenoids in berry at 58 DPA (panel a) and at 65 DPA (panel b). The "early" group includes carotenoids from Geranylgeranyl pyrophosphate to lycopene, "β-branch", from β-carotene to neoxanthin and "α-branch", from δ-carotene to lutein.

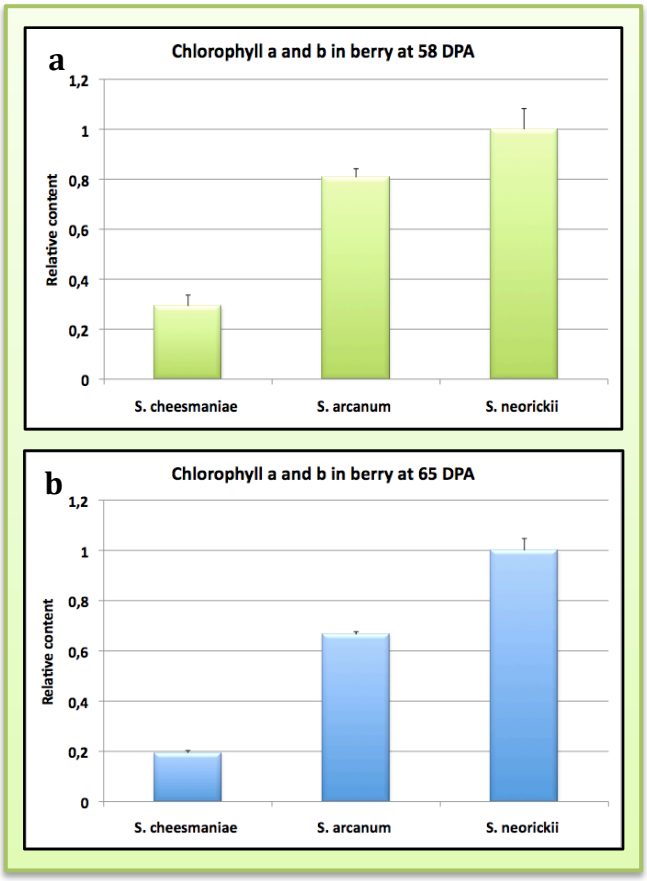


Fig. 47: HPLC analysis: Chlorophylls (a and b) composition in the orange-fruited species *S. cheesmaniae* and the green-fruited species *S. neorickii and S. arcanum* in berry at 58 (panel a) and at 65 DPA (panel b). The data were normalized with respect to *S. neorickii*.

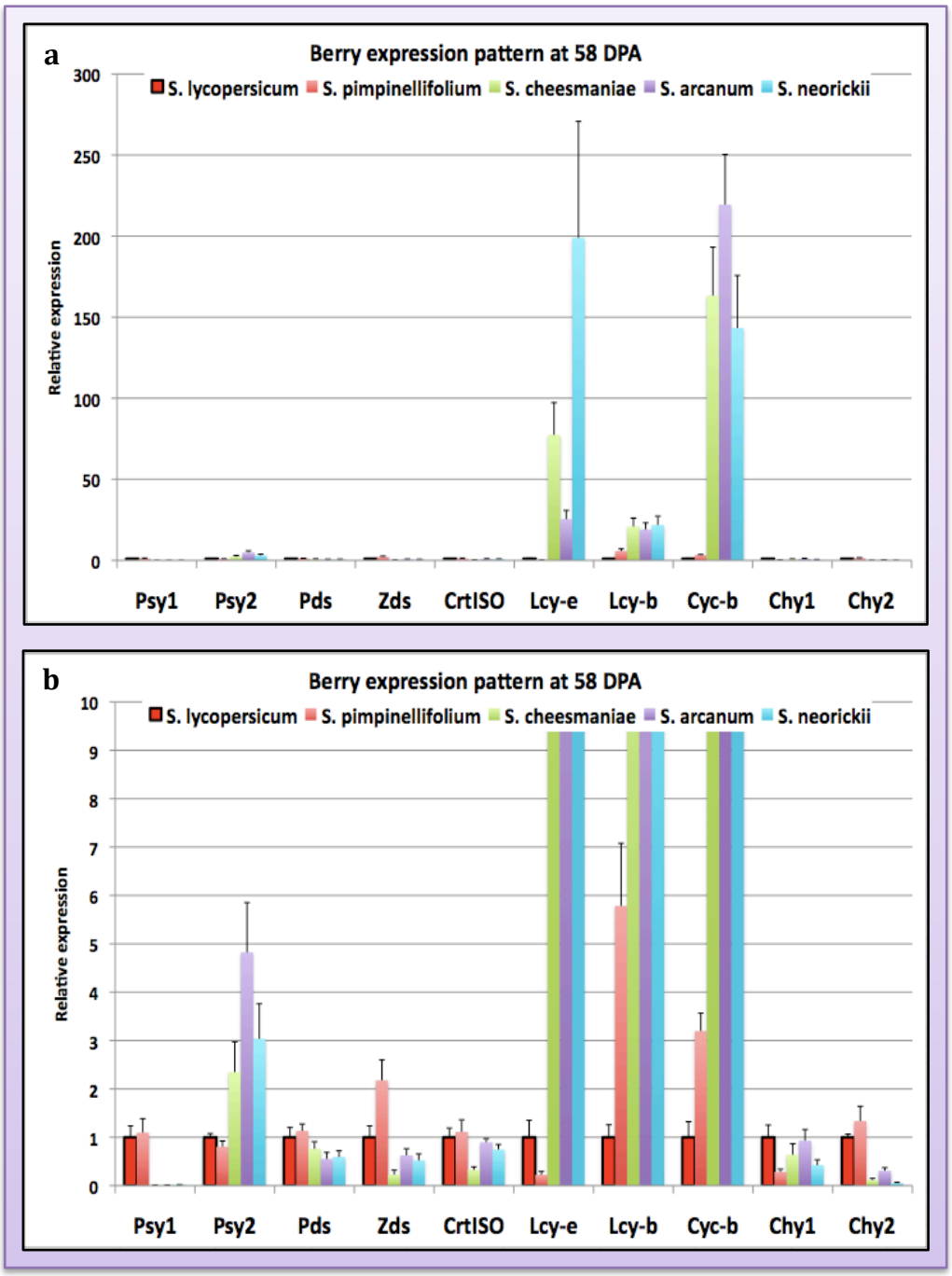


Fig. 48: Carotenoid gene expression in *S. lycopersicum* and its wild relatives in berry at 58 DPA. Transcript levels were measured via Real-Time PCR and were first normalized for expression of the housekeeping *Actin* gene, and then for the expression levels in the *S. lycopersicum*; panel b. enlargement of panel a.

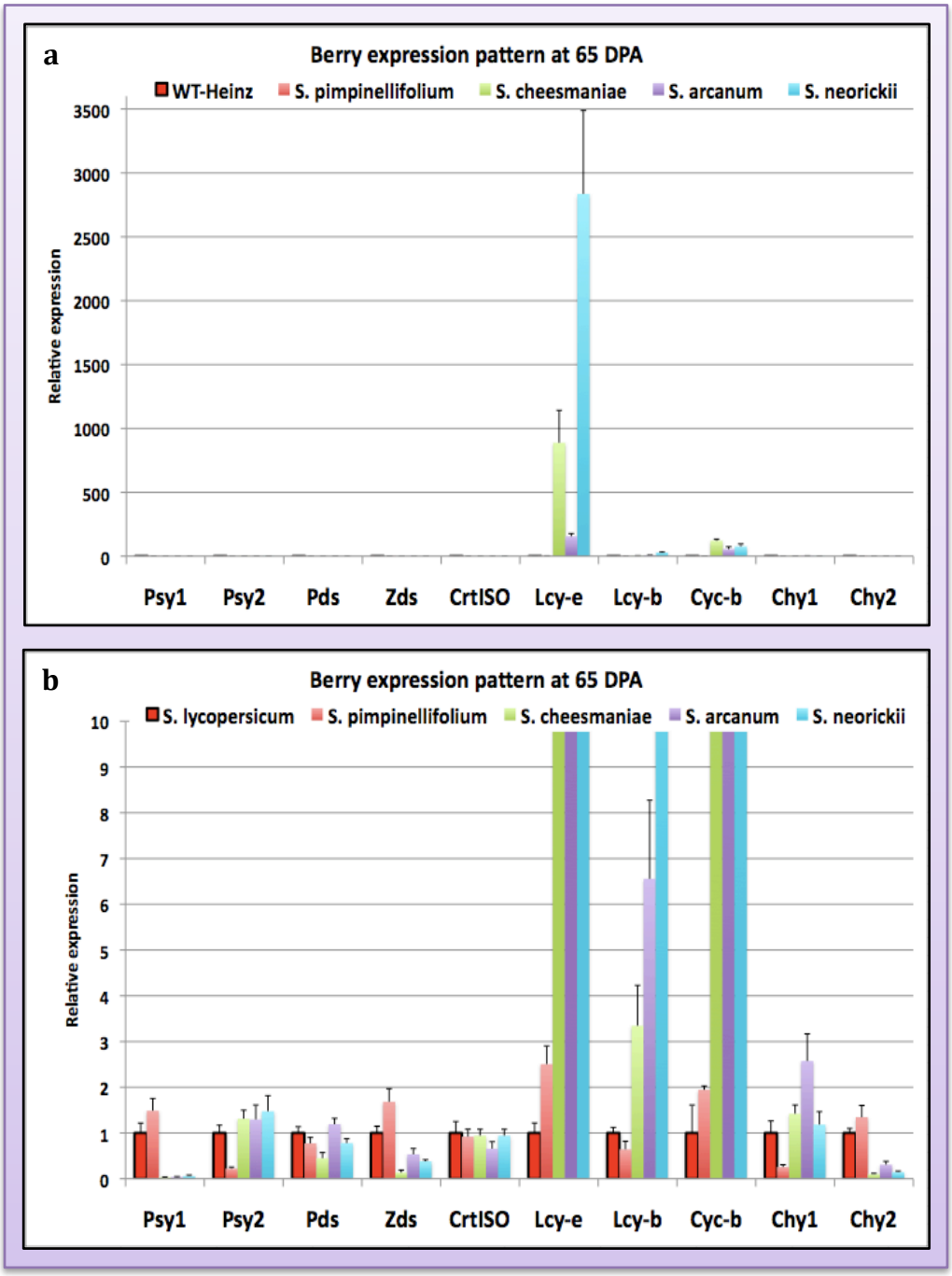


Fig. 49: Carotenoid gene expression profiling in *S. lycopersicum* and its wild relatives in berry at 65 DPA. Transcript levels were measured via Real-Time PCR and were first normalized for expression of the housekeeping *Actin* gene, and then for the expression levels in the *S. lycopersicum*; panel b. enlargement of panel a.

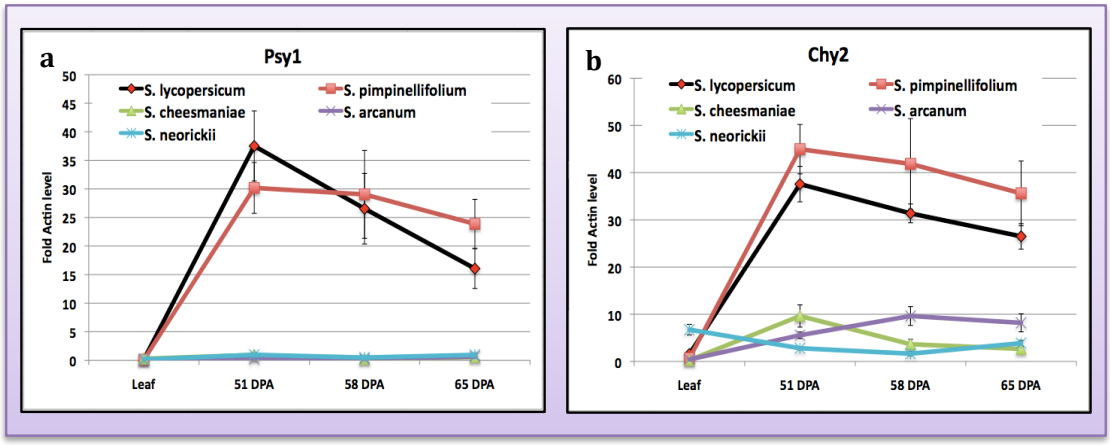


Fig. 50: Time course of the expression level of *PSY1* (panel a) and *CHY2* (panel b).

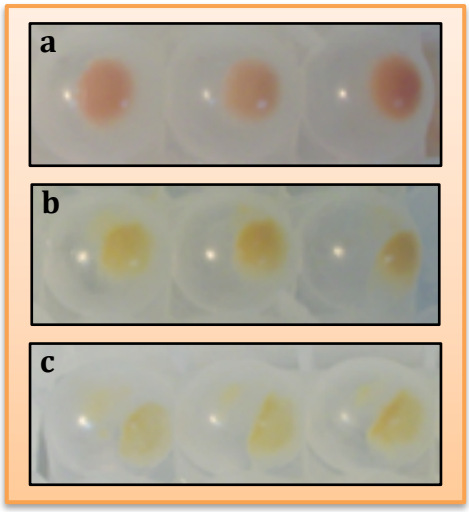


Fig. 51: Coloured pellet of *E. coli*-pFabeR cells transformed with pQE-50/LCY-e after the IPTG induction: Panel a. negative control of the transformation: the cells accumulate only lycopene (red colour); the transformed cells with the LCY-e of *S. arcanum* (panel b) and *S. lycopersicum* (panel c) show orange colour.

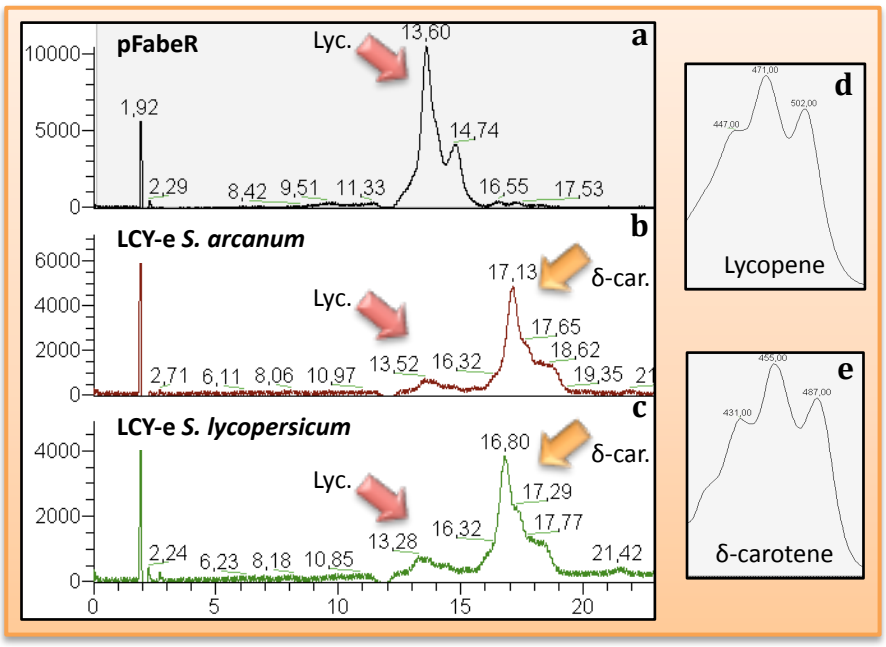


Fig. 52: HPLC chromatograms of carotenoids extracted from *E. coli* after induction with IPTG. Panel a. pFabeR: the cells accumulate only lycopene; panel b. and c: pFabeR + pQE-50/LCY-e (*S. arcanum*) and pFabeR + pQE-50/LCY-e (*S. lycopersicum*): both lycopene and δ-carotene are present; panel d. absorption spectrum of lycopene; panel e. absorption spectrum of δ-carotene.

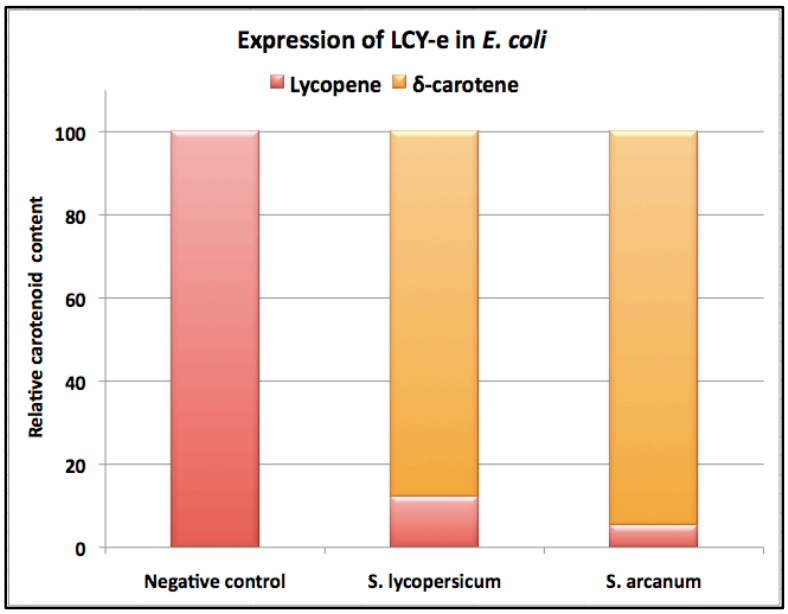


Fig. 53: HPLC analysis of extracted carotenoid after the induction from *E. coli*-pFabeR: the cells was transformed with pQE-50/LCY-e plasmid of *S. lycopersicum* and *S. arcanum*. Negative control: *E. coli*-pFabeR without pQE-50/LCY-e.

9. References

1. Emmanuel, E. and A.A. Levy, *Tomato mutants as tools for functional genomics.* Curr Opin Plant Biol, 2002. **5**(2): p. 112-7.
2. Giovannoni, J.J., *Fruit ripening mutants yield insights into ripening control.* Curr Opin Plant Biol, 2007. **10**(3): p. 283-9.
3. Tanksley, S.D., et al., *High density molecular linkage maps of the tomato and potato genomes.* Genetics, 1992. **132**(4): p. 1141-60.
4. Tanksley, S.D., *Mapping polygenes.* Annu Rev Genet, 1993. **27**: p. 205-33.
5. Van der Hoeven, R., et al., *Deductions about the number, organization, and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing.* Plant Cell, 2002. **14**(7): p. 1441-56.
6. An, G., B.D. Watson, and C.C. Chiang, *Transformation of Tobacco, Tomato, Potato, and Arabidopsis thaliana Using a Binary Ti Vector System.* Plant Physiol, 1986. **81**(1): p. 301-305.
7. Alba, R., et al., *Transcriptome and selected metabolite analyses reveal multiple points of ethylene control during tomato fruit development.* Plant Cell, 2005. **17**(11): p. 2954-65.
8. Moore, S., et al., *Use of genomics tools to isolate key ripening genes and analyse fruit maturation in tomato.* J Exp Bot, 2002. **53**(377): p. 2023-30.
9. Mueller, L.A., René Klein Lankhorst, Steven D. Tanksley, James J. Giovannoni, Ruth White, Julia Vrebalov, Zhangjun Fei, Joyce van Eck, Robert Buels, Adri

A. Mills, Naama Menda, Isaak Y. Tecle, Aureliano Bombarely, Stephen Stack, Suzanne M. Royer, Song-Bin Chang, Lindsay A. Shearer, Byung Dong Kim, Sung-Hwan Jo, Cheol-Goo Hur, Doil Choi, Chang-Bao Li, Jiuhai Zhao, Hongling Jiang, Yu Geng, Yuanyuan Dai, Huajie Fan, Jinfeng Chen, Fei Lu, Jinfeng Shi, Shouhong Sun, Jianjun Chen, Xiaohua Yang, Chen Lu, Mingsheng Chen, Zhukuan Cheng, Chuanyou Li, Hongqing Ling, Yongbiao Xue, Ying Wang, Graham B. Seymour, Gerard J. Bishop, Glenn Bryan, Jane Rogers, Sarah Sims, Sarah Butcher, Daniel Buchan, James Abbott, Helen Beasley, Christine Nicholson, Clare Riddle, Sean Humphray, Karen McLaren, Saloni Mathur, Shailendra Vyas, Amolkumar U. Solanke, Rahul Kumar, Vikrant Gupta, Arun K. Sharma, Paramjit Khurana, Jitendra P. Khurana, Akhilesh Tyagi, Sarita, Parul Chowdhury, Smriti Shridhar, Debasis Chattopadhyay, Awadhesh Pandit, Pradeep Singh, Ajay Kumar, Rekha Dixit, Archana Singh, Sumera Praveen, Vivek Dalal, Mahavir Yadav, Irfan Ahmad Ghazi, Kishor Gaikwad, Tilak Raj Sharma, Trilochan Mohapatra, Nagendra Kumar Singh, Dora Szinay, Hans de Jong, Sander Peters, Marjo van Staveren, Erwin Datema, Mark W.E.J. Fiers, Roeland C.H.J. van Ham, P. Lindhout, Murielle Philippot, Pierre Frasse, Farid Regad, Mohamed Zouine, Mondher Bouzayen, Erika Asamizu, Shusei Sato, Hiroyuki Fukuoka, Satoshi Tabata, Daisuke Shibata, Miguel A. Botella, M. Perez-Alonso, V. Fernandez-Pedrosa, Sonia Osorio, Amparo Mico, Antonio Granell, Zhonghua Zhang, Jun He, Sanwen Huang, Yongchen Du, Dongyu Qu, Longfei Liu, Dongyuan Liu, Jun Wang, Zhibiao Ye, Wencai Yang, Guoping Wang, Alessandro Vezzi, Sara Todesco,

Giorgio Valle, Giulia Falcone, Marco Pietrella, Giovanni Giuliano, Silvana Grandillo, Alessandra Traini, Nunzio D'Agostino, Maria Luisa Chiusano, Mara Ercolano, Amalia Barone, Luigi Frusciante, Heiko Schoof, Anika Jocker, Rémy Bruggmann, Manuel Spannagl, Klaus X.F. Mayer, Roderic Guigò, Francisco Camara, Stephane Rombauts, Jeffrey A. Fawcett, Yves Van de Peer, Sandra Knapp, Dani Zamir, and Willem Stiekema, *A Snapshot of the Emerging Tomato Genome Sequence.* The Plant Genome, 2009. **2**(1): p. 78Ð92.

10. Chetelat, R.T., *Revised list of miscellaneous stocks.* Tomato Genetics Cooperative Reports, 2005. **55**: p. 48Ð69.
11. Atherton, J.G. and J. Rudich, *The tomato crop.* Springer editor, 1986: p. 121-125.
12. De Jong, W.S., et al., *Candidate gene analysis of anthocyanin pigmentation loci in the Solanaceae.* Theor Appl Genet., 2004. **108**(3): p. 423-32.
13. Ho, L.C. and J.D. Hewitt, *Fruit development. In 'The Tomato Crop: a Scientific Basis for Improvement'.* (Eds J. C.Atherton and J. Rudich.) (Chapman and Hall: London.), 1986: p. 202-206.
14. Jones, J.B., *Tomato Plant Culture.* CRC Press, LLC., Boca Raton, FL., 1999: p. 1-3.
15. Jones, B.J., *Tomato Plant Culture In the Field, Greenhouse, and Home Garden.* CRC Edition, 1998.
16. Gillaspy, G., H. Ben-David, and W. Gruissem, *Fruits: A Developmental Perspective.* Plant Cell, 1993. **5**(10): p. 1439-1451.

17. Giovannoni, J.J., *Genetic regulation of fruit development and ripening.* Plant Cell, 2004. **16 Suppl**: p. S170-80.

18. Tanksley, S.D., *The genetic, developmental, and molecular bases of fruit size and shape variation in tomato.* Plant Cell, 2004. **16 Suppl**: p. S181-9.

19. Gorguet, B., A.W. van Heusden, and P. Lindhout, *Parthenocarpic fruit development in tomato.* Plant Biol (Stuttg), 2005. **7**(2): p. 131-9.

20. Yang, S.F., *Biosynthesis and action of ethylene.* HortScience, 1985. **20**: p. 41-45.

21. Solano, R. and J.R. Ecker, *Ethylene gas: perception, signaling and response.* Curr Opin Plant Biol, 1998. **1**(5): p. 393-8.

22. Alexander, L. and D. Grierson, *Ethylene biosynthesis and action in tomato: a model for climacteric fruit ripening.* J Exp Bot, 2002. **53**(377): p. 2039-55.

23. Giovannoni, J., *Molecular Biology of Fruit Maturation and Ripening.* Annu Rev Plant Physiol Plant Mol Biol, 2001. **52**: p. 725-749.

24. Bartley, G.E. and B.K. Ishida, *Ethylene-sensitive and insensitive regulation of transcription factor expression during in vitro tomato sepal ripening.* J Exp Bot, 2007. **58**(8): p. 2043-51.

25. Solano, R., et al., *Nuclear events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSEFACTOR1.* Genes and Development, 1998. **12**: p. 3703-3714.

26. Wilkinson, J.Q., et al., *An ethylene-inducible component of signal transduction encoded by never-ripe.* Science, 1995. **270**(5243): p. 1807-9.

27. Vrebalov, J., et al., *A MADS-box gene necessary for fruit ripening at the tomato*

ripening-inhibitor (rin) locus. Science, 2002. **296**(5566): p. 343-6.

28. Giovannoni, J.J., et al., *Molecular genetic analysis of the ripening-inhibitor and non-ripening loci of tomato: a first step in genetic map-based cloning of fruit ripening genes.* Mol Gen Genet, 1995. **248**(2): p. 195-206.
29. Barry, C.S. and J.J. Giovannoni, *Ripening in the tomato Green-ripe mutant is inhibited by ectopic expression of a protein that disrupts ethylene signaling.* Proc Natl Acad Sci U S A, 2006. **103**(20): p. 7923-8.
30. Barry, C.S., et al., *Ethylene insensitivity conferred by the Green-ripe and Never-ripe 2 ripening mutants of tomato.* Plant Physiol, 2005. **138**(1): p. 267-75.
31. Liu, Y., et al., *Manipulation of light signal transduction as a means of modifying fruit nutritional quality in tomato.* Proc Natl Acad Sci USA, 2004. **101**: p. 9897-902.
32. Lenucci, M.S., et al., *Antioxidant composition in cherry and high-pigment tomato cultivars.* J Agric Food Chem, 2006. **54**(7): p. 2606-13.
33. Guil-Guerrero, J.L. and M.M. Rebolloso-Fuentes, *Nutrient composition and antioxidant activity of eight tomato (Lycopersicon esculentum) varieties.* Journal of Food Composition and Analysis, 2009. **22**: p. 123Ð129.
34. Azari, R., et al., *Light signaling genes and their manipulation towards modulation of phytonutrient content in tomato fruits .* Biotechnology Advances, 2010. **28**(1): p. 108-118.
35. Beecher, G.R., *Nutrient content of tomatoes and tomato products.* Proc Soc Exp Biol Med, 1998. **218**(2): p. 98-100.

36. Di Mascio, P., S. Kaiser, and H. Sies, *Lycopene as the most efficient biological carotenoid singlet oxygen quencher.* Arch Biochem Biophys, 1989. **274**(2): p. 532-8.

37. Giovannucci, E., *Tomatoes, tomato-based products, lycopene, and cancer: review of the epidemiologic literature.* J. Natl. Cancer Inst., 1999. **91**: p. 317-331.

38. Agarwal, S. and A.V. Rao, *Tomato lycopene and its role in human health and chronic diseases.* Can. Med. Assoc. J., 2000. **163**: p. 739-744.

39. Wikstršm, N., V. Savolainen, and M.W. Chase, *Evolution of the angiosperms: calibrating the family tree.* Proc. R. Soc. Lond. B, 2001. **268**: p. 2211-2220.

40. Nesbitt, T.C. and S.D. Tanksley, *Comparative sequencing in the genus Lycopersicon. Implications for the evolution of fruit size in the domestication of cultivated tomatoes.* Genetics, 2002. **162**(1): p. 365-79.

41. Rick, C.M., *The tomato.* Science, 1978. **239**: p. 66-76.

42. Rick, C.M., J.W. Uhlig, and A.D. Jones, *High alpha-tomatine content in ripe fruit of Andean Lycopersicon esculentum var. cerasiforme: developmental and genetic aspects.* Proc Natl Acad Sci U S A, 1994. **91**(26): p. 12877-81.

43. Muller, C.H., *A revision of the genus Lycopersicon.* USDA Mis Publ, 1940. **382**: p. 1-28.

44. Cox, S., *From discovery to modern commercialism: the complete story behind Lycopersicon esculentum.* http://www.landscapeimagery.com/articles.html, 2000.

45. Rick, C.M. and R.T. Chetelat, *Utilization of related wild species for tomato*

improvement. Acta Hort, 1995. **412**: p.:21Ð38.

46. Spooner, D.M., G.J. Anderson and Jansen, R.K., *Chloroplast DNA evidence for the interrelationships of tomatoes, potatoes and pepinos (Solanaceae).* American Journal of Botany, 1993. **80**(676-688).

47. Bohs, L.a.O., R., *Phylogenetic relationships in Solanum (Solanaceae) based on ndhF sequences.* Sys. Bot., 1997. **22**: p. 5-17.

48. Olmstead, R.a.P., J., *Implications for the phylogeny, classification, and biogeography of Solanum from cpDNA restriction site variation.* Sys. Bot., 1997. **22**: p. 19-29.

49. Peralta, I.E., S. Knapp, and D.M. Spooner, *New Species of Wild Tomatoes (Solanum Section Lycopersicon: Solanaceae) from Northern Peru.* Systematic Botany, 2005. **30**(2): p. 424Ð434.

50. Peralta, I.E. and D.M. Spooner, Granule-bound starch synthase (GBSSI) gene phylogeny of wild tomatoes (*Solanum* L. section *Lycopersicon* [Mill.] Wettst. subsection *Lycopersicon*). American Journal of Botany, 2001. **88**(10): p. 1888Ð1902.

51. Darwin, S.C., S. Knapp, and I.E. Peralta, *Taxonomy of tomatoes in the Galapagos Islands: native and introduced species of Solanum section Lycopersicon*

(Solanaceae). Systematics and Biodiversity, 2003. **1**(1): p. 29Ð53.

52. Knapp, S. and S.C. Darwin, *Proposal to conserve the name Solanum cheesmaniae (L. Riley) Fosberg against S. cheesmanii Geras. (Solanaceae).* Taxon, 2007. **55**: p. 806Ð807.

53. Spooner, D.M., I.E. Peralta, and S. Knapp, *Comparisons of AFLPs with other markers for phylogenetic inference in wild tomatoes [SolanumL. section Lycopersicon (Mill.)Wettst.].* Taxon, 2005. **54**: p. 43Ð61.

54. USDA, A., National Genetic Resources Program, *Tomato Crop Germplasm Committee Report*, G.R.I.N.-. (GRIN), Editor. 2004.

55. Cong, B., L.S. Barrero, and S.D. Tanksley, *Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication.* Nat Genet, 2008. **40**(6): p. 800-4.

56. Peralta, I.E., D.M. Spooner, and S.K. Knapp, *Taxonomy of Wild Tomatoes and their Relatives*, ed. S.B.M.-S. (Solanaceae). Vol. 84. 2008.

57. Chetelat, R.T., et al., *Distribution, ecology and reproductive biology of wild tomatoes and related nightshades from the Atacama Desert region of northern Chile.* Euphytica, 2009. **167**: p. 77-93.

58. Dehan, K. and M. Tal, *Salt tolerance in the wild relatives of the cultivated tomato: responses of Solanum pennellii to high salinity.* Irrigation Science, 1978. **1**(1): p. 71-76.

59. Shalata, A. and M. Tal, *The effect of salt stress on lipid peroxidation and antioxidants in the leaf of the cultivated tomato and its wild salt tolerant relative Lycopersicon pennellii.* Physiol Plant, 2002. **104**(2): p. 169-174.

60. Venema, J.H., et al., *Differential response of domestic and wild Lycopersicon species to chilling under low light: growth, carbohydrate content, photosynthesis and the xanthophyll cycle.* Physiologia Plantarum, 1999. **105**: p. 81-88.

61. Stevens, M.A. and C.M. Rick, *Genetics and breeding. In: Atherton JG, Rudich J (eds) The tomato crop: a scientific basis for improvement.* Chapman and Hall London, 1986: p. 35Ð109.

62. Robert, V.J.M., et al., *Marker-assisted introgression of blackmold resistance QTL alleles from wild Lycopersicon cheesmanii to cultivated tomato (L.Êesculentum) and evaluation of QTL phenotypic effects.* Molecular Breeding, 2001. **8**(3): p. 217-233.

63. Huang, C.C., et al., *The resistance to powdery mildew (Oidium lycopersicum) in shape Lycopersicon species is mainly associated with hypersensitive response.* European Journal of Plant Pathology, 2004. **104**(4): p. 399-407.

64. Chetelat, R.T., et al., *A male-fertile Lycopersicon esculentum? Solanum lycopersicoides hybrid enables direct backcrossing to tomato at the diploid level.* Euphytica, 1997. **95**(1): p. 99-108.

65. Zamir, D., *Improving plant breeding with exotic genetic libraries.* Nat Rev Genet, 2001. **2**(12): p. 983-9.

66. Monforte, A.J. and S.D. Tanksley, *Development of a set of near isogenic and backcross recombinant inbred lines containing most of the Lycopersicon hirsutum genome in a L. esculentum genetic background: a tool for gene mapping and gene discovery.* Genome, 2000. **43**(5): p. 803-13.

67. Tanksley, S.D., et al., *Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative L. pimpinellifolium.* TAG Theoretical and Applied Genetics, 1996. **92**(2): p. 213-224.

68. Paran, I., et al., *Recombinant inbred lines for genetic mapping in tomato.* TAG

Theoretical and Applied Genetics, 1995. **90**(3-4): p. 542-548.

69. Frary, A., et al., *Fine mapping of quantitative trait loci for improved fruit characteristics from Lycopersicon chmielewskii chromosome 1.* Genome, 2003. **46**(2): p. 235-43.

70. Fulton, T.M., et al., *Advanced backcross QTL analysis of a Lycopersicon esculentum x L. parviflorum cross.* Theor Appl Genet, 2000. **100**: p. 1025-1042.

71. Palmer, J.D. and D. Zamir, *Chloroplast DNA evolution and phylogenetic relationships in Lycopersicon.* Proc Natl Acad Sci USA, 1982. **79**: p. 5006Ð5010.

72. Miller, J.C. and S.D. Tanksley, *RFLP analysis of phylogenetic relationships and genetic variation in the genus Lycopersicon.* Theor Appl Genet, 1990. **80**: p. 437Ð448.

73. Breto, M.P., M.J. Asins, and E.A. Carbonell, *Genetic variability in Lycopersicon species and their genetic relationships.* Theor Appl Genet, 1993. **86**: p. 113Ð120.

74. Alvarez, A.E., et al., *Use of microsatellites to evaluate genetic diversityand species relationships in the genus Lycopersicon.* Theor Appl Genet, 2001. **103**: p. 1283Ð1292.

75. Tomes, M.L., F.W. Quackenbush, and M. McQuistan, *Modification and Dominance of the Gene Governing Formation of High Concentrations of BETA-Carotene in the Tomato.* Genetics, 1954. **39**(6): p. 810-7.

76. Tomes, M.L., *Delta-Carotene in the Tomato.* Genetics, 1969. **62**(4): p. 769-

780.

77. Tomes, M.L., et al., *The Inheritance of Carotenoid Pigment Systems in the Tomato.* Genetics, 1953. **38**(2): p. 117-27.

78. Ronen, G., et al., *Regulation of carotenoid biosynthesis during tomato fruit development: expression of the gene for lycopene epsilon-cyclase is down-regulated during ripening and is elevated in the mutant Delta.* Plant J, 1999. **17**(4): p. 341-51.

79. Ronen, G., et al., *An alternative pathway to beta -carotene formation in plant chromoplasts discovered by map-based cloning of beta and old-gold color mutations in tomato.* Proc Natl Acad Sci U S A, 2000. **97**(20): p. 11102-7.

80. Hanson, P.M., et al., *Variation for antioxidant activity and antioxidants in tomato.* Journal of the American Society for Horticultural Science, 2004. **129**(5): p. 704-711.

81. Chen, F.Q., et al., *Mapping of QTLs for lycopene and other fruit traits in a Lycopersicon esculentum x L. pimpinellifolium cross and comparison of QTLs across tomato species.* Molecular Breeding, 1999. **5**: p. 283Ð299.

82. Rick, C.M. and S. tanksley, *Genetic variation in Solanum pennellii - Comparisons with 2 other sympatric tomato species.* Plant Syst Evol, 1981. **139**: p. 11Ð45.

83. Rick, C.M., J.F. Fobes, and S.D. Tanksley, *Evolution of mating systems in Lycopersicon hirsutum as deduced from genetic variation in electrophoretic andmorphological characters.* Plant Syst Evol, 1979. **132**: p. 279Ð298.

84. Rick, C.M., et al., *Genetic andbiosystematic studies on two new sibling speces*

of Lycopersicon from interandean Peru. Theor Appl Genet, 1976. **47**: p. 55Ð68.

85. Marshall, A.V., et al., *Molecular systematics of Solanum section Lycopersicum (Lycopersicon) using the nuclear ITS rDNA region.* Theor Appl Genet, 2001. **103**: p. 1216Ð1222.

86. Baudry, E., et al., *Species and recombination effects on DNA variability in the tomato genus.* Genetics, 2001. **158**: p. 1725Ð1735.

87. Roselius, K., W. Stephan, and T.R. Stadle, *The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species.* Genetics, 2005. **171**: p. 753Ð763.

88. Rothberg, J.M. and J.H. Leamon, *The development and impact of 454 sequencing.* Nat Biotechnol, 2008. **26**(10): p. 1117-24.

89. Pettersson, E., J. Lundeberg, and A. Ahmadian, *Generations of sequencing technologies.* Genomics, 2009. **93**(2): p. 105-11.

90. Peterson, D.G., et al., *DNA content of heterochromatin and euchromatin in tomato (Lycopersicon esculentum) pachytene chromosomes.* Genome, 1996. **39**(1): p. 77-82.

91. Wang, Y., et al., *Euchromatin and pericentromeric heterochromatin: comparative composition in the tomato genome.* Genetics, 2006. **172**(4): p. 2529-40.

92. Fulton, T.M., et al., *Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants.* Plant Cell, 2002. **14**(7): p. 1457-67.

93. Isaacs, N.W., et al., *Light-harvesting mechanisms in purple photosynthetic bacteria.* Curr Opin Struct Biol, 1995. **5**(6): p. 794-7.

94. Ohmiya, A., *Carotenoid cleavage dioxygenases and their apocarotenoid products in plants.* Plant Biotechnology, 2009. **26**: p. 351-358.

95. Giuliano, G., S. Al-Babili, and J. von Lintig, *Carotenoid oxygenases: cleave it or leave it.* Trends Plant Sci, 2003. **8**(4): p. 145-9.

96. Mathieu, S., et al., *Flavour compounds in tomato fruits: identification of loci and potential pathways affecting volatile composition.* J Exp Bot, 2009. **60**(1): p. 325-37.

97. Goff, S.A. and H.J. Klee, *Plant volatile compounds: sensory cues for health and nutritional value?* Science, 2006. **311**(5762): p. 815-9.

98. Romer, S. and P.D. Fraser, *Recent advances in carotenoid biosynthesis, regulation and manipulation.* Planta, 2005. **221**(3): p. 305-8.

99. Hirshberg, J., *Carotenoid biosynthesis in flowering plants.* Current Opinion in Plant Biology, 2001. **4**: p. 210-218.

100. Bouvier, F., et al., *Oxidative tailoring of carotenoids: a prospect towards novel functions in plants.* Trends Plant Sci, 2005. **10**(4): p. 187-94.

101. Hosseini, M., A. Hemati Kakhki, and A.-R. Karbasi, *STUDY OF SOCIO-ECONOMIC EFFECTS OF TEN YEARS RESEARCH ON SAFFRON (CROCUS SATIVUS L.).* ISHS Acta Horticulturae 850: III International Symposium on Saffron: Forthcoming Challenges in Cultivation, Research and Economics, 2009.

102. Horton, P. and A. Ruban, *Molecular design of the photosystem II light-*

harvesting antenna: photosynthesis and photoprotection. J Exp Bot, 2005. **56**(411): p. 365-73.

103. Szabo, I., E. Bergantino, and G.M. Giacometti, *Light and oxygenic photosynthesis: energy dissipation as a protection mechanism against photo-oxidation.* EMBO Rep, 2005. **6**(7): p. 629-34.

104. Alboresi, A., et al., *Antenna complexes protect Photosystem I from photoinhibition.* BMC Plant Biol, 2009. **9**: p. 71.

105. Floss, D.S. and M.H. Walter, *Role of carotenoid cleavage dioxygenase 1 (CCD1) in apocarotenoid biogenesis revisited.* Plant Signal Behav, 2009. **4**(3): p. 172-5.

106. Umehara, M., et al., *Inhibition of shoot branching by new terpenoid plant hormones.* Nature, 2008. **455**(7210): p. 195-200.

107. Gomez-Roldan, V., et al., *Strigolactone inhibition of shoot branching.* Nature, 2008. **455**(7210): p. 189-94.

108. Fan, J., et al., *Abscisic acid has a key role in modulating diverse plant-pathogen interactions.* Plant Physiol, 2009. **150**(4): p. 1750-61.

109. Salt, S.D., S. Tuzun, and J. Kuc, *Effects of beta-ionone and abscisic acid on the growth of tobacco and resistance to blue mold: mimicry of effects of stem infection by Peronospora tabacina.* Physiol. Mol. Plant Pathol., 1986. **28**: p. 287-297.

110. Fester, T., W. Maier, and D. Strack, *Accumulation of secondary compounds in barley and wheat roots in response to inoculation with an arbuscular mycorrhizal fungus and co-inoculation with rhizosphere bacteria.*

Mycorrhiza, 1999. **8**: p. 241-246.

111. Yeum, K.J. and R.M. Russell, *Carotenoid bioavailability and bioconversion.* Annu Rev Nutr, 2002. **22**: p. 483-504.

112. DellaPenna, D. and B.J. Pogson, *Vitamin synthesis in plants: tocopherols and carotenoids.* Annu Rev Plant Biol, 2006. **57**: p. 711-38.

113. West, K.P., Jr., *Vitamin A deficiency disorders in children and women.* Food Nutr Bull, 2003. **24**(4 Suppl): p. S78-90.

114. Block, G., *The data support a role for antioxidants in reducing cancer risk.* Nutr Rev, 1992. **50**(7): p. 207-13.

115. Ames, B.N., M.K. Shigenaga, and T.M. Hagen, *Oxidants, antioxidants, and the degenerative diseases of aging.* Proc Natl Acad Sci U S A, 1993. **90**(17): p. 7915-22.

116. Lorenzo, Y., et al., *The carotenoid beta-cryptoxanthin stimulates the repair of DNA oxidation damage in addition to acting as an antioxidant in human cells.* Carcinogenesis, 2009. **30**(2): p. 308-14.

117. Polidori, M.C., *Antioxidant micronutrients in the prevention of age-related diseases.* J Postgrad Med, 2003. **49**(3): p. 229-35.

118. Lam, J., et al., *Micronutrient supplementation in children and adults with HIV infection.* Cochrane Database Syst Rev., 2005. **19**(4).

119. Hubschman, J.P., S. Reddy, and S.D. Schwartz, *Age-related macular degeneration: experimental and emerging treatments.* Clin Ophthalmol, 2009. **3**: p. 167-74.

120. Hubschman, J.P., S. Reddy, and S.D. Schwartz, *Age-related macular*

degeneration: current treatments. Clin Ophthalmol, 2009. **3**: p. 155-66.

121. GrŸnewald, K., J. Hirschberg, and C. Hagen, *Ketocarotenoid biosynthesis outside of plastids in the unicellular green alga Haematococcus pluvialis.* J Biol Chem, 2001. **276**: p. 6023-6029.

122. Giuliano, G., et al., *Metabolic engineering of carotenoid biosynthesis in plants.* Trends Biotechnol, 2008. **26**(3): p. 139-45.

123. Lichtenthaler, H.K., et al., *Biosynthesis of isoprenoids in higher plant chloroplasts proceeds via a mevalonate-independent pathway.* FEBS Lett, 1997. **400**(3): p. 271-4.

124. Rohmer, M., *Mevalonate-independent methylerythritol phosphate pathway for isoprenoid biosynthesis. Elucidation and distribution.* Pure Appl. Chem., 2003. **75**(Nos. 2Ð3): p. 375Ð387.

125. Okada, K., et al., *Genetic evidence for the role of isopentenyl diphosphate isomerases in the mevalonate pathway and plant development in Arabidopsis.* Plant Cell Physiol, 2008. **49**(4): p. 604-16.

126. Araki, N., et al., *Temperature-sensitive Arabidopsis mutant defective in 1-deoxy-D-xylulose 5 phosphate synthase within the plastid non-mevalonate pathway of isoprenoid biosynthesis.* Physiol Plant,, 2000. **108**: p. 19-24.

127. Lois, L.M., et al., *Carotenoid biosynthesis during tomato fruit development: regulatory role of 1-deoxy-D-xylulose 5-phosphate synthase.* Plant J, 2000. **22**(6): p. 503-13.

128. Schwender, J., et al., *Cloning and heterologous expression of a cDNA encoding 1-deoxy-D-xylulose-5-phosphate reductoisomerase of Arabidopsis thaliana.*

FEBS Lett, 1999. **455**(1-2): p. 140-4.

129. Rodriguez-Concepcion, M., et al., *1-Deoxy-D-xylulose 5-phosphate reductoisomerase and plastid isoprenoid biosynthesis during tomato fruit ripening.* Plant J, 2001. **27**(3): p. 213-22.

130. Eisenreich, W., F. Rohdich, and A. Bacher, *Deoxyxylulose phosphate pathway to terpenoids.* Trends Plant Sci, 2001. **6**(2): p. 78-84.

131. Cunningham, F.X., Jr., T.P. Lafond, and E. Gantt, *Evidence of a role for LytB in the nonmevalonate pathway of isoprenoid biosynthesis.* J Bacteriol, 2000. **182**(20): p. 5841-8.

132. Cunningham, F.X. and E. Gantt, *Genes and Enzymes of Carotenoid Biosynthesis in Plants.* Annu Rev Plant Physiol Plant Mol Biol, 1998. **49**: p. 557-583.

133. Phillips, M.A., et al., *The Arabidopsis thaliana type I Isopentenyl Diphosphate Isomerases are targeted to multiple subcellular compartments and have overlapping functions in isoprenoid biosynthesis.* Plant Cell, 2008. **20**(3): p. 677-96.

134. Camara, B., *Plant phytoene synthase complex, component enzymes, immunology, and biogenesis.* Methods Enzymol, 1993. **214**: p. 352-365.

135. Bartley, G.E. and P.A. Scolnik, *cDNA cloning, expression during development, and genome mapping of PSY2, a second tomato gene encoding phytoene synthase.* J Biol Chem, 1993. **268**(34): p. 25718-21.

136. Giorio, G., A.L. Stigliani, and C. D'Ambrosio, *Phytoene synthase genes in tomato (Solanumlycopersicum L.) - new data on the structures, the deduced*

amino acid sequences and the expression patterns. Febs J, 2008. **275**(3): p. 527-35.

137. Fraser, P.D., et al., *Carotenoid biosynthesis during tomato fruit development.* Plant Physiol, 1994. **105**: p. 405-413.

138. Fraser, P.D., et al., *Manipulation of phytoene levels in tomato fruit: effects on isoprenoids, plastids, and intermediary metabolism.* Plant Cell, 2007. **19**(10): p. 3194-211.

139. Pecker, I., et al., *A single polypeptide catalyzing the conversion of phytoene to zeta-carotene is transcriptionally regulated during tomato fruit ripening.* Proc Natl Acad Sci U S A, 1992. **89**(11): p. 4962-6.

140. Bartley, G.E., P.A. Scolnik, and P. Beyer, *Two Arabidopsis thaliana carotene desaturases, phytoene desaturase and zeta-carotene desaturase, expressed in Escherichia coli, catalyze a poly-cis pathway to yield pro-lycopene.* Eur J Biochem, 1999. **259**(1-2): p. 396-403.

141. Isaacson, T., et al., *Cloning of tangerine from tomato reveals a carotenoid isomerase essential for the production of beta-carotene and xanthophylls in plants.* Plant Cell, 2002. **14**(2): p. 333-42.

142. Isaacson, T., et al., *Analysis in vitro of the enzyme CRTISO establishes a poly-cis-carotenoid biosynthesis pathway in plants.* Plant Physiol, 2004. **136**(4): p. 4246-55.

143. Pecker, I., et al., *Cloning and characterization of the cDNA for lycopene beta-cyclase from tomato reveals decrease in its expression during fruit ripening.* Plant Mol Biol, 1996. **30**(4): p. 807-19.

144. Lefebvre, V., et al., *The capsanthin-capsorubin synthase gene: a candidate gene for the y locus controlling the red fruit color in pepper.* Plant Mol Biol, 1998. **36**(5): p. 785-9.

145. Hugueney, P., et al., *Metabolism of cyclic carotenoids: a model for the alteration of this biosynthetic pathway in Capsicum annuum chromoplasts.* Plant J, 1995. **8**(3): p. 417-24.

146. Inoue, K., *Carotenoid hydroxylation--P450 finally!* Trends Plant Sci, 2004. **9**(11): p. 515-7.

147. Fiore, A., et al., *Elucidation of the beta-carotene hydroxylation pathway in Arabidopsis thaliana.* FEBS Lett, 2006. **580**(19): p. 4718-22.

148. Bouvier, F., et al., *Xanthophyll biosynthesis: molecular and functional characterization of carotenoid hydroxylases from pepper fruits (Capsicum annuum L.).* Biochim Biophys Acta, 1998. **1391**(3): p. 320-8.

149. Hirschberg, J., *Molecular biology of carotenoid biosynthesis.*, in *Carotenoids*, L.-J.S. Edited by Britton G, Pfander H. Basel: Birkhauser Verlag, Editor. 1998. p. 149-194.

150. Galpaz, N., et al., *A chromoplast-specific carotenoid biosynthesis pathway is revealed by cloning of the tomato white-flower locus.* Plant Cell, 2006. **18**(8): p. 1947-60.

151. Tian, L., et al., *The Arabidopsis LUT1 locus encodes a member of the cytochrome p450 family that is required for carotenoid epsilon-ring hydroxylation activity.* Proc Natl Acad Sci U S A, 2004. **101**(1): p. 402-7.

152. Burbidge, A., et al., *Structure and expression of a cDNA encoding zeaxanthin*

epoxidase, isolated from a wilt-related tomato (Lycopersicon esculentum Mill.) library. Journal of Experimental Botany, 1997. **48**(9): p. 1749-1750.

153. Bouvier, F., et al., *Identification of neoxanthin synthase as a carotenoid cyclase paralog.* Eur J Biochem, 2000. **267**(21): p. 6346-52.

154. Al-Babili, S., et al., *Identification of a novel gene coding for neoxanthin synthase from Solanum tuberosum.* FEBS Lett, 2000. **485**(2-3): p. 168-72.

155. Lincoln, R.E., et al., *Provitamin A and vitamin C in the genus Lycopersicon.* Bot Gaz, 1943. **105**: p. 113Ð115.

156. Kohler, G.W., et al., *Selection and breeding for high beta-carotene content (provitamin A) in tomato.* Bot Gaz, 1947. **109**: p. 219-225.

157. Rick, C.M., *New mutants.* Tomato Genet Coop Rep, 1956. **6**: p. 22-23.

158. Chalukova, M., *Carotenoid composition of the fruits of hybrids between Lycopersicon esculentum and some wild species of the genus Lycopersicon, IV. Progenies of lycopene*

and β-carotene BC1P1 hybrids of L. chmielewskii. Genet Breed, 1988. **21**: p. 49Ð57.

159. Stommel, J.R. and K.G. Haynes, *Inheritance of beta carotene*

content inthewild tomato species Lycopersicon cheesmanii. J Hered, 1994. **85**: p. 401Ð404.

160. Thompson, A.E., et al., *Inheritance of crimson fruit color in tomatoes.* Proc Am Soc Hort Sci, 1967. **91**: p. 495Ð504.

161. MacArthur, J., *Linkage groups in the tomato.* J Genet, 1934. **29**: p. 123Ð133.

162. Boileau, A.C., et al., *Cis-lycopene is more bioavailable than trans-lycopene in vitro and in vivo in lymph-cannulated ferrets.* J Nutr, 1999. **129**(6): p. 1176-

81.

163. Boileau, T.W., A.C. Boileau, and J.W. Erdman, Jr., *Bioavailability of all-trans and cis-isomers of lycopene.* Exp Biol Med (Maywood), 2002. **227**(10): p. 914-9.

164. Tomes, M.L., *Temperature inhibition of carotene synthesis in tomato.* Bot Gaz, 1963. **124**: p. 180Ð185.

165. van Tuinen, A., et al., *The mapping of phytochrome genes and photomorphogenic mutants of tomato.* Theor Appl Genet, 1997. **94**(1): p. 115-22.

166. Cookson, P.J., et al., *Increases in cell elongation, plastid compartment size and phytoene synthase activity underlie the phenotype of the high pigment-1 mutant of tomato.* Planta, 2003. **217**(6): p. 896-903.

167. Mustilli, A.C., et al., *Phenotype of the tomato high pigment-2 mutant is caused by a mutation in the tomato homolog of DEETIOLATED1.* Plant Cell, 1999. **11**(2): p. 145-57.

168. Konsler, T.R., *Three mutants appearing in ÔManapalÕ tomato.* HortScience, 1973. **8**: p. 331-333.

169. Kerckhoffs, L.H.J. and R.E. Kendrick, *Photocontrol of anthocyanin biosynthesis in tomato.* J Plant Res, 1997. **110**: p. 141Ð149.

170. Galpaz, N., et al., *Abscisic acid deficiency in the tomato mutant high-pigment 3 leading to increased plastid number and higher fruit lycopene content.* Plant J, 2008. **53**(5): p. 717-30.

171. Rick, C.M. and L. Butler, *Cytogenetics of the tomato.* Adv Genet, 1956. **8**: p.

267Ð382.

172. Jenkins, J.A. and G. Mackinney, *Carotenoids of the Apricot Tomato and Its Hybrids with Yellow and Tangerine.* Genetics, 1955. **40**(5): p. 715-20.

173. Zscheile, F.P. and J.W. Lesley, *Pigment analysis of sherry: Flesh color mutation resembling yellow in the tomato.* J Hered, 1967. **58**: p. 193Ð194.

174. Josse, E.M., et al., *A plastid terminal oxidase associated with carotenoid desaturation during chromoplast differentiation.* Plant Physiol, 2000. **123**(4): p. 1427-36.

175. Sambrook, J., E.F. Fritsch, and T. Maniatis, *Molecular Cloning. A Laboratory Manual (Second Edition).* ed. C.S.H.C.S.H.L. Press. 1989.

176. Lopez-Gomez, R. and M.A. Gomez-Lim, *A method for extracting intact RNA from fruits rich in polysaccharides using ripe mango mesocarp.* Hort Science, 1992. **27**(440-442).

177. Estrada, A.F., et al., *Novel apocarotenoid intermediates in Neurospora crassa mutants imply a new biosynthetic reaction sequence leading to neurosporaxanthin formation.* Fungal Genet Biol, 2008. **45**(11): p. 1497-1505.

178. Yang, Z. and J.P. Bielawski, *Statistical methods for detecting molecular adaptation.* Trends Ecol Evol, 2000. **15**(12): p. 496-503.

179. Saitou, N. and M. Nei, *The neighbor-joining method: a new method for reconstructing phylogenetic trees.* Mol Biol Evol, 1987. **4**(4): p. 406-25.

180. Porra, R.J., W.A. Thompson, and P.E. Kriedermann, *Determination of accurate extinction coefficients and simultaneous equation for assaying*

chlorophyll a and b extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. Biochim Biophys Acta, 1989(97): p. 384-394.

181. Rodriguez-Amaya, D.B., *A guide to carotenoid analysis in foods*, ed. O.R.I.H.N. Institute. 2001, Washington, D. C.

182. von Heijne, G. and K. Nishikawa, *Choloplast transit peptides. The perfect random coil?* Febs J, 1991. **278**(1): p. 1-3.

183. Bruce, B.D., *Chloroplast transit peptides: structure, function and evolution.* Trends in Cell Biology, 2000. **10**(10): p. 440-447.

184. Bousquet, J., et al., *Extensive variation in evolutionary rate of rbcL gene sequences among seed plants.* Proc Natl Acad Sci U S A, 1992. **89**(16): p. 7844-8.

185. Qin, G., et al., *Disruption of phytoene desaturase gene results in albino and dwarf phenotypes in Arabidopsis by impairing chlorophyll, carotenoid, and gibberellin biosynthesis.* Cell Res, 2007. **17**(5): p. 471-82.

186. Dong, H., et al., *The Arabidopsis Spontaneous Cell Death1 gene, encoding a zeta-carotene desaturase essential for carotenoid biosynthesis, is involved in chloroplast development, photoprotection and retrograde signalling.* Cell Res, 2007. **17**(5): p. 458-70.

187. Dall'Osto, L., et al., *Lutein is needed for efficient chlorophyll triplet quenching in the major LHCII antenna complex of higher plants and effective photoprotection in vivo under strong light.* BMC Plant Biol, 2006. **6**: p. 32.

188. Schauer, N., D. Zamir, and A.R. Fernie, *Metabolic profiling of leaves and fruit*

of wild species tomato: a survey of the Solanum lycopersicum complex. J Exp Bot, 2004. **56**(410): p. 297-307.

189. Grumet, R., J.F. Fobes, and R.C. Herner, *Ripening Behavior of Wild Tomato Species.* Plant Physiol, 1981. **68**(6): p. 1428-1432.

190. Loivamaki, M., et al., *Arabidopsis, a model to study biological functions of isoprene emission?* Plant Physiol, 2007. **144**(2): p. 1066-78.

191. Gancel, A. L., Alter, P., Dhuique-Mayer, C., Ruales, J. and Vaillant, F., *Identifying carotenoids and phenolic compounds in naranjilla (Solanum quitoense Lam. var. Puyo hybrid), an Andean fruit.* J Agric Food Chem, 2008, **56**(24): p.11890-99

192. Giovannoni, J. J., *Breeding new life into plant metabolism.* Nat Biotechnol, 2006, **24**(4): p. 418-9

10. Published paper

The tomato genome is being sequenced by an international consortium of 10 countries, including Italy. Our lab takes part to this project and a preliminary paper on the tomato genome has been published in the course of 2009.
Giulia Falcone took part of the team responsible for the BAC analysis, mapping and GenBank submission.

Mueller, L.A., Ren□ Klein Lankhorst, Steven D. Tanksley, James J. Giovannoni, Ruth White, Julia Vrebalov, Zhangjun Fei, Joyce van Eck, Robert Buels, Adri A. Mills, Naama Menda, Isaak Y. Tecle, Aureliano Bombarely, Stephen Stack, Suzanne M. Royer, Song-Bin Chang, Lindsay A. Shearer, Byung Dong Kim, Sung-Hwan Jo, Cheol-Goo Hur, Doil Choi, Chang-Bao Li, Jiuhai Zhao, Hongling Jiang, Yu Geng, Yuanyuan Dai, Huajie Fan, Jinfeng Chen, Fei Lu, Jinfeng Shi, Shouhong Sun, Jianjun Chen, Xiaohua Yang, Chen Lu, Mingsheng Chen, Zhukuan Cheng, Chuanyou Li, Hongqing Ling, Yongbiao Xue, Ying Wang, Graham B. Seymour, Gerard J. Bishop, Glenn Bryan, Jane Rogers, Sarah Sims, Sarah Butcher, Daniel Buchan, James Abbott, Helen Beasley, Christine Nicholson, Clare Riddle, Sean Humphray, Karen McLaren, Saloni Mathur, Shailendra Vyas, Amolkumar U. Solanke, Rahul Kumar, Vikrant Gupta, Arun K. Sharma, Paramjit Khurana, Jitendra P. Khurana, Akhilesh Tyagi, Sarita, Parul Chowdhury, Smriti Shridhar, Debasis Chattopadhyay, Awadhesh Pandit, Pradeep Singh, Ajay Kumar, Rekha Dixit, Archana Singh, Sumera Praveen, Vivek

Dalal, Mahavir Yadav, Irfan Ahmad Ghazi, Kishor Gaikwad, Tilak Raj Sharma, Trilochan Mohapatra, Nagendra Kumar Singh, Dora Szinay, Hans de Jong, Sander Peters, Marjo van Staveren, Erwin Datema, Mark W.E.J. Fiers, Roeland C.H.J. van Ham, P. Lindhout, Murielle Philippot, Pierre Frasse, Farid Regad, Mohamed Zouine, Mondher Bouzayen, Erika Asamizu, Shusei Sato, Hiroyuki Fukuoka, Satoshi Tabata, Daisuke Shibata, Miguel A. Botella, M. Perez-Alonso, V. Fernandez-Pedrosa, Sonia Osorio, Amparo Mico, Antonio Granell, Zhonghua Zhang, Jun He, Sanwen Huang, Yongchen Du, Dongyu Qu, Longfei Liu, Dongyuan Liu, Jun Wang, Zhibiao Ye, Wencai Yang, Guoping Wang, Alessandro Vezzi, Sara Todesco, Giorgio Valle, **Giulia Falcone**, Marco Pietrella, Giovanni Giuliano, Silvana Grandillo, Alessandra Traini, Nunzio D'Agostino, Maria Luisa Chiusano, Mara Ercolano, Amalia Barone, Luigi Frusciante, Heiko Schoof, Anika Jocker, Rémy Bruggmann, Manuel Spannagl, Klaus X.F. Mayer, Roderic Guigò, Francisco Camara, Stephane Rombauts, Jeffrey A. Fawcett, Yves Van de Peer, Sandra Knapp, Dani Zamir, and Willem Stiekema,

A Snapshot of the Emerging Tomato Genome Sequence.

The Plant Genome, 2009. **2**(1): p. 78Đ92.

11. Acknowledgments

Il mio primo ringraziamento è per i miei tutor, il Prof. Roberto Bassi per la sua supervisione e il Prof. Giovanni Giuliano per i suoi molti insegnamenti e per avermi dato la possibilità di svolgere questa tesi nei suoi laboratori, dove non mancano mai cose nuove da imparare.

Grazie al mio collega/amico Elio Fantini (3G) che ha condiviso con me questo percorso e che mi ha iniziato all'arte della serra, senza di lui, quanto meno all'inizio, le mie piante non sarebbero sopravissute!

Grazie al mio collega/amico Marco Pietrella per…tutto: per aver contribuito al lavoro bioinformatico, per avermi insegnato tanto, per avermi consigliato tanto e sopratutto per avermi sopportato.

Grazie a tutti gli altri miei compagni di laboratorio, in particolare Alessia (sv), Gianfranco e Federico, per il loro fondamentale aiuto con le analisi biochimiche, Silvia, Mireia, Sara e Paola per la loro disponibilità e il loro prezioso aiuto.

E infine ma non per ultimo Grazie alla mia famiglia che mi ha sostenuto e incoraggiato nei momenti di difficoltà e al mio Michele per la sua presenza così..solare.