Dissecting regulatory eQTLs of the carotenoid biosynthetic pathway in tomato fruit using *S.lycopersicum* x *S. pennellii* introgression lines

Jie Li

A thesis submitted to the University of East Anglia for the degree of Doctor of Philosophy

John Innes Centre

Norwich

June 2018

Abstract

Tomato (*Solanum lycopersicum*) ripening involves a number of physiological processes that include the visible breakdown of chlorophyll and build-up of carotenoids, with massive accumulation of antioxidant compounds such as lycopene and β-carotene (provitamin A) within the chromoplasts. Although the catalytic steps of the carotenoid biosynthetic pathway have been well-characterised, the regulatory mechanisms that control carotenoid accumulation remain poorly understood.

In this study, I used expression quantitative trait loci (eQTL) analysis to obtain a comprehensive understanding of the genetic basis of regulation of carotenoid biosynthesis in tomato fruit using the *S. lycopersicum* x *S. pennellii* introgression population. In total, 31 *cis*-eQTLs related to 18 carotenoid biosynthetic genes were identified, and the isoforms of some structural genes functional during fruit ripening were identified in this analysis. Six *trans*-eQTL hotspots were identified for lycopene biosynthesis. Co-expression analysis of one of the *trans*-eQTL candidates, residing in IL2-1, revealed that a basic helix-loop-helix transcription factor, SlHONG likely acts as a positive regulator of lycopene accumulation in fruit, by activating the expression of the structural genes involved in lycopene biosynthesis. SlRIN, a master regulator of fruit ripening, binds directly to two adjacent CArG motifs in the promoter of *SlHONG*, to control the expression of *SlHONG* in an ethylene-independent manner. Genome-scale phylogenetic analysis of bHLH proteins in Arabidopsis and tomato found SlHONG was the most similar protein to BEE2 in Arabidopsis, which is involved in brassinosteroid signalling by forming the BEE complex together with another three bHLH transcription factors. bHLH transcription factors potentially involved in the BEE complex with SlHONG in tomato, were identified by phylogenetic analysis and interactions between the candidate proteins were detected. SlHONG was identified as a regulator of lycopene biosynthesis in tomato fruit, and may be under the control of multi-layered regulatory mechanisms.

Acknowledgements

This is the last bit to finish the whole thesis. It is 22:59 and I am sitting in the office where I finished almost all the writing. This is the fifth year since I came to Norwich to start my PhD back to 2013 the journey started. When I am almost there and look back, there are so many people I have like to thank, without whom all of these would not be possible.

Firstly, I would like to thank my supervisor, Cathie, who is probably doing the last check for my thesis now. The journey of my PhD study was not a peaceful one. There were many high highs and low lows, and Cathie was the one who help me, supported me and encouraged me all the way here. Even when I was very disappointed in myself, she was the one who trusted me and was always there to get me out of the darkness. 'Science must be fun.' That is the best interpretation I have ever come across about research, and will be the rule for my scientific career. I benefit a lot from her enthusiasm for science, dedication to work and wisdom of life. I also would like to thank my mentors, Dr Nick Brewin, Dr Kate Conway, Dr Stephen Bornemann and Dr Yiliang Ding who supported and help me with their sincere and valuable advice when I got confused about my PhD life.

I have been fortunate to have so many brilliant and passionate people around me to share my science with. All the members of Cathie's group have been great, supportive and helpful all the time, who always can inspire me with new ideas during the discussion. I would like to thank kind Dr Eugenio Butelli, Dr Yang Zhang, Dr Paul Baily, Dr Wanmei Jin for their kindly help and contribution to my project. I also would like to thank Dr Jie Luo, Dr Shouchuang Wang and Dr Wei Chen as fantastic collaborators in the project.

I would like to thank my parents who always encourage me to do things what I think is right. They are my strongest backing all the time even though they are far in China. I would like to thank all my friends in UK and all over the world. It feels so good to have you in my life.

After nearly five years, my PhD study is about to be finished. However, my scientific career has just started.

Table of Contents

List of Figures

List of Tables

Abbreviations

ABA	abscisic acid
ACC	1-Aminocyclopropane-1-carboxylate
ACS	ACC synthase
AD	activation domain
Ade (A)	aminodecanoic acid
AMD	age-related macular degeneration
ARF4	*Auxin Response Factor 4*
AU	absorption Units
B	breaker
BCAA	branched-chain amino acid
BCH	ferredoxin-dependent non-heme β-ring hydroxylases
BD	DNA-binding domain
BEE1	*BR-ENHANCED EXPRESSION 1*
BEE2	*BR-ENHANCED EXPRESSION 2*
BEE3	*BR-ENHANCED EXPRESSION 3*
bHLH	basic helix-loop-helix
BHT	butylated hydroxytoluene
BOC1	β-carotene oxygenase 1
BR	Brassinosteroid
bZIP	basic leucine zipper
BZR1	*BRASSINAZOLE RESISTANT1*
Carb	Carbenicillin
ChIP	chromatin immunoprecipitation
CHY	carotenoid hydroxylase
cis-eQTL	*cis*-acting QTL
Cm	Chloramphenicol
CMK	4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase

CMS	4-(cytidine 5'-diphospho)-2-C-methyl-D synthase
Cnr	*Colourless non-ripening*
CPD	*Constitutive Photomorphogenesis and Dwarfism*
CRISPR	clustered, regularly interspaced, short palindromic repeat
crRNA	CRISPR RNA
CRTISO	carotenoid isomerase
CTR1	*constitutive triple response 1*
CVD	cardiovascular disease
CYP450	Cytochrome P450 hydroxylases
CYP97C	carotene ε-ring hydroxylase
DET1	*DE-ETIOLATED 1*
DLR	Dual-Luciferase® Reporter
DOXP	1-deoxy-D-xylulose-5-phosphate
DPA	days post anthesis
DR	*Delila*/*Rosia1*
DSB	double strand breaks
DW	dry weight
DXR	1-deoxy-D-xylulose 5-phosphate reductoisomerase
DXR	1-deoxy-D-xylulose 5-phosphate reductoisomerase
DXS	1-deoxyxylulose-5-phosphate synthase
DXS	1-deoxyxylulose-5-phosphate synthase
EBR	2,4-Epibrassinolide
eQTL	expression quantitative trait locus
ERF	Ethylene Responsive Factor
ETR	ethylene receptor
EV	empty vector
FAO	Food and Agriculture Organization
FaST-LLM	Factored Spectrally Transformed Linear module
FUL1	*FRUITFULL 1*
FUL2	*FRUITFULL 2*

Gent	Gentamicin
GGDP	geranylgeranyl diphosphate
GGPP	geranyl-geranyl diphosphate
GGPPS	geranyl-geranyl diphosphate (GGPP) synthase
GGPPS	geranyl-geranyl diphosphate (GGPP) synthase
GLK1	*Golden1-Like*
GLK2	*Golden2-Like*
Gr	*Green ripe*
GSL	glucosinolate
GWAS	genome-wide association studies
HD	homeodomain
HDR	1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate reductase
HDR	4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase
HDS	hydroxymethylbutenyl 4-diphosphate synthase
HGA	homogentisic acid
HGGT	homogentisate geranylgeranyl transferase
His (H)	histidine
HMBPP	hydroxymethylbutenyl 4-diphosphate
Hp1	*high pigment 1*
hp2	*high pigment 2*
Hp3	*high pigment 3*
HPLC	High Performance Liquid Chromatography
HR	homology-dependent recombination
HY5	*Elongated Hypocotyl 5*
HYD-B	β-Carotene hydroxylases
HYDE	ε-carotene hydroxylases
IAA	indole-3-acetic acid (auxin)
IG	Immature green
IPP	isopentenyl diphosphate isomerases
KO	knock out

LC/GC-MS	Liquid/Gas chromatography-mass spectrometry
Leu (L)	leucine
LLM	linear mixed model
LYCB	lycopene β-cyclase
LYCE	lycopene ε-cyclase
MAF	minor allele frequency
MCS	ME-cPP synthase
MEP	methylerythritol 4-phosphate
MDA	malondialdehyde
MG	mature green
mGWAS	metabolite-based genome-wide association studies
MTBE	methyl tert-butyl ether
MVA	Mevalonate
MYB	myoblastoma
NCED	9-cisepoxycarotenoid dioxygenase
NHEJ	error-prone nonhomologous end-joining
NMR	nuclear magnetic resonance
NPQ	Non-Photochemical Quenching
Nr	*Never ripe*
og	*old gold*
Or	*Orange*
P	pericarp
PAM	protospacer adjacent motif
PCR	polymerase chain reaction
PDS	phytoene desaturase
PIF1	*Phytochrome-interacting Factor 1*
PSY	phytoene synthase
PTOX	alternative oxidase
QTG	quantitative trait gene
QTL	quantitative trait loci

r	*yellow flesh*
RDA	recommended daily allowances
Rif	Rifampicin
rin	*ripening inhibitor*
ROS	reactive oxygen species
RT	retention time
RT-qPCR	Real-time quantitative PCR
SGN	SOL Genomic Network
SlAPRR2-like	*Arabidopsis Pseudo-Response Regulator 2-like*
SNP	single nucleotide polymorphism
t	*tangerine*
TAGL1	*Tomato AGAMOUS-like 1*
TF	Transcription factor
TILLING	Targeting Induced Local Lesions IN Genomes
tracrRNA	trans-activating crRNA
trans-eQTL	*trans*-acting QTL
Trp (W)	tryptophan
TSS	transcription start site
TTG1	*Transparent Testa Glabra 1*
TTG2	*Transparent Testa Glabra 2*
TTG8	*Transparent Testa Glabra 8*
UAS	Upstream Activator Sequences
VDE	violaxanthin deepoxidase
VIGS	virus-induced gene silencing
wf	*white flower*
ZDS	ζ -carotene desaturase
ZEP	zeaxanthin epoxidase
ZISO	ζ-carotene isomerase

Chapter 1

General Introduction

1.1 Carotenoids

Carotenoids are a subgroup of isoprenoids containing more than 700 members widely distributed in bacteria, fungi, algae and plants. Carotenoids are mainly composed of 40 carbons forming their polyene backbones with double bonds, and some of them have rings at the ends. The conjugate double bonds contribute to the absorption of the visible light, resulting in the colour of this group of chemicals which ranges from colourless to yellow, orange and red, which is reflected in the pigmentation of organs such as flowers and fruits.

1.1.1 Carotenoids in human health

Out of more than 700 carotenoids produced by plants, six (α-carotene, β-carotene, lycopene, lutein, zeaxanthin, and β-cryptoxanthin) seem to be important for human health. These also represent the most abundant carotenoids found in humans (Maiani *et al*., 2009). More than 40 carotenoids are taken up from a typical human diet, most of which are obtained from fruit and vegetables (Mangels *et al*., 1993; Johnson, 2002). Consumption of carotenoids in the human diet is believed to provide health benefits in protection against certain kinds of disease, especially eye disease and certain cancers. Carotenoids have a wide range of biological activities in promoting human health, such as the well-known provitamin A activity, enhancement of the immune system and antioxidant ability. The most studied dietary carotenoids are lycopene, β-carotene, lutein, zeaxanthin and β-cryptoxanthin.

During the past few decades, there have been many publications describing the antioxidant and pro-oxidant activity of different kinds of carotenoids (Oguz, 2017; Tanumihardjo, 2013; Young and Lowe, 2001; Cvetkovic et al., 2013; Palozza, 1998). However, scepticism about whether carotenoids have antioxidant capacity in vivo remains, which mainly because of the different methodologies used in the research to confirm activity (Rice-Evans et al., 1997; Briviba et al., 2004). Since the first paper published in 1968 characterised the antioxidant ability of carotenoids in vitro provided the evidence to the support the role of carotenoids

in photosensitised oxidation, extensive studies have been performed in vitro, involving widespread use of liposomes, which have been clearly demonstrated to incorporate carotenoids and showed that carotenoids offer protection against AAPH-induced lipid peroxidation (Anderson and Krinsky, 1973; Liebler et al., 1997; Albrecht et al., 2000). Similar studies have been conducted using other tissues as well (Zhang and Omaye, 2001, 2000). Building on studies on various animal species (Bhuvaneswari et al., 2001; Palozza et al., 2000; Matos et al., 2001), the antioxidant effects of carotenoids have also been studied in humans. Carotenoid-deficient diets can elevate levels of malondialdehyde (MDA), which is a widely used biomarker of oxidative stress status in human, a situation which can be reversed by the supplementation with a mixture of carotenoids (Dixon et al., 1994, 1998). Decreases in the levels of MDA have been observed in other studies involving supplementation with β-carotene at various dosages (Winklhofer-Roob et al., 1995; Meraji et al., 1997; Lepage et al., 1996). Furthermore, changes in oxidative stress biomarkers, such as DNA damage level, were detected in other experiments after treating with either certain types of carotenoids, such as β-carotene and lycopene, or a mixed carotenoid preparation (Collins et al., 1998; Torbergsen and Collins, 2000; Lorenzo et al., 2009; Lee et al., 2000). The antioxidant activity provided by lycopene can help protect against degenerative diseases by neutralizing free radicals in the body, resulting in the prevention of DNA damage in the cells and the improvement of cell function. Lycopene can also protect skin from UV damage, which helps the skin to look youthful and prevent sunburn (Wang and Chen, 2006). These results show that carotenoids are very likely acting as antioxidants in humans.

Various epidemiological studies have shown a reduced risk of cancer associated with intake of carotenoid-rich diets, suggesting that natural carotenoids have anticarcinogenic capacity. It has been observed that a reduced risk of breast cancer was related to increased intake of β-carotene (Cho et al., 2003; Eliassen et al., 2015). A negative relationship between β-carotene consumption and the risk of lung cancer has also been reported (Virtamo et al., 2003; Omenn et al., 1996; Cook et al., 2000; Lee et al., 1999), although a Cochrane review in 2012 of 78 randomised trials with 296,707 participants concluded that antioxidant supplements have little or no effect on health outcomes and an increased risk of mortality associated with consumption of beta-carotene (Bjelakovic et al., 2015). The United States Preventive Services Task Force (USPSTF) recommended against β-carotene supplements for

the prevention of cardiovascular disease or cancer (Moyer, 2014). In contrast, α-carotene showed higher potency than β-carotene in prevention of tumorigenesis in lung, liver and skin (Nishino et al., 2009). It has been shown that ripe tomato fruits and its related products provide approximately 85% of the lycopene found in the human diet (Canene-Adams et al., 2005). Lycopene has been shown to reduce the risk of prostate cancer (Etminan et al., 2004; Stacewicz-Sapuntzakis and Bowen, 2005). The earliest case-control study in 1979 evaluating the relationship between a lycopene-rich diet and cancer risk reported 40% reduction in the risk of oesophageal cancer (Cook-Mozaffari et al., 1979). Other reverse associations between lycopene intake or serum lycopene concentrations and cancer risk have been reported, such as gastric cancer, breast cancer and bladder cancer (Buiatti et al., 1989; Helzlsouer et al., 1989; Zhang et al., 1997). Another two well-studied carotenoids, lutein and β-cryptoxanthin, also showed strong antitumor activity when applied separately or combined with other carotenoids (Satia et al., 2009; Yuan et al., 2003).

A group of carotenoids, called provitamin A, containing a β-ionone ring at the end(s) of the isoprenoid chain, can be enzymatically cleaved by β-carotene oxygenase 1 (BOC1) in mammalian systems to produce vitamin A, including β-carotene, α-carotene, β-cryptoxanthin, α-cryptoxanthin and γ-carotene (von Lintig and Vogt, 2000). Deficiency in vitamin A is the major cause of blindness in children, and is also associated with an increased burden of infectious disease, night blindness, xerophthalmia (dry eye syndrome), and increased risk of mortality (West, 2003). Once consumed, β-carotene can be cleaved to form two molecules of retinol (vitamin A), and consequently is considered a better source of vitamin A than α-carotene or β-cryptoxanthin which can also be metabolised to a single molecule of retinol following digestion. Lutein and zeaxanthin, which are richest in green leafy vegetables (such as spinach, broccoli, peas and lettuce), protect against the development of age-related macular degeneration (AMD), due to their selective accumulation in the macula of the retina of the eye (Bone & Landrum 1992).

Some epidemiological studies have reported an inverse association between the intake of carotenoids and other antioxidant vitamins in the diet and the risk of cardiovascular disease (CVD) (Mayne, 1996). According to the antioxidant properties of carotenoids, carotenoids

have been assumed to play a role in the prevention of CVD. A significant inverse association between the intake of β-carotene and α-carotene and coronary artery disease risk in women has been observed (Osganian et al., 2003). Similarly, higher concentrations of lycopene in plasma were reported associated with a significantly lower risk of CVD, and significantly elevated mortality was observed in subjects with initially low plasma carotenoid concentrations (Sesso et al., 2004; Gey et al., 1993). However, the health benefits of dietary carotenoids still need further investigation, because there are some epidemiological and intervention trials which found no significant association between dietary intake of carotenoids or plasma/serum concentration of carotenoids with corresponding risk of CVD or the mortality resulting from heart disease (Hennekens et al., 1996; Greenberg et al., 1996).

Due to their chemical and biological properties, other health benefits of carotenoids have been reported. For example, because of their ability to absorb light, studies suggest that carotenoids may play a role in the protection of skin during sun exposure (Mathews-Roth, 1993), and the consumption of carotenoids, like β-carotene, lycopene, phytoene and phytofluene, can alter sensitivity to UV light and help prevent UV-light-induced erythema in healthy individuals (Stahl et al., 2001; Aust et al., 2005; Mathews-Roth, 1983). Lycopene has also been reported to protect against the development of type 2 diabetes, an observation which remains controversial because other studies do not support this prospective association, making further investigation necessary (Valero et al., 2011; Wang et al., 2006). As one of the major groups of phytonutrients, carotenoids have been widely studied and their health benefits are generally accepted especially for dietary consumption from fruits and vegetables. Therefore, understanding the genetic and regulatory basis of carotenoid biosynthesis in plants is important.

1.1.2 Carotenoids in plants

Carotenoids can be produced in both photosynthetic and non-photosynthetic organs of plants. Carotenoids serve an integral role in the photosystem in higher plants, the

composition of which is highly conserved in different plant species, and includes carotenoids such as lutein, β-carotene, neoxanthin and violaxanthin.

In photosynthetic green tissues, carotenoids perform essential roles as accessory light-harvesting pigments, in photosystem assembly and in photoprotection (Domonkos et al., 2013). Carotenoids are thought to provide the first level of defence against photodamage caused by reactive oxygen species (ROS), which are by-products of photosynthesis, photosynthetic electron transport and oxygenation of Rubisco in photorespiration (Species et al., 2001; Peterhansel and Maurino, 2011). ROS can induce damage to lipids and proteins, leading to subsequent reduced photosynthetic efficiency and impaired growth and development. Carotenoids serve a crucial role in protecting plants from such damage, which can be achieved in two ways, physical quenching and chemical quenching (Stahl and Sies, 2003; Krieger-Liszkay, 2005). Non-Photochemical Quenching (NPQ) is mechanism exploited by plants and algae to dissipate the extra light energy absorbed which can't be processed in photosynthesis. Xanthophylls take part, either directly or indirectly, in NPQ in photosystem II, through the xanthophyll cycle which exists in all land plants (Janik et al., 2016; Jahns and Holzwarth, 2012). As an antioxidant, zeaxanthin serves multiple roles in photoprotection of the photosynthetic apparatus, involving NPQ and the memory of photo-oxidative stress (Janik et al., 2016; Jahns and Holzwarth, 2012) Carotenoids can absorb light in the 450 – 550 nm range, which is the most abundant light reaching earth but which cannot be absorbed by major chlorophylls. Thus, carotenoids function as accessory light-harvesting pigments, extending the spectral range over which light can be absorbed by the photosynthetic apparatus (Green and Parson, 2003; Fromme, 2008).

In non-photosynthetic tissues, many carotenoids confer colour phenotypes, and serve as pigments in flowers to attract animal pollinators and in fruit to attract animal dispersors. They may be important in regulating the rate of post-harvest fruit softening because of their antioxidant capacity. In addition, carotenoids serve as precursors for the biosynthesis of two phytohormones, abscisic acid (ABA) and strigolactones, which are widely involved in the regulation of plant growth, development and stress responses (Walter and Strack, 2011).

Figure 1.1 Vegetables and fruits are rich sources of pigments, including carotenoid pigments. In this picture, the colour of tomatoes, pineapple, citrus, cucurbitaceae vegetables (such as melon and squash), and maize are due to accumulation of different kinds of carotenoids. Carotenoids are also present in green vegetables such as cucumber and runner beans. Apples, avocado, plums, strawberries and raspberries are coloured by anthocyanins. Image from the JIC image library, taken by Andrew Davis.

1.1.3 Plant sources of carotenoids

Carotenoids are produced in chloroplasts which are abundant in leafy vegetables and chromoplasts (particularly abundant in red, orange and yellow fruit and vegetables). The composition of carotenoids in chloroplasts is fairly constant but the carotenoid composition of chromoplasts differs widely between different plant species.

Vegetables are major sources of dietary carotenoids (Figure 1.1). Tomato (*Solanum lycopersicum*) ripening involves a number of physiological processes that include the visible breakdown of chlorophyll and build-up of carotenoids, with massive accumulation of antioxidant compounds such as lycopene and β-carotene within the plastids (Egea et al., 2010). At the red-ripe stage of fruit development, lycopene accumulates to a high level due to the differential expression of genes encoding enzymes involved in conversion of lycopene to other carotenoids (Hirschberg, 2001). It has been shown that ripe tomato fruits and relelated products provide approximately 85% of the lycopene found in the human diet (Canene-Adams et al., 2005). Due to the different types of carotenoids that can accumulate, tomato displays various colours in fruits, such as yellow, tangerine, orange, orange-red and red, with distinct carotenoid profiles. The fruit of the *yellow flesh* natural mutant (*r* locus) (Figure 1.2) displays a pale-yellow colour with undetectable levels of lycopene and very low β-carotene levels (Fray and Grierson, 1993). The *tangerine* (*t*) mutant accumulates pro-lycopene in fruit resulting in its tangerine colour (Isaacson, 2002a) (Figure 1.2). *Beta* and *Delta* mutants have fruits of orange and orange-red due to enhanced production of β-carotene and δ-carotene respectively, at the expense of lycopene (Ronen et al., 2000a, 1999). The fruit of *old-gold* has elevated accumulation of lycopene leading to a deep crimson colour, due to the abolition of the production of beta-carotene (Ronen et al., 2000a).

Pepper (*Capsicum annuum*) is another vegetable containing a wide range of carotenoids in fruit of different species. Differential accumulation of carotenoids gives pepper fruits various colours; yellow, orange and red. The yellow-fruited species predominantly produce

β-carotene and lutein and are devoid of capsanthin, a characteristic carotenoid highly accumulated in red-fruited varieties (Ha et al., 2007; Rodriguez-Uribe et al., 2012). A more complex carotenoid profile is found in orange pepper fruit, which exhibit a variety-dependent colouration, and are composed of all of the carotenoids mentioned in red or yellow varieties (Rodriguez-Uribe et al., 2012; Guzman et al., 2010). Generally, red-fruited varieties contain higher amounts of total carotenoids than the non-red ones (Ha et al., 2007).

Among other vegetables that accumulate carotenoids, sweet potato and carrot contain very high levels of carotenoids with β-carotene as the dominant form (Figure 1.1). Carrots synthesize diverse carotenoids leading to their edible parts being a range of yellow, orange and red colours, resulting from the accumulation of lutein, β-carotene and lycopene and β-carotene respectively (Rodriguez-Concepcion and Stange, 2013).

Cauliflower varieties carrying a dominant mutation, *Orange* (*OR*), produce dramatically increased levels of β-carotene compared to the white variety containing negligible amounts of carotenoids (Li et al., 2001). Neoxanthin and violaxanthin are abundant in dark leafy vegetables, such as spinach and kale, as well as broccoli.

Apart from vegetables, many fruit species are major sources of carotenoids. The predominantly red-fleshed watermelon accumulates lycopene as the major carotenoid in fruit. There are also white-, yellow- and orange-coloured watermelons containing β/ε-carotene, lutein, phytoene, violaxanthin or neoxanthin depending on different varieties (Lv et al., 2015). Similarly, melon fruit display various colours, such as white, green or orange (Figure 1.1), closely related to their carotenoid composition. Among different cultivars, the orange-fruit accumulates the highest amount of total carotenoids with β-carotene as the primary carotenoid (Burger et al., 2009). This dominant phenotype is determined by an SNP in the melon *Orange* gene, affecting multiple cellular processes including sugar and carotenoid metabolism (Chayut et al., 2015; Tzuri et al., 2015).

Wild Type
LA3802

Delta
LA2996A
(D)

Beta
LA0316
(B)

tangerine
LA0351
(t)

yellow flesh
(r) LA3417

Wild Type

Never ripe
(Nr)

Green ripe
(Gr)

ripening
inhibitor (rin)

Cnr showing
epigenetic
revertant
sector

Colourless-
non ripening
(Cnr)

Figure 1.2 Phenotypes of mutations affecting carotenoid production in tomato fruit.

The top row shows mutations in genes encoding enzymes of carotenoid biosynthesis. *Delta* is a dominant allele of *LCYE* increasing the levels of δ-carotene, *Beta* is a dominant allele of *LCYB* increasing the levels of β-carotene, *tangerine* (*t*) affects the activity of CRTISO and *yellow flesh* (*r*) affects the activity of phytoene synthase 1 (*PSY1*) in tomato fruit. These pictures are from the Tomato Genetic Resource Center and the numbers refer to stock centre lines carrying these mutations.

The second row shows ripening mutants, *Never ripe* (*Nr*) encodes an ETR1-like ethylene receptor (Wilkinson et al., 1995), *Green ripe* (*Gr*) encodes a protein of unknown biochemical function that influences ethylene responsiveness of fruit (Barry and Giovannoni, 2006), *ripening inhibitor* (*rin*) encodes a MADS-domain transcription factor that is a positive regulator of ethylene production and ripening of tomato fruit (Vrebalov et al., 2002).

The third row shows the mutant phenotype of *Colourless non-ripening* (*Cnr*) which encodes an SBP-like transcription factor which positively regulates ripening (Manning et al., 2006). The picture on the left shows epigenetic reversion of the mutant allele (Manning et al., 2006). The other pictures of ripening mutants are from Giovannoni, (2007).

Due to complexity of the genetic background in citrus, this genus contains the most diverse carotenoid profiles with the widest range of carotenoids found in fruit (Kato, 2012). For instance, pummelo and grapefruit are rich in phytoene, phytofluene and ζ/β-carotene while citrus fruits such as mandarin and orange predominantly accumulate other types of carotenoids, like violaxanthin, lutein, β-cryptoxanthin and zeaxanthin (Kato, 2012). There are other fruits that are also rich in carotenoids, including peach, papaya, mango and apricot.

Flowers are organs which often show high accumulation of carotenoids, which contribute particularly to the colour of flowers. The most abundant carotenoids present in flower petals are xanthophylls, as well as their epoxides, especially in yellow flowers, while orange flowers are mainly rich in carotenes (Zhu et al., 2010). Unlike fruits and vegetables, red colour is mostly dependent on the accumulation of lycopene, although in red Asiatic hybrid lily, the colour of petals is determined by other dark-coloured carotenoids, capsanthin and capsorbulin, which normally give orange to red colours depending on the composition and content (Yamagishi et al., 2010).

Because they are lipophilic, the bioavailability of dietary carotenoids can be influenced by a wide range of factors. The bioavailabilities of β-carotene and lycopene from papaya have been reported to be approximately 3-fold higher than the same compounds from tomatoes (Schweiggert *et al*., 2014). Even within the same species, natural variation and different cultivation practices can affect the composition and content of carotenoids in crops (Pott *et al*., 2003; Lenucci *et al*., 2006), which influence their bioavailability indirectly. Each step of food processing, from postharvest storage, thermal processing, to product storage, may also affect carotenoid stability and bioavailability (Maiani *et al*., 2009).

1.2 Carotenoid metabolic pathways

The massive accumulation of carotenoids in tomato fruit is associated with a very active endogenous isoprenoid biosynthetic pathway in the plastid (Figure 1.3). The red colour of

lycopene begins to be seen from the breaker stage when the ripening process is triggered by ethylene. Several mutants have been found which cause abnormal carotenoid accumulation with impaired ripening, such as *yellow flesh* (*r*), *old gold* (*og*), *tangerine* (*t*), *Never-ripe* (*Nr*), *Green-ripe* (*Gr*), *Colour-less non-ripening* (*Cnr*) and *ripening-inhibitor* (*rin*) mutants (Figure 1.2) (Fray and Grierson, 1993; Ronen et al., 2000a; Isaacson, 2002a; Barry and Giovannoni, 2006; Lanahan, 1994; Liu et al., 2004; Manning et al., 2006; Mustilli, 1999). The carotenoid biosynthetic pathway begins in the plastid with the formation of phytoene from geranylgeranyl diphosphate (GGPP) from the central isoprenoid pathway, called the 1-deoxy-D-xylulose-5-phosphate (DOXP) or the MEP pathway in contrast to the cytoplasmically-localised mevalonic acid pathway.

1.2.1 DOXP/MEP pathway

In higher plants, isopentenyl diphosphate (IPP), and its double-bond isomer dimethylallyl diphosphate (DAMPP) serve as building blocks for carotenoid biosynthesis (Chappell et al., 1995). Two distinct and compartmentalised pathways exist in plants for IPP and DAMPP production: the mevalonate (MVA) pathway is active in the cytosol and the 1-deoxy-D-xylulose-5-phosphate (DOXP) or methylerythritol 4-phosphate (MEP) pathway is active in plastids (Rodriguez-Concepcion, 2002; Lichtenthaler et al., 1997). Biosynthesis of plant carotenoids takes place in plastids, and is dependent on the precursors produced through DOXP/MEP pathway (Milborrow and Lee, 1998). The DOXP/MEP pathway supplies precursors not only for carotenoid biosynthesis, but also for the synthesis of gibberellins, tocopherols, isoprenes, diterpenes and monoterpenes, reflecting its central role in the terpenoid secondary metabolic network (Rodríguez-Concepción, 2010).

The DOXP/MEP pathway is intitiated by condensation of pyruvate and glyceraldenhyde 3-phosphate to form deoxy-D-xylulose 5-phosphate (DXP), catalysed by DXP synthase (DXS), which is considered to be the first rate-determing step in the regulation of flux in the DOXP/MEP pathway in plants (Lichtenthaler, 1999; Carretero-Paulet et al., 2006; Estévez et al., 2001) (Figure 1.3, Figure 3.4 B). This initial catalytic step is performed in a tissue- and developmental stage- specific manner. There are three *SlDXS* genes (*Solyc01g067890*,

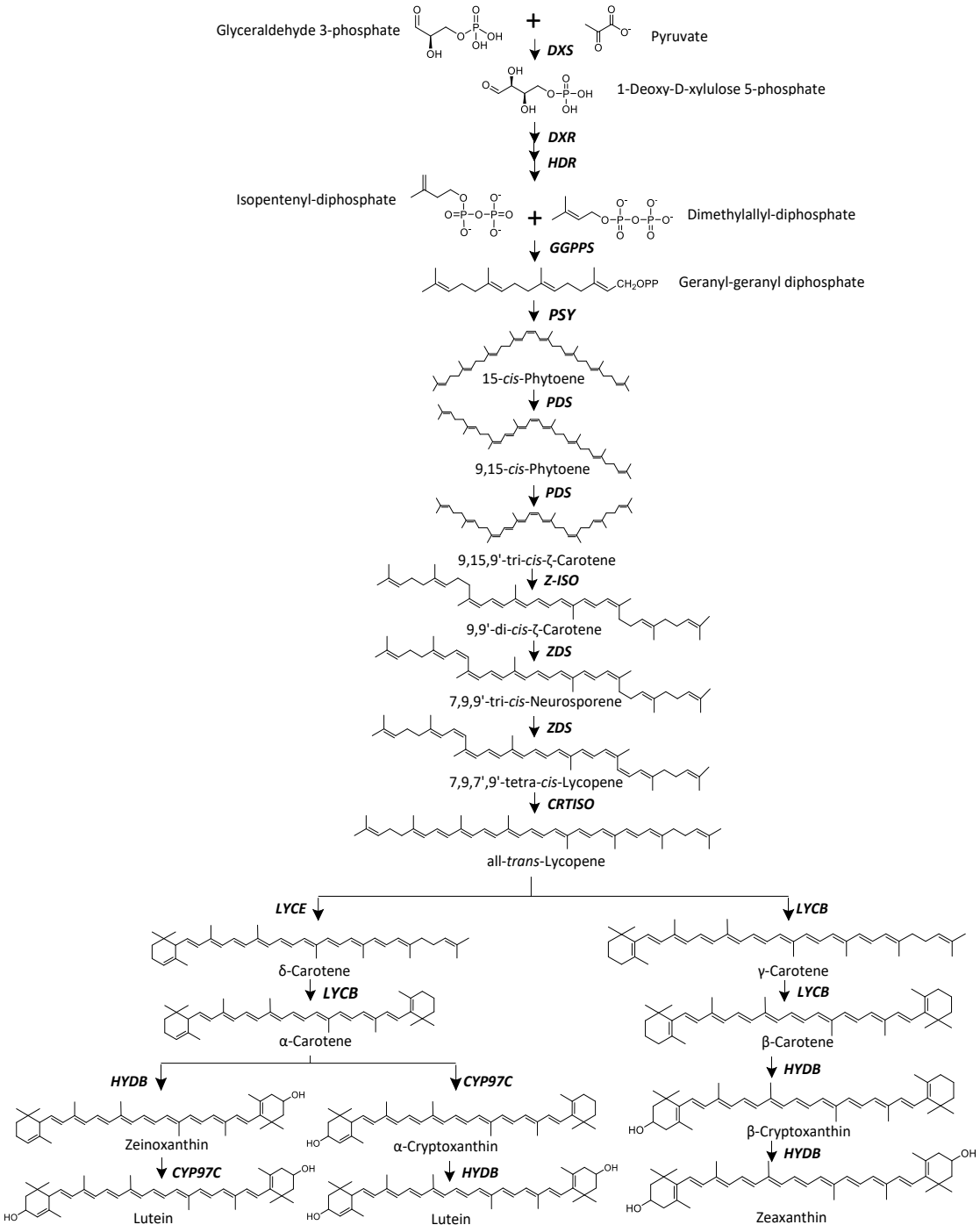


Figure 1.3 The carotenoid biosynthetic pathway in plants. DXS, 1-deoxyxylulose-5-phosphate synthase; DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; HDR, 1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate reductase; GGPPS, geranyl-geranyl diphosphate (GGPP) synthase; PSY, phytoene synthase; PDS, phytoene desaturase; ZDS, ζ -carotene desaturase; Z-ISO, ζ-carotene isomerase; CRTISO, carotenoid isomerase; LYCB, lycopene β-cyclase; LYCE, lycopene ε-cyclase; CYP97C, carotene ε-ring hydroxylase; HYD-B, β-Carotene hydroxylases. Redrawn from Zhu *et al*., (2010). This picture was reused from Martin and Li (2017) (Appendix 5), with the copyright license number 4431440777168 (Appendix 6).

Solyc08g066950, *Solyc11g010850*) annotated in the tomato genome (Tomato Genome Consortium, 2012). Studies have showed that expression of *SlDXS1* (*Solyc01g067890*) is positively correlated with the formation of lycopene during fruit ripening (Lois et al., 2000), whereas the transcripts of *Solyc11g010850* (*SlDXS2)*, are abundant in young leaves, trichomes and petals but not in fruit (Paetzold et al., 2010, Tomato eFP Browser). A third *DXS* gene (Solyc08g066950) has been proposed, but its activity and functions have not yet been fully elucidated. DXP is then converted to Methylerithritol 4-phosphate (MEP) via an intermolecular rearrangement followed by a reduction catalysed by DXP reductoisomerase (DXR). Subsequently, MEP is converted into hydroxymethylbutenyl diphophate (HMBPP) through four catalytic reactions, via intermediates cytidine diphosphomethylerythritol (CDP-ME), CDP-ME 2-phosphate (CDP-MEP), and methylerythritol 2,4-cyclodiphophate (ME-cPP), generated by 4-(cytidine 5'-diphospho)-2-C-methyl-D synthase (CMS), 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase (CMK), ME-cPP synthase (MCS) and HMBPP synthase (HDS). Eventually, HMBPP is converted to isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) by HMBPP reductase (HDR). Similarly, DXS, DXR and HDR are also regarded as flux-limiting enzymes in DOXP/MEP pathway controlling the production of the DOXP/MEP-derived precursors for plastid isoprenoid biosynthesis (Botella-Pavía et al., 2004; Seetang-Nun et al., 2008; Kim et al., 2009). IPP and DAMPP can be interconverted by isopentenyl diphosphate isomerase (IPI). Geranylgeranyl diphosphate (GGPP) synthase (GGPPS) contributes to the formation of the isoprenoid precursor of carotenoid biosynthesis, GGPP, by condensing IPP and DMAPP. There are two functionally-characterised *GGPPS* genes, *SlGGPPS1* and *SlGGPPS2*, in tomato. *SlGGPP2* functions specifically in chromoplasts in flowers and fruits, while *SlGGPP1* is primarily expressed in leaves (Ament et al., 2006).

1.2.2 Lycopene biosynthesis

Two molecules of GGPP are condensed in a head to tail manner to from a colourless compound, phytoene, which is probably the most important step in the carotenoid biosynthesis pathway (Cunningham and Gantt, 1998). This two-step reaction is catalysed by the enzyme phytoene synthase (PSY) (Figure 1.3, Figure 3.5). Tomato has two paralogs of

the *PSY* gene with characterised functions, *SlPSY1* (Solyc03g031860) and *SlPSY2* (Solyc02g081330). The *SlPSY1* gene encodes a fruit-ripening-specific isoform, whilst *SlPSY2* predominates in green tissues, including mature green fruit and has no role in carotenogenesis in ripening fruit (Bartley and Scolnik, 1993; Fraser et al., 1999). A third *PSY* gene in tomato, *SlPSY3* (Solyc01g005940), has been found in the tomato genome, and is predicted to encode a PSY protein controlling carotenoid biosynthesis in root tissues responding to abiotic stress, which is in line with the induced production of carotenoid-derived hormones, ABA and strigolactones, in response to stressful conditions. (Li et al., 2008a). Phytoene undergoes a series of four desaturation reactions leading to phytofluene, ζ-carotene, neurosporene and finally lycopene. These four sequential desaturations are catalysed by two related enzymes in plants; phytoene desaturase (PDS) and ζ-carotene desaturase (ZDS). A carotenoid isomerase (CRTISO) activity is additionally required to transform the poly cis-lycopene (pro-lycopene) to the all trans-isomers.

Several mutants in tomato affecting genes underlying lycopene biosynthesis have been identified. *SlPSY*, encoding the enzyme controlling the key point in carotenoid biosynthesis, has been extensively studied since the identification of loss-of-function mutant of *SlPSY1*, *yellow flesh* (locus *r*) (Figure 1.2), resulting in fruit devoid of lycopene accumulation in fruit (Fray and Grierson, 1993). The fact that transgenic silencing of *SlPSY1* led to a very similar yellow-coloured fruit phenotype when fully ripe with only 3% total carotenoids left (Ray et al., 1992), and a *SlPSY1* knock-out mutant, identified by Targeting Induced Local Lesions IN Genomes (TILLING), had yellow flesh with undetectable levels of carotenoids (Gady et al., 2012), further confirmed the tissue- and development-specific role of SlPSY1 in the regulation of carotenoid production in tomato fruit. CRTISO was identified through map-based cloning of the locus *t* which is affected in the *tangerine* mutant, which has orange fruits, pale-yellow flowers, and yellowish young leaves because of the over-accumulation of prolycopene instead of lycopene (Figure1.2) (Fantini et al., 2013). Recently, two functional *CRTISO* genes were identified in tomato, as well as in Arabidopsis and grape, indicating possible competing steps in the metabolic pathway towards all-*trans*-lycopene. The CRTISO step is followed by the branch point in biosynthesis to form lutein or β-carotene from lycopene (Fantini et al., 2013) (Figure 1.3, Figure 3.5). In a virus- induced gene silencing study, three metabolic units involved in lycopene accumulation were characterised in

tomato fruit, which contain *SlPSY1*, *SlPDS*/*SlZISO* and *SlZDS*/*SlCRTISO*, controlling the biosynthesis of intermediates 15-*cis*-phytoene, ,9,9'-di-*cis*-ζ-carotene and all-*trans*-lycopene respectively (Fantini et al., 2013), indicating that there is cross talk underlying lycopene accumulation in tomato fruit.

1.2.3 Lycopene catabolism and carotenoid derivative biosynthesis

Lycopene accumulation in tomato fruit is the outcome of both its increased biosynthesis and reduced degradation. Cyclization of lycopene serves as a branch point in the carotenoid biosynthetic pathway, largely contributing to the diversity of carotenoids, but is almost completely switched off in ripe fruit of *S.lycopersicum*, although this pathway remains active in some wild relatives, such as *S. pennellii*, with green fruit. There are two pathways which branch from lycopene, distinguished by the presence of different cyclic groups at the ends of the products of cyclization of lycopene; either with addition of a β-ring and/or an ε-ring. One route leads to the production of β-carotene, xanthophylls and precursors for carotenoid derivatives, such as ABA and strigolactones. The other is responsible for the production of α-carotene and lutein. In detail, lycopene β-cyclase (LCYB) catalyses the conversion of lycopene to produce β-carotene in a two-step reaction that creates one β-ionone ring at each end of the lycopene molecule. In the other pathway branch, δ-carotene is produced by the addition of one ε-ring to lycopene in the presence of lycopene ε-cyclase (LCYE) (Cunningham and Gantt, 2001; Ronen et al., 2000). There are three LCYB genes annotated in tomato genome (Tomato Genome Consortium, 2012). *SlCYCB* (*Solyc06g074240*) was characterised by mapped-based cloning in tomato colour mutants, *old-gold* (*og*) and *Beta* (Ronen et al., 2000a) (Figure1.2). Similarly, *SlLCYE* (*Solyc12g008980*) was identified from a dominant mutant *Delta*, in which the expression level of *SlLCYE* was more than 30-fold increased accompanied by elevated levels of δ-carotene (Ronen et al., 1999) (Figure1.2).

α-Carotene and β-carotene may then be hydroxylated further to form xanthophylls, typically by the addition of hydroxyl, epoxy or keto groups, which are the major carotenoids

that play important roles in photosystems in green tissues. Generally, two types of carotenoid hydroxylase (CHY) have been identified with distinct evolutionary backgrounds controlling xanthophylls biosynthesis in plants. One group contains ferredoxin-dependent non-heme β-ring hydroxylases (BCH type) and the other group are composed of P450-type (CYP97 family) β/ε-ring hydroxylases. There are two and three CHY enzymes belonging to the BCH family and CYP97 family respectively, working in a tissue-specific manner (Pogson, 1996; Tian and Dellapenna, 2001). The divergent functions of the two types of CHYs exist in tomato as well, as characterised by genetic and phylogenetic analysis. *SlBCH2*, encoding a chromoplast-specific BHC enzyme, was identified by genetic mapping of the locus controlling the phenotype of the *white flower* (*wf*) mutant. In contrast, *SlBCH1* is predominantly expressed in green tissues (Galpaz, 2006). Hydroxylases in the CYP97 family have also been identified in tomato, including CYP97C11 and CYP97A29, genes of which have very similar expression patterns across the plant, except in roots and ripening fruits. CYP97A29 is more active or more highly expressed than CY97C11 (Stigliani et al., 2011).

Zeaxanthin, generated by hydroxylation of β-carotene, can be further metabolised to violaxanthin, catalysed by zeaxanthin epoxidase (ZEP). The dynamic composition of xanthophylls is regulated in response to chloroplast redox status and light intensity, through the xanthophyll cycle, which is involved in photoprotection of the photosynthetic apparatus under light stress (Jahns and Holzwarth, 2012; Niyogi et al., 1998). In response to strong light, zeaxanthin can be converted to violaxanthin via antheraxanthin as an intermediate, and this process can be reversed by violaxanthin deepoxidase (VDE). In the tomato *hp3* mutant, carrying a loss-of-function mutation in *SlZEP*, ABA levels are significantly reduced compared to wildtype, suggesting a crucial role of this enzyme in controlling the metabolic flux to ABA (Galpaz et al., 2008). In contrast, under high-intensity light, overexpression of VDE can alleviate the photoinhibition of photosystems as a result of a more active xanthophyll cycle. (Han et al., 2010). Violaxanthin can be converted to neoxanthin by neoxanthin synthase (NSY). This step is catalysed by NSY encoded by the *ABA4* gene in Arabidopsis, and three homologous proteins with putative NSY activities have been identified in tomato, the functions of which need further investigation (North et al., 2007). Neoxanthin, as well as its precursor, violaxanthin, can be cleaved by 9-*cis*-epoxycarotenoid dioxygenase (NCED) and further metabolised to form ABA. In tomato, there are three *NCED*

genes annotated in the genome (Tomato Genome Consortium, 2012). Apart from NCEDs, carotenoid cleavage dioxygenases (CCDs) are also responsible for the turnover of carotenoids and production of apocarotenoids, such as strigolactones and β-citraurin. There are seven *CCD* genes in the tomato genome based on their sequence similarity of Arabidopsis *CCD*s (Tomato Genome Consortium, 2012). However, the precise roles of different CCDs in tomato still needs to be elucidated.

1.3 Tomato fruit ripening and over-ripening

Based on the requirement for ethylene production for the initiation of fruit ripening, fleshy fruit are classified into two subgroups, climacteric and non-climacteric. The initiation of climacteric fruit ripening is concurrent with the increased production of ethylene, as well as increased respiration, and includes fruits of tomato, apple, avocado and banana, for which ethylene synthesis is essential and a block in ethylene production or signalling normally results in a delay in ripening (Klee and Giovannoni, 2011). In contrast, for nonclimacteric fleshy fruit such as grape and strawberry, ethylene is not necessary for the initiation of ripening and there is no change in respiration observed at this stage (Klee and Giovannoni, 2011). Exogenous application of ethylene can initiate and accelerate ripening of climacteric fruits, but this does not work for nonclimacteric fruits (Klee and Giovannoni, 2011).

Many species in the Solanaceae family serve important roles in the human diet, including tomatoes (*Solanum lycopersicum*), eggplants (*Solanum melongena*), potatoes (*Solanum tuberosum*) and peppers (*Capsicum annuum*). Tomato is one of the most popular vegetables in the human diet, with an average production of 152 million tons per year according to data from the Food and Agriculture Organization (FAO, 2010). As a typical climacteric fruit, with a strict requirement for ethylene production for fruit ripening, tomato is the most studied model for fleshy fruit development and maturation, as well as ethylene biosynthesis and signalling (Alexander, 2002; Giovannoni, 2004). Tomato fruit ripening is a dynamic and complex developmental program involving the coordination of numerous metabolic pathways influencing morphology, flavour, colours, aroma and texture, many of which

attributes enhance fruit nutritional value and attractiveness, thereby promoting seed dispersal and consumption (Goff & Klee 2006).

1.3.1 Ripening of tomato fruit

Tomato fruit ripening is complex and dynamic process involving changes in a wide range of physiological processes, such as ethylene production and perception (Seymour et al., 2013), shifts in cell wall composition (Rose et al., 2004; Scheible and Pauly, 2004), chloroplast to chromoplast transition (Beyer et al., 1994), as well as changes in morphology and metabolism (Klee and Giovannoni, 2011).

1.3.1.1 Ethylene biosynthesis and signalling

The ripening of tomato is entirely dependent on ethylene production and perception. Ethylene is, chemically, the simplest plant hormone, but plays central roles in the regulation of diverse biological processes, including responses to biotic and abiotic stresses, fruit development and ripening, as well as organ abscission (Klee and Giovannoni, 2011). Therefore, the synthesis and perception of ethylene is regulated precisely by the coordination of different factors across multiple levels. Ethylene is synthesised in two steps. 1-Aminocyclopropane-1-carboxylate (ACC) is generated by ACC synthase (ACS) using S-adenosylmethionine as a substrate, and is subsequently converted to ethylene by ACC oxidase (ACO) (Lieberman and Kunishi, 1966). The two enzymes involved in ethylene production, ACS and ACO, are encoded by multi-gene families (Rottmann et al., 1991). In tomato, eight *ACS* genes have been characterised, each of which functions in a tissue- and/or stimuli-specific pattern, and four *ACO* genes have been identified (Barry et al., 1996; Nakatsuka et al., 1998; Yip et al., 1992; Barry et al., 2000). The formation of ACC by ACS is usually the rate-limiting step, although certain ACOs are ethylene-inducible and the silencing of *ACO1*, the most active *ACO* gene, inhibits ethylene production and the following ripening of fruit (Barry et al., 1996; Hamilton et al., 1990). During ripening, two *ACS* genes, *SlACS2* and *SlACS4*, are significantly induced, indicating that the expression of ACS is a key point in the regulation of ethylene biosynthesis (Barry et al., 1996).

After synthesis, ethylene is perceived by ethylene receptors (ETRs), a group of copper-binding membrane-associated proteins, which function as negative regulators of ethylene signalling, located in the endoplasmic reticulum. In the absence of ethylene, ETRs are in the 'on' state, and inhibit subsequent ethylene signalling. They are de-activated by binding ethylene to allow downstream signalling processes (Cherian et al., 2014; Klee and Giovannoni, 2011). There are seven *ETR* genes identified in tomato, *SlETR1*, *SlETR2*, *NR* (*SlETR3*), *SlETR4*, *SlETR5*, *SlETR6* and *SlETR7* (Lashbrook et al., 1998; Tieman and Klee, 1999; Wilkinson et al., 1995), among which *SlETR4*, *Nr* (*SlETR3*) and *SlETR6* are significantly transcriptionally induced during ripening (Kevany et al., 2007).

ETRs interact with the constitutive triple response 1 (CTR1) protein, a Raf kinase-like protein, which works as an inhibitor of ethylene responses, and this repression is relieved when CTR1 binds to ETRs to induce the signalling (Seymour et al., 2013; Zhu et al., 2015; Klee and Giovannoni, 2011). At least three genes encoding CTR1 have been identified in tomato (Adams-Phillips et al., 2004). A set of transcription factors process the signalling at the downstream of CTR1, which I will discuss in the section on transcriptional regulation of fruit ripening (See 'Chapter 1, 1.3.2 Transcription factors controlling fruit ripening').

1.3.1.2 Metabolic changes during tomato fruit ripening

During tomato fruit ripening, metabolite levels change dramatically. Among primary metabolites, the levels of nearly all sugars increase during ripening, including fructose, glucose, mannose, maltose and others, among which glucose and fructose are the two most abundant sugars in ripe fruit, constituting 2%-4% of the fruit fresh weight (Carrari and Fernie, 2006; Lee et al., 2012). Although amino acids do not show profound changes during ripening, some amino acids are reduced, such as asparagine, alanine, arginine and proline, while others increase, like lysine methionine, cysteine. The levels of all the intermediates of the TCA cycle are also decreased (Carrari and Fernie, 2006). However, the regulatory mechanisms underlying these changes are not well understood.

During fruit ripening, chloroplasts are converted to chromoplasts, chlorophylls are broken down and carotenoids and their derivatives start to accumulate in ripening fruit, leading to the colour change from green through orange to red when fully ripe. Up to the breaker stage, 85% of pigments in tomato fruit are chlorophylls and the composition of carotenoids is very similar to that of photosynthetic tissues like leaves. Lycopene and β-carotene begin to accumulate as the ripening process is triggered by ethylene. The changes in the colour of ripening fruit are mainly dependent on the dynamic composition of chlorophyll and carotenoid pigments (Klee and Giovannoni, 2011; Carrari and Fernie, 2006).

Flavonoids are another group of pigments in tomato fruit, which are synthesized via the phenylpropanoid pathway. Flavonoids are initially produced in the epidermis of fruit and then transported to the cuticle during ripening, where naringenin, naringenin chalcone, rutin and kaempferol are the most abundant flavonoids (Laguna et al., 1999; Mintz-Oron et al., 2008). The yellow colour contributed by flavonoids is normally obscured by carotenoids, but in '*pink*' tomatoes, which are popular in Asia, a mutation in the *y* gene encoding the SlMYB12 transcription factor that controls flavanol biosynthesis, results in loss of flavonoids in the fruit and a change in fruit colour (Adato et al., 2009; Ballester et al., 2010).

Apart from the metabolite changes mentioned above, the abundance of compounds that contribute to the flavour and smell of tomato fruit, also changes dramatically during ripening, such as the saponin α-tomatine which declines in ripe fruit and volatiles associated with fruit flavour which increase in ripe fruit (Klee and Giovannoni, 2011).

1.3.2 Transcription factors controlling fruit ripening

As a typical climatric plant, tomato fruit ripening is triggered and controlled by ethylene signalling, the molecular basis of which has been discussed in Section 1.3.1, as well as the transcription factors involved. Recently, with the collection and identification of mutants affecting fruit ripening, additional regulatory mechanisms coordinating ethylene biosynthesis and perception have been identified. Serval mutants exhibiting pleiotropic

phenotypes have been identified, such as ripining-*inhibitor* (*rin*), *Never-ripe* (*Nr*), *Colour non-ripening* (*Cnr*) and *Green-ripe* (*Gr*) (Figure 1.2). RIN (ripening inhibitor), a MADS-box transcription factor acts as a master regulator of fruit ripening. In the same way as for many transcription factors in this family, RIN has been shown to interact with the promoters of *ACS2* and *ACS4*, the ethylene biosynthetic genes, by direct binding to CArG motifs (Martel et al., 2011). *RIN* is expressed specifically in fruit tissues from the breaker stage, and has been demonstrated to regulate carotenoid accumulation by interacting directly with the *SlPSY1* promoter (Zhong et al., 2013; Fujisawa et al., 2013). *rin*, the corresponding mutant of *RIN*, failed in ethylene production and lycopene accumulation in its fruits, which have yellow colour (Figure 1.2) (Giovannoni et al., 1995; Vrebalov et al., 2002). Other MADS-domain transcription factors have been identified which regulate the ripening process. *Tomato AGAMOUS-like 1* (*TAGL1*), is highly expressed during carpel development and at the onset of ripening. *TAGL1* knock-down lines generated by RNAi show inhibition of ethylene biosynthesis and carotenoid accumulation, as well as reduction in carpel thickness (Itkin et al., 2009; Pan et al., 2010). FRUITFULL 1 and 2 (FUL1 and FUL2) have been found to regulate positively the ethylene signalling pathway during ripening following characterisation of the fruits of FUL1/2-suppressed mutants (Vrebalov et al., 2009; Shima et al., 2014). In a yeast two-hybrid assay, TAGL and FUL1/2 interact with RIN, suggesting that the MADS-domain transcription factor works in a combinatorial manner in the regulation of fruit ripening (Martel et al., 2011; Liu et al., 2015a). *Cnr* encodes an SPB transcription factor, functional downstream of RIN in the network controlling ethylene production and fruit ripening, but recruitment of RIN to target loci is dependent on the presence of Cnr which alters the promoter methylation of RIN's target genes (Martel et al., 2011; Manning et al., 2006). AP2a, a homolog of the Arabidopsis APETALA2, has been identified recently as a negative ripening regulator, and functions downstream of RIN and CNR (Chung et al., 2010; Karlova et al., 2011; Fujisawa et al., 2013). The transcription factors involved in ethylene biosynthesis signalling have been discussed in Section 1.3.1: Ethylene biosynthesis and signalling.

1.3.3 Tomato fruit over-ripening

Unlike ripening, over-ripening (senescence) of tomato fruit has not been investigated very extensively, but is important to the plant as well as for its impact on commercial value. There is no agreed definition of over-ripening of tomato fruit, but generally, over-ripening is defined as programmed senescence of the fruits, in which fully ripe fruit begin various physiological processes which can shorten the shelf life. Several symptoms can be observed during over-ripening, such as fruit softening, cell wall degradation, increased susceptibility to postharvest pathogens and declining activity of antioxidants (Brummell and Harpster, 2001; Shah et al., 2012; Jimenez et al., 2002). As fruit over-ripen, there are broad changes in metabolite profiles. As discussed above, carotenoid accumulation is closely related to fruit ripening, which reaches the maximum before the initiation of over-ripening when levels decline. Therefore, the transcriptional regulation of carotenoid biosynthesis discussed in this thesis does not extend to over-ripening.

1.4 Transcription factors

Transcription factors (TFs) are a group of proteins regulating gene expression that work as enhancers or repressors of transcription of their target genes. TFs, by definition, bind to DNA through specific binding motifs, which are often, although not exclusively, located upstream of the target gene coding sequences. Transcriptional activators may bind Upstream Activator Sequences (UAS) in the promoters of their target genes. Repressors bind to Repressor motifs, usually also located in the promoter regions of target genes. Both types of TF bind DNA through conserved DNA binding domains, which allow TFs to be grouped into specific families. Those conserved across eukaryotes include MYB (myoblastosis), bHLH (basic helix-loop-helix), HD (homeodomain), bZIP (basic leucine zipper), MADS (from the founding members of this family: MCM1 from *Saccharomyces cerevisiae*, AGAMOUS from *Arabidopsis thaliana*, DEFICIENS from *Antirrhinum majus*, SRF from human). The regulatory activity of TFs is achieved by direct or indirect interaction with the basal transcriptional machinery that modulates the rate of transcriptional initiation by RNA Polymerase II.

Plant growth and development is an extremely dynamic process involving responses to a wide range of internal and external stimuli, which requires timely and accurate regulation of gene expression. It has been shown that transcriptional regulation of gene expression plays an essential role in controlling the signalling cascades governing physiological processes. TF-encoding genes constitute 6% - 10% of any plant genome according to different databases (Libault et al., 2009; Riechmann et al., 2000).

1.4.1 Transcription factors in plants

Transcription factors (TFs) play diverse and important roles throughout plant development and in response to various internal and environmental signals, making it crucial to characterise the functions of transcription factors to understand the mechanisms controlling development and physiological changes (Lyzenga and Stone, 2012; Singh et al., 2002; Cominelli et al., 2010). From the pioneering work on TF characterisation in Arabidopsis, the number of transcription factors identified has risen to ~1500 to ~2000 (Riechmann et al., 2000; Davuluri et al., 2003; Guo et al., 2005; Hermoso, 2004; Riaño-Pachón et al., 2007). Similar studies are progressing in other plant species, such as rice, tobacco and maize, from which around 2500, 2500 and 2300 TFs respectively, have been identified (Caldana et al., 2007; Rushton et al., 2008; Jiang et al., 2012). The number of genes encoding TFs in *Drosophila*, *C. elegans* and human are much lower than in plants (Riechmann et al., 2000; Lambert et al., 2018). In Arabidopsis, plant-specific TF families comprise approximately 50% of the total TFs identified, and include members of the SBP (SQUAMOSA-promoter Binding Protein), TCP (Teosinte branched1/Cincinnata/proliferating cell factor), WRKY, AP2-EREBP (APETALA12-ETHYLENE RESPONSIVE ELEMENT BINDING PROTEINS), ABI3-VP1 (ABA INSENSITIVE3-VIVAPARIOUS1), Dof (DNA binding with one finger), EIL (Ethylene-Insensitive 3 Like), NAC, and YABBY families (Riechmann et al., 2000). The other TFs belong to families found also in other eukaryotes and include members of the MYB, bHLH, bZIP, HD and MADS families. Most TFs act as enhancers or repressors to regulate the expression of their target genes by binding to specific motifs in the promoter region. The characterisation of DNA binding motifs for 63 TFs has been undertaken for Arabidopsis using a combination of different approaches (Franco-Zorrilla et al., 2014).

1.4.2 Transcription factor complexes in plants

One of the best studied TF complexes is the MYB-bHLH-WD40 (MBW) complex controlling the expression of biosynthetic genes involved in the flavonoid pathway (Gonzalez et al., 2008). In this complex, a WD40 protein, for example, the Transparent Testa Glabra1 (TTG1) protein forms a complex with bHLH proteins and R2R3-MYB proteins to regulate the expression of genes encoding enzymes in the anthocyanin and proanthocyanidin pathways (Koes et al., 2005). Members of the MADS-box TF family form homo- and heterodimers (Pellegrini et al 1995) and even larger complexes have been suggested to ensure regulation of floral organ specificity (Immink et al., 2009; Bartlett, 2017). Heterodimerization is also necessary for the interaction of Auxin Response-Factors (ARFs) with the auxin/indole-3-acetic acid (AUX/IAA) proteins. In this example, when short lived AUX/IAA proteins are degraded (following exposure to auxins), ARF proteins are released and able to bind to specific DNA regions and regulate the expression of auxin target genes (Ulmasov et al., 1999).

1.4.3 Transcriptional regulation of plant metabolism

Usually, the biosynthesis and accumulation of metabolites occurs in developmental stage- and/or tissue- specific patterns. For example, the induction of the carotenoid biosynthetic pathway in tomato fruit is closely related to fruit ripening and carotenoid profiles in photosynthetic green tissues, such as leaves and immature fruits, are very distinct from those in flowers and ripe fruits. The spatial and temporal induction of metabolites is under transcriptional regulation by transcription factors. TFs from many transcription factor families have been identified in plants including bHLH, MYB, AR2/ERF, and WRKY families. bHLH transcription factors represent a group of proteins that participate a wide range of biological processes in plants, including regulation of metabolic pathways.

One widely-investigated transcription factor family involved in plant metabolism is the MYB transcription factor family. The production of many secondary metabolites are under the

control of MYB transcription factors, including anthocyanin, proanthocyanidin, glucosinolate, and phenylpropanoid metabolites (Teng, 2005; Gonzalez et al., 2008; Butelli et al., 2012; Nesi et al., 2001; Celenza, 2005). In Arabidopsis, TT2 controls the accumulation of proanthocyanidins in seed coats in combination with a bHLH TF and a WD40 protein by regulating *BANYULS* expression. *BANYULS* encodes a leucoanthocyanidin reductase (Nesi et al., 2001). Several MYB transcription factors have been identified that activate anthocyanin biosynthesis, including AtMYB75, AtMYB90, AtMYB113 and AtMYB114 (Borevitz et al., 2000; Gonzalez et al., 2008; Stracke et al., 2001). Again, these MYB proteins interact with bHLH TFs and WD40 proteins in a regulatory complex.

In Arabidopsis, the TT2/TT8/TTG1 complex fine-tunes biosynthesis of proanthocyanidins in seeds (Baudry et al., 2004). In petunia the AN2/AN1/AN11 complex regulates anthocyanin accumulation in the corolla of flowers, by controlling expression of the biosynthetic genes, DFR and CHSJ (Quattrocchio, 1993; Spelt et al., 2002).

1.5 Regulation of carotenoid metabolism in tomato

1.5.1 Transcriptional regulation of carotenoid biosynthesis

Several transcription factors have been identified that function in the transcriptional regulation of carotenoid accumulation in tomato fruit, most of which are associated with indirect regulation through their control of fruit ripening (Figure 1.4). Ethylene Response Factor 6 (SlERF6), has been shown to act as a negative regulator of carotenoid biosynthesis, and reducing its expression in tomato fruit leads to enhanced carotenoid accumulation and elevated ethylene levels, accompanied by up-regulated expression of *SlDXS1* and several ethylene biosynthetic genes, including, *SlACO1*, *SlACO2* and *SlACS2* (Lee et al., 2012). Another ERF gene, *SlERF.B3* shows expression linked to fruit development and it plays a crucial role in the regulation of several ethylene-related genes (Liu et al., 2013). Dominant suppression of *SlERF.B3* results in significantly decreased accumulation of lycopene and its precursors, with dramatically reduced transcript levels of *SlPSY1* and *SlPDS* (Liu et al., 2014b).

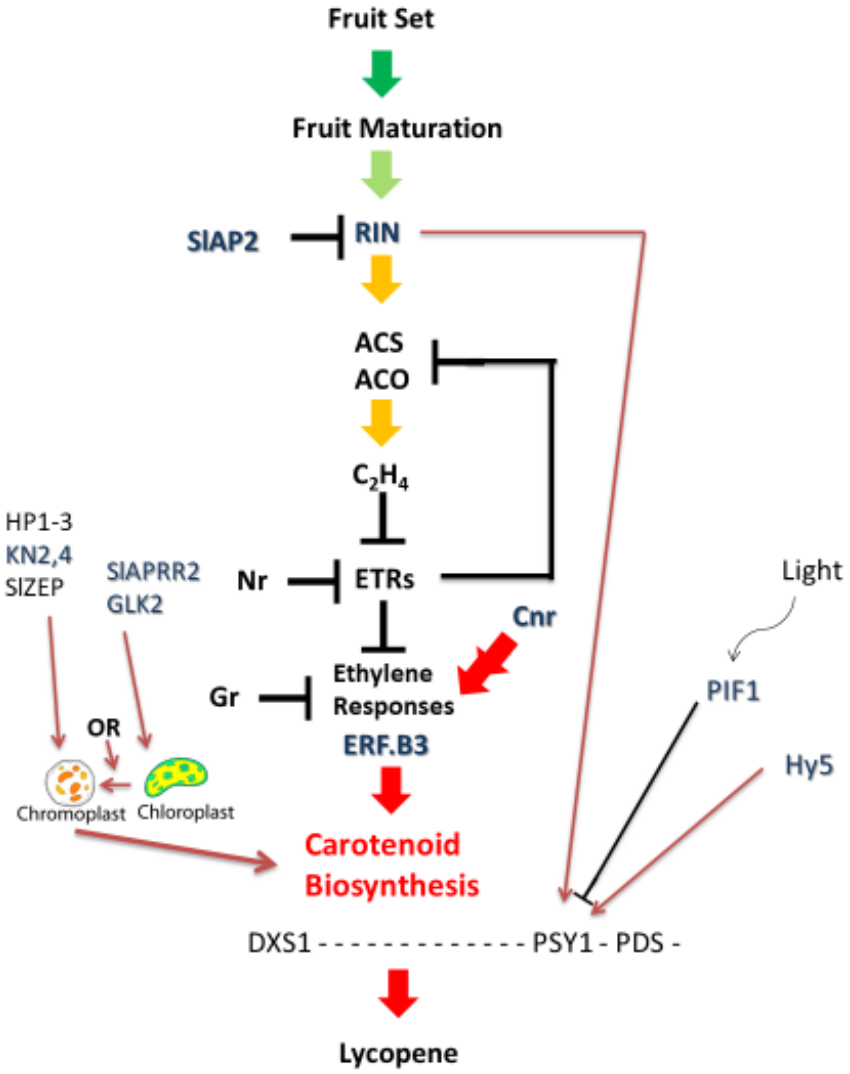


Figure 1.4 Diagram outlining the transcriptional control of fruit ripening and lycopene biosynthesis in tomato, based on the current literature.

The regulatory relationships between transcription factors operational in controlling fruit ripening, ethylene biosynthesis and lycopene biosynthesis are illustrated. Transcription factors are indicated in dark blue lettering. Never ripe (Nr) and Green ripe (Gr) are involved in ethylene perception as are ETRs.

Cnr encodes an SBP transcription factor, *RIN* encodes a MADS domain transcription factor, *SlAP2* and *ERF.B3* encode ethylene response factor/AP2 type transcription factors, *Hy5* encodes a bZIP transcription factor, *SlAPRR2* encodes a Pseudo-Response Regulator, *KN2* and *KN4* encode homeobox transcription factors, *PIF1* encodes a bHLH transcription factor responsive to light, *GLK2* encodes a member of the Myb superfamily of transcription factors.

The colours of the arrows, dark green, green, yellow and red, represent the fruit colours, indicating the developmental stages of tomato fruits.

However, in *SlERF.B3* dominant repression mutants, the opposite effects were observed on ethylene production in an earlier study (Liu et al., 2013, 2014b). Overexpression of *SlAP2a*, which encodes a negative regulator of ethylene production, impacted fruit pigmentation during ripening resulting in significantly decreased expression of *SlPSY1*, reduced lycopene and elevated β-carotene accumulation and an orange-fruit phenotype (Chung et al., 2010). These three transcription factors all belong to the AP2 family of transcription factors and function in regulating carotenoid accumulation through their control of ethylene production and the resulting ripening processes. There is no evidence to suggest that they regulate the expression of the structural genes of carotenoid biosynthesis through direct binding.

Phytochrome-interacting Factor 1 (PIF1) is a helix-loop-helix transcription factor involved in light-related responses. It has been demonstrated that in Arabidopsis, PIF1 represses expression of *PSY* through direct binding, leading to reduced production of carotenoids under shade conditions (Toledo-Ortiz et al., 2010). SlPIF1a is a PIF1 homolog in tomato, which negatively regulates carotenoid production in tomato fruit by specifically and directly interacting with *SlPSY1* by binding to a PBE box in its promoter (Llorente et al., 2016). Another transcription factor involved in light signalling, Long Hypocotyl 5 (HY5), works antagonistically with PIF1 in response to light signalling in Arabidopsis (Bou-Torrent et al., 2015). Down-regulation of *SlHY5* in tomato results in inhibited seedling photomorphogenesis and decreased carotenoid accumulation (Liu et al., 2004). Several other transcription factors involved in light signalling influence carotenoid biosynthesis in tomato fruit (Davuluri et al., 2005, 2004; Liu et al., 2004).

Overexpression of the *SlAPRR2-like* gene (*Arabidopsis Pseudo-Response Regulator 2-like*) under the control of the CaMV 35S promoter resulted in increased carotenoid content by affecting plastid number and size in tomato fruit (Pan et al., 2013). *Golden2-Like* (*SlGLK2*), which encodes a member of the Myb transcription factor superfamily, promoted the development of chloroplasts by regulating the expression of related chloroplast genes. Overexpression of *SlGLK2* increased the transcript level of genes encoding proteins functioning in photosynthesis, resulting in elevated carotenoid levels in ripe tomato fruit

(Powell et al., 2012). These two transcription factors, SlGLK2 and SlAPRR2, taken together with the recently identified *SlTKN2* and *SlTKN4* genes, encoding KNOTTED1-like homeobox transcription factors, influence carotenoid content in fruit by regulating plastid development, which can partially determine the level and rate of carotenoid accumulation (Nadakuduti et al., 2014; Galpaz et al., 2008).

1.5.2 The role of plastids in the regulation of carotenoid biosynthesis

Carotenoids are mainly biosynthesized and sequestered in plastids, in chloroplasts and chromoplasts. One of the major developmental processes during fruit ripening is the breakdown of chloroplasts and formation of chromoplasts, leading to dramatic changes in the carotenoid profiles of tomato fruit. There is a positive relationship between the number of chloroplasts at the green stage and the number of chromoplasts when the fruit is ripe (Isaacson, 2002a; Galpaz et al., 2008). Therefore, formation and transition of plastids, as well as their division and differentiation, are closely related to carotenoid sequestration and stable storage, processes which play a crucial role in carotenoid accumulation (Li and Yuan, 2013).

During carotenoid biosynthesis in tomato fruit, chromoplasts act as metabolic sinks, the formation and development of which have significant impact on carotenoid accumulation. The *Or* (*Orange*) gene is involved in the regulation of this process. The *Or* gene was originally cloned from cauliflower using a dominant mutant allele (Li et al., 2001). The *Or* gene encodes a DNAJ cysteine-rich domain-containing protein (Lu et al., 2006), and can trigger the transition from non-coloured plastids into chromoplasts accompanied by significantly increased accumulation of β-carotene in cauliflower, potato and melon (Tzuri et al., 2015; Lu et al., 2006; Lopez et al., 2008). Apart from its role in plastid differentiation, Or is a major post-transcriptional regulator of PSY (Zhou et al., 2015; Chayut et al., 2017). Further studies are needed to investigate the regulatory mechanism underlying enhanced chromoplast formation by Or.

In tomato *high-pigment* mutants, the significant enhancement of carotenoid accumulation is the result of the increased number and size of chromoplasts. Relatively larger chromoplasts are observed in the *high pigment 2* (*hp2*) mutant, which bears a mutation in *DE-ETIOLATED 1* (*DET1*) leading to higher carotenoid accumulation (Mustilli, 1999; Liu et al., 2004). The *hp3* mutant, which carries a loss-of-function mutation in *SlZEP*, accumulates more carotenoids as well as ABA, with an enlarged plastid compartment (Galpaz et al., 2008).

Uniform ripening mutant (locus *U*) possesses increased intensity and altered pattern of plastids. A Golden2-Like (SlGLK2) protein, encoded by *Solyc10g008160*, is a member of the Myb transcription factor superfamily, and promotes the development of chloroplasts by regulating the expression of related chloroplast genes. Overexpression of *SlGLK2* increases the transcript level of genes encoding proteins functioning in photosynthesis, resulting in elevated carotenoid levels in ripe tomato fruit (Powell et al., 2012).

1.5.3. Regulation of carotenoid metabolism in response to the environment

In plants, carotenoids play an important role in response to external signals, including temperature, biotic and abiotic stress and light intensity.

It has been shown that carotenoid biosynthesis is affected by light signalling in tomato fruit. Repression of negative regulators of light signalling pathways, such as SlDDB1 and SlDET1, can induce lycopene accumulation in fruit, while silencing of *SlHY5*, encoding the Long Hypocotyl 5 TF, has the opposite effect (Davuluri et al., 2004, 2005; Liu et al., 2004). Carotenoids function in photoprotection of the photosystems under light stress through the xanthophyll cycle (Jahns and Holzwarth, 2012; Niyogi et al., 1998). High light intensity can induce the production of xanthophylls to protect the photosynthetic apparatus against photodamage (Galpaz et al., 2008).

Temperature is a crucial factor influencing development of tomato fruit. It has been observed that under high or low temperature conditions, lycopene accumulation in fruit is inhibited, by stimulating the biosynthesis of β-carotene at the expense of lycopene (Hamauzu et al., 1998; Dumas et al., 2003). In addition, oxidative stress and CO2 can affect the composition and content of different carotenoids in tomato fruit (Jimenez et al., 2002; Zhang et al., 2014).

1.5.4 Hormonal regulation of carotenoid metabolism

Carotenoid accumulation during fruit ripening is influenced by many internal and external factors, among which phytohormones play an important role in the signaling networks (Lee et al., 2012). For example, the regulatory mechanism promoting carotenoid biosynthesis in tomato fruit by ethylene has been extensively studied. Lycopene and β-carotene accumulation is dependent on ethylene production, which is also correlated with expression levels of biosynthetic genes, *SlPSY1* and *SlPDS* (Marty et al., 2005). Many transcription factors regulating carotenoid biosynthesis are also involved in controlling ethylene production or signaling. For example, SlRIN, a MADS-domain transcription factor, functions as a master regulator of fruit ripening, and controls the expression of *SlPSY1* by direct binding to its CArG motif in its promoter (Fujisawa et al., 2013). rin, bearing a semidominant allele of *RIN*, shows severe defects in lycopene accumulation resulting in yellow-coloured fruits (Ito et al., 2017; Giovannoni et al., 1995). Three ethylene recptors, SlETR4, SlETR6 and SlETR3/Nr, are highly expressed in fruit. Never ripe (Nr) mutant, which is an ethylene-insensitive, fails to accumulate lycopene in its fruit (Lanahan, 1994). One of the ethylene response transcription factors, Ethylene Responsive Factor 6 (ERF6), negatively regulates carotenoid accumulation in fruit. *SlERF6*-suppressed fruit show increased ethylene levels and enhanced carotenoid accumulation (Lee et al., 2012).

The roles of auxin, abscisic acid and jasmonic acid in the regulation of fruit ripening, as well as carotenoid accumulation, have also been characterized (Jones et al., 2002; Galpaz et al., 2008; Liu et al., 2012). Suppression of *Auxin Response Factor 4* (*ARF4*) results in fruits with

dark-green colour with increased numbers of chloroplasts, indicating that ARF4 affects carotenoid accumulation by regulating plastid formation and transition to chromoplasts (Jones et al., 2002). This effect is thought to involve repression of the expression of *SlGLK*1 by ARF4. GLK1 is a transcription factor that promotes chloroplast development, and repression involves binding of ARF4 to an ARF-binding motif in its promoter (Sagar et al., 2013).

Brassinosteroids are a group of steroid compounds in plants, which participate in a wide range of physiological processes, regulating different aspects of plant growth, development and reproduction. It has been shown that brassinosteroids regulate carotenoid accumulation in tomato fruit (Vardhini and Rao, 2002), and increased brassinosteroid levels can elevate carotenoid production as well as the expression of the structural genes (Liu et al., 2014a; Nie et al., 2017; Zhu et al., 2015). Several key components involved in brassinosteroid biosynthesis and signalling have been characterised (Holton et al., 2007; Jones et al., 2002; Koka et al., 2000; Bishop et al., 1999). 2,4-Epibrassinolide (EBR) has been shown to regulate carotenoid accumulation (Vardhini and Rao, 2002). Furthermore, EBR-treated pericarp discs of *Nr* accumulate more carotenoids than those of the control, suggesting the existence of a BR-induced carotenoid accumulation pathway independent of NR-mediated ethylene signal transduction (Liu et al., 2014a). The transcription factor BRASSINAZOLE RESISTANT1 (BZR1) is a key component of BR signaling. Tomato fruit overexpressing the *Arabidopsis BZR1-1D* gene exhibit enhanced carotenoid accumulation and increased soluble solid, soluble sugar, and ascorbic acid contents during fruit ripening (Liu et al., 2014a).

1.6 Metabolic engineering of carotenoids in plants

1.6.1 Biofortification of provitamin A

Improving the provitamin A content of staple crops has focussed on enhancing the levels of β-carotene in the parts of the crop that people consume, either endosperm of rice, sorghum

or maize seeds or the tubers of cassava. β-carotene can be cleaved, once consumed, to form two molecules of retinol (vitamin A), and consequently is considered a better source of vitamin A than α-carotene or β-cryptoxanthin which can also be metabolised to retinol following digestion. β-carotene levels can be enhanced by increasing synthesis (a 'pull' strategy), increasing storage or reducing catabolism in the target tissues ('protect' strategies). In fact, almost all biotechnological biofortification strategies have initially pursued the 'pull' strategy of increased synthesis, because this can be effective in tissues such as rice endosperm where no β-carotene is normally synthesised. For Golden Rice this involved expressing the gene encoding phytoene synthase (PSY) from daffodil and the gene encoding the multifunctional enzyme carotene desaturase, *CrtI*, from the bacterium *Erwinia uredovora,* which is capable of performing all the desaturation and isomerization reactions necessary to form lycopene. Activity of lycopene β-cyclase was found to be unnecessary for the formation of β-carotene in rice endosperm, probably due to endogenous activity of this enzyme in the endosperm (Ye *et al.,* 2000). The result was Golden Rice 1 which had a maximum carotenoid content of 1.6 μg/g dry weight (DW), with 50% of this as β-carotene (Ye *et al.*, 2000). Further optimisation of synthesis involved replacing the daffodil *PSY* with the gene encoding the more efficient PSY enzyme from maize together with the *CrtI* gene from *Erwinia uredorova,* both driven by the endosperm-specific rice glutelin promoter, to produce Golden Rice 2 which maximally accumulated 37 μg total carotenoids per g of grain, of which >80% was β-carotene, a level conservatively estimated to supply 50% of the recommended daily allowances (RDA) for provitamin A in 100g of rice (Paine *et al.,* 2005).

Sorghum is the second most important cereal crop in Africa with 300 million people dependent on it as a staple, and it is seriously deficient in provitamin A, iron and zinc (Che *et al*., 2016). In sorghum, use of the same strategy as for Golden Rice 2 resulted in poor accumulation of β-carotene (<2 μg/g), not due to a failure in synthesis, but due to degradation of β-carotene by oxidation in older, mature seed. This was significant in sorghum because this crop is traditionally stored, post-harvest, for several months prior to consumption. The expression of the maize *PSY* gene under the control of the endosperm-specific promoter in sorghum declined to zero by seed maturation, meaning that no more β-carotene was synthesised, post-harvest. Levels of β-carotene could be enhanced by reducing oxidative loss (a 'protect' strategy), by increasing the synthesis of tocotrienol and tocopherol (vitamin E) lipophilic antioxidants. This was achieved by expressing the gene

encoding homogentisate geranylgeranyl transferase (HGGT), which catalyses the first step committed to tocotrienol biosynthesis involving the condensation of homogentisic acid (HGA) and geranylgeranyl diphosphate (GGDP), under the control of an endosperm-specific promoter from barley in sorghum. Levels of α-tocotrienol, α-tocopherol, and γ-tocopherol, were increased 27.3-, 1.8-, and 1.7-fold, respectively and consequently, levels of β-carotene were increased to 7-12 μg/g DW, as a result of reduced loss of β-carotene through oxidation (Che *et al.,* 2016).

In cassava a strategy of expressing PSY in roots gave increases of 10-20-fold in total carotenoids, whereas high level expression of the gene encoding 1-Deoxy-D-Xylulose 5-Phosphate Synthase (DXS) together with *PSY* (Figure 1.3) gave 15- to 30-fold higher carotenoid concentrations in roots than those in storage roots from non-transformed plants, achieving concentrations >50 μg/g DW (Sayre *et al.,* 2011). Although levels were reported stable in field trials, these varieties have not been taken forward due to the availability of β-carotene enriched, 'golden cassava' germplasm for breeding (Welsch *et al*., 2010).

In staple crops where there is significant natural variation in β-carotene levels, the 'protect' strategy of reducing catabolism has proven most effective in breeding for high β-carotene varieties. Thus, selection of weaker alleles of lycopene ε-cyclase (*LCYE*), an enzyme which converts all-*trans* lycopene to α-carotene (Figure 1.3), has been shown to enhance the levels of accumulation of β-carotene in maize kernels (Harjes *et a*l., 2008). Similarly rare, weak alleles of *crtRB1*, which encodes Crtβ-carotene hydroxylase1 that converts β-carotene to β-cryptoxanthin and zeaxanthin (Figure 1.3), have been associated with higher levels of β-carotene in maize, and have been proposed for use in selecting new varieties of maize associated with yet higher levels of provitamin A (Yan *et al.,*2010).

1.6.2 Metabolic engineering of carotenoid biosynthesis

Engineering carotenoid biosynthesis has been successfully achieved in several fruits and vegetables, in addition to the bio-fortification programs for β-carotene (provitamin A), described earlier. Massive accumulation of carotenoids in plants is associated with a very active endogenous isoprenoid biosynthetic pathway in plastids. The main approach to engineer the carotenoid pathway, as for β-carotene biofortification, has been to overexpress one or multiple biosynthetic genes, (from either plants or bacteria) or combinations of genes from the two sources. *PSY* is thought to catalyse the key step in carotenoid biosynthesis (Cunningham & Gantt, 1998), which has made it the primary target for metabolic engineering. Overexpression of a bacterial *CrtB* (*PSY*) in canola (*Brassica napus*), driven by a seed-specific, napin promoter and fused to a plastid targeting peptide, achieved visibly orange seeds with up to 50-fold increase in carotenoid content (Shewmaker *et al*., 1999). Plastid-targeted expression of the *CtrB* (*PSY*) - *CrtI* (*PDS*) - *CrtY* (*LYCB*) bacterial mini-pathway in canola seeds led to increased β-carotene levels and higher β- to *α*-carotene ratios (Ravanello *et al*., 2003). Engineering several genes encoding biosynthetic enzymes has also been used in soybean, tobacco, lettuce and other species to increase specific carotenoid levels, in targeted tissues (Hasunuma *et al*., 2008; Kim *et al*., 2012; Harada *et al*., 2014).

Tomato ripening involves the visible breakdown of chlorophyll and build-up of carotenoids, with massive accumulation of antioxidant components such as lycopene and β-carotene within the chromoplasts (Egea *et al*., 2010). Different approaches combining conventional breeding and genetic engineering have been used to increase carotenoid content in tomato fruit (Zamir, 2001; Fraser, *et al*., 2009). Although *PSY1* has been shown to be a key regulator of carotenoid accumulation in tomato fruit, overexpression of *CtrB* (*PSY*) from *Erwinia* under the control of a fruit-specific promoter, did not increase lycopene content, but resulted in higher phytoene and, β-carotene with higher total carotenoid levels. Transgenic tomatoes with overexpressed *CtrI (PDS)* had an increased content of β-carotene, amounting to up to 45% total carotenoids, at the expense of reduced lycopene content (Romer *et al*., 2000). In these two cases, unchanged or decreased lycopene content together with higher β-

carotene accumulation suggested the limitations of adopting a 'pull strategy' alone, especially when the target compound (in these cases, lycopene) is not the end product of the pathway. In contrast, silencing of *LCYB* using antisense technology gave elevated lycopene levels in tomato fruit (a 'protect strategy') (Rosati *et al*., 2000).

Several studies have described regulation of carotenoid biosynthesis at the molecular level in plants (Cunningham & Gantt, 1998; Hirschberg, 2001; Liu *et al*., 2004b). Different types of regulatory mechanisms, operating at the transcriptional and posttranscriptional levels, have been suggested to be involved in accumulation of specific carotenoids (Sauret-Gueto *et al*., 2006; Lee *et al*., 2012). Over-expression of the *Orange* (*Or*) gene from cauliflower, in potato plants triggers β-carotene accumulation in tubers, by increasing PSY stability and chromoplast generation (Li *et al*., 2012). However, the regulatory mechanisms that control carotenoid accumulation remain poorly understood. Metabolic engineering of carotenoid biosynthesis would benefit enormously from the identification of regulatory mechanisms, particularly transcription factors specifically controlling carotenoid biosynthesis, to overcome the limitations on flux in particular tissues that are consumed as foods.

1.7 Approaches to the identification of transcription factors regulating carotenoid metabolism in tomato

1.7.1 Expression quantitative trait loci (eQTL) analysis

Traits are often inherited, influenced by environment and controlled by multiple genes of small or large effect. A powerful approach to understanding the genetic basis of such complex traits is through identification of quantitative trait loci (QTL) associated with the trait. Phenotypic differences may be the result of sequence polymorphisms and related changes in gene product activity, particularly when located in the coding sequences of the associated genes, but can also be attributed to variations in mRNA transcript abundance among individuals responding to a wide range of internal and external stimuli.

An expression Quantitative Trait Locus (eQTL) is a position in the genome where the polymorphism causes differences in the abundance of steady-state transcripts. Various phenotypic differences have been found that result from changes in gene expression patterns, such as disease resistance, plant development, flowering time (Werner et al., 2005; Zhang et al., 2006; Svistoonoff et al., 2007).

Expression QTL (eQTL) analysis is a refinement of the QTL approach involving the identification of regions of the genome which confer significant alterations in the expression of a given gene. This approach has been used extensively in the medical field and is starting to be used more widely in plants. eQTLs can be classified into two major groups, *cis*-eQTLs and *trans*-eQTLs, based on the proximity to the gene regulated. For *cis*-eQTLs, the transcript level of the gene of interest is regulated by polymorphisms physically located within or in close proximity to the gene. Such *cis*-eQTLs can be generated by variations in the coding sequence (such as those causing nonsense mutations that alter transcript levels through processes such as nonsense-mediated decay) or in the promoter region containing the regulatory elements. In contrast, a *trans*-eQTL is an eQTL physically located in a portion of the genome which does not coincide with the location of the gene being regulated (ie *in trans* to the regulated gene). Normally, the phenotypic variations caused by *cis*-eQTLs (differences in transcript abundance) are larger than those caused by the *trans*-located ones (Brem and Kruglyak, 2005; Hughes et al., 2006; Joosen et al., 2013; Keurentjes et al., 2007).

Many of the conducted eQTL analyses have shown that eQTLs are not evenly distributed throughout the genome. Regions where more eQTLs cluster than other locations, are called 'eQTL hotspots'. Compared to *cis*-eQTL hotspots, which most likely result from gene enrichment or infrequent recombination, *trans*-eQTL hotspots are of greater functional interest as locations of genes encoding regulators or transcription factors regulating the expression of a suit of genes functioning in the same biological process or metabolic pathway. Therefore, eQTL analysis offers a strategy well-suited to the identification of transcription factors controlling metabolism, with greater likelihood of the characterisation of regulators controlling specific pathways compared to other strategies. eQTL analysis has been applied in the study of different plant species (Potokina et al., 2008; Keurentjes et al.,

2007; West et al., 2007). In most instances the number of eQTLs for a given gene or the eQTL hotspots for a certain metabolic pathway are relatively small (between 1 and 5), which can speed up the cloning of the corresponding genes underlying the phenotypic effect.

1.7.2 Metabolite-based Genome-wide Association Studies (mGWAS)

Traits can be qualitative or quantitative. Quantitative traits, such as content of certain metabolites, are usually controlled by a number of genes, and can be placed on a continuum. In plants, a lot of phenotypic variation is controlled by qualitative trait loci (QTL). With the development of high-throughput genotyping technologies, such as next-generation sequencing, genome-wide association studies (GWAS), have been used extensively for understanding the genetic basis of complex traits in various plants (Todesco et al., 2010; Buckler et al., 2009; Kump et al., 2011; Huang et al., 2010). The identification of high-density single nucleotide polymorphisms (SNPs) allows genome-scale scans to identify small haplotype blocks which are significantly associated with variation in quantitative traits. The high diversity and quantitative properties of metabolite abundance measurements make them heritable traits that can be studied by GWAS. Metabolite-based genome-wide association studies provide a combinatorial approach for GWAS that can be linked with high-throughput metabolite profiling as performed by High Performance Liquid Chromatography (HPLC) and Liquid/Gas chromatography-mass spectrometry (LC/GC-MS). mGWAS was first developed for the model plant, *Arabidopsis thaliana*, followed by successful applications in a number of different plant species, including rice, tomato and maize (Chan et al., 2011; Chen et al., 2014; Riedelsheimer et al., 2012; Rosenwasser et al., 2014; Zhu et al., 2018). mGWAS offers a relatively new technique for identification of regulators of carotenoid accumulation in tomato.

1.7.3 Virus-Induced Gene Silencing (VIGS)

Virus-Induced Gene Silencing (VIGS) is an efficient method for rapid characterisation of gene function in plants (Baulcombe, 1999). Tobacco Rattle Virus (TRV) is one of the viruses

most commonly used to trigger post-translational silencing of target gene in fruit tissues and has been successfully applied in tomato fruit (Orzaez, 2005). Engineered TRV has been used to study the function of genes involved in fruit-specific processes by agroinjection (Orzaez, 2005; Fantini et al., 2013). The irregular distribution of VIGS limits the accurate quantification of the effects of gene silencing in fruit. An improved visual reporter VIGS system for tomato fruit was developed by monitoring anthocyanin production in transgenic *Del*/*Ros1* tomatoes in which Delila and Rosea 1, transcription factors from *Antirrhinum majus*, are overexpressed specifically in fruit resulting in induced accumulation of anthocyanin which confers purple colour throughout the fruit (Figure 1.5) (Orzaez et al., 2009). VIGS-silencing of *Del* and *Ros1* in purple tomatoes blocks the production of anthocyanin, leading to red sectors, which can be distinguished easily from the non-silenced sectors which remain purple on the same fruit. Silencing of target genes-of-interest (GOI), together with *Del* and *Ros1* in *Del/Ros1* purple tomatoes, allows the quantitative assessment of GOI function based on the colour phenotype of the silenced red sectors.

1.7.4 Genome editing by CRISPR /Cas9

An engineered form of the clustered, regularly interspaced, short palindromic repeat (CRISPR) system of *Streptococcus pyogenes* has been shown to function in plants (Nekrasov et al., 2013; Belhaj et al., 2013; Jinek et al., 2013; Upadhyay et al., 2013; Xie and Yang, 2013). CRISPR systems provide a unique prokaryotic defence against invading DNAs, such as plasmids and viruses (Barrangou et al., 2007). In the type II CRISPR system, the Cas9 protein is directed to target sites in the genome by short RNAs-CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA)-functioning as an endonuclease (Jinek et al., 2012). The presence of a conserved sequence motif (NGG) recognised as a protospacer adjacent motif (PAM) downstream of the target spacer sequence is also essential for cleavage (Gasiunas et al., 2012). In the reconstituted system, Cas9 can form a complex with a synthetic single-guide RNA (sgRNA), comprising a fusion of crRNA and tracrRNA (Jinek et al., 2012). The double strand breaks (DSB) introduced by the Cas9 nuclease guided by sgRNA can lead to various sequence mutations as a result of DNA damage repair by error-prone nonhomologous end-joining (NHEJ) or homology-dependent recombination (HR). CRISPR-Cas system has been demonstrated to work efficiently for genome editing in bacterial, yeast,

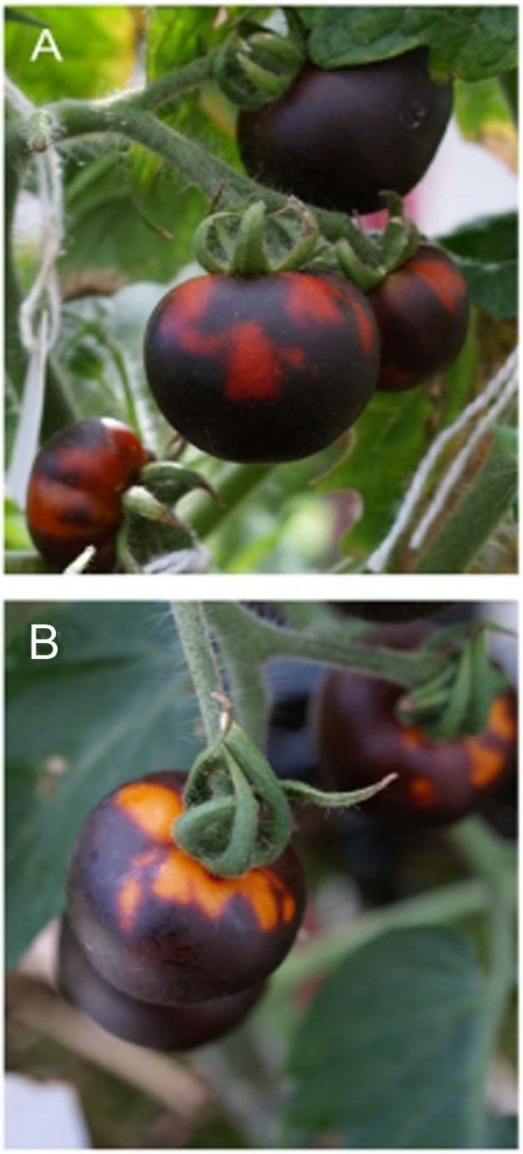


Figure 1.5 Phenotypes of Fruit VIGS in the Del/Ros1 purple tomato background (Orzaez et al., 2009). A) Shows red sectors in fruit agro-inoculated with TRV2 carrying fragments of the *Del* and *Ros1* genes. The red sectors show the tissues in which *Del* and *Ros1* have been silenced. B) Shows yellow sectors in fruit agro-inoculated with TRV2 carrying a fragment of the *PDS* gene as well as fragments of the *Del* and *Ros1* genes. The yellow sectors show that *PDS* is silenced co-incidentally with *Del* and *Ros1* and inhibits carotenoid/lycopene production. This shows that fruit VIGS in *Del/Ros1* tomatoes provides an excellent system for analysing transient loss of gene-of-interest (GOI) function (from Orzaez et al., 2009).

and animal systems (Cong et al., 2013; Gasiunas et al., 2012; Jinek et al., 2013) and has been applied to different plants, such as *Arabidopsis thaliana*, *Nicotiana benthamiana*, tomato, rice and wheat (Jiang et al., 2013b; Feng et al., 2014; Upadhyay et al., 2013). This method is well-suited for de novo generation of knock-out mutations of candidate regulatory genes controlling carotenoid biosynthesis in tomato.

Together, eQTL mapping, VIGS silencing and CRISPR/Cas9 genome editing provided the core technologies I employed to address the transcriptional regulation of carotenoid metabolism during tomato fruit ripening, which is the topic of this PhD thesis.

Chapter 2

General Materials and Methods

2.1 Materials

2.1.1 Plant Materials

Wild-type tomato varieties, MicroTom and MoneyMaker, were used in VIGS experiments. Gene time-course expression analysis and stable gene transformations were from lab stocks of tomato seeds (MoneyMaker, M82 and MicroTom). Seeds of transgenic *E8:Del/Ros1* in both MicroTom and MoneyMaker genetic backgrounds were obtained from lab stocks generated by Dr. Eugenio Butelli (Butelli et al., 2008).

The tomato introgression population *Solanum lycopersicum* x *Solanum pennellii*, and tomato introgression population *Solanum lycopersicum* x *Solanum lycopersicoides* were provided by Dani Zamir at the Hebrew University of Jerusalem (Eshed and Zamir, 1995).

Tomato wild relatives *Solanum corneliomulleri* (LA0107), *Solanum peruvianum* (LA1278), *Solanum pennellii* (LA0716), *Solanum arcanum* (LA2172), *Solanum chmielewskii* (LA1208), *Solanum habrochaites* (LA1778), *Solanum chilense* (LA1969), *Solanum pimpinellifolium* (LA1578) and *Solanum neorichii* (LA2133) were supplied by the Tomato Genetic Resource Center (http://tgrc.ucdavis.edu/).

2.1.2 Bacterial and Yeast Strains

Information on the bacterial and yeast strains used in this thesis is listed in **Table 2.1**.

Table 2.1 Bacteria and Yeast Strains Used

Strain	Type	Antibody Resistance	Use
DH5α	*E.coli*	n.a.	Normal *E.coli* for propagation of plasmid vectors
DB3.1	*E.coli*	Chloramphenicol	ccdB resistant *E.coli* strain, for propagation of plasmids containing Gateway destination cassettes
GV3101 :pMP90	*Agrobacterium tumefaciens*	Gentamicin Rifampicin	*Agrobacterium tumefaciens* strain for plant transient transformation experiments using agroinfiltration
AGL-1	*Agrobacterium tumefaciens*	Carbenicillin Rifampicin	*Agrobacterium tumefaciens* strain used for stable tomato transformation
Y187	*Saccharomyces cerevisiae*	n.a.	Yeast two-hybrid assay

2.1.3 Chemicals

All chemicals used were analytical reagent grade were purchased from Invitrogen, Roche, Qiagen, New England Biolabs, Sigma and Promega. (E/Z)-Phytoene (78903), lycopene (L9879), β-carotene (C4582), lutein (7168), α-carotene (50887) and zeaxanthin (14681) used as standards were of HPLC grade and were purchased from Sigma Ltd.

2.1.4 Antibiotics

Antibiotics used to select bacteria and transgenic plants are listed in **Table 2.2**.

Table 2.2 Stock and Working Concentrations of Antibodies

Antibodies	Solvent	Use	Stock Conc.	Working Conc.
Ampicillin	H_20	Bacterial screening	100 mg/mL	100 µg/mL
Chloramphenicol	Ethanol	Bacterial screening	34 mg/mL	34 µg/mL
Gentamicin	H_20	Bacterial screening	100 mg/mL	20 µg/mL
Kanamycin	H_20	Bacterial screening	100 mg/mL	50 µg/mL
		Transgenic plant screening	100 mg/mL	100 µg/mL
Rifampicin	Methanol	Bacterial screening	25 mg/mL	25 µg/mL
Streptomycin	H_20	Bacterial screening	100 mg/mL	100 µg/mL
Carbenicillin	H_20	Bacterial screening	100 mg/mL	50 µg/mL

2.1.5 Plasmids

Vectors and constructed plasmids used in this thesis are listed in **Appendix 1**.

2.1.6 Media Preparation Recipes

The recipes for the media used in this thesis are listed in **Appendix 2**.

2.2 Methods

2.2.1 Primer design

The primers used in this thesis were designed using Primer Blast from NCBI website (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) followed by a quality check using

NetPrimer (http://www.premierbiosoft.com/netprimer/netprlaunch/Help/xnetprlaunch.html). The following principles were applied during the primer design: for regular PCR and qPCR, the optimised primer length was between 20-28 bp (except for the primers used for Gateway cloning, GoldenGate cloning, GoldenBraid cloning and other specific purposes); the Tm of primers was ~60 °C with a GC content between 40-60%; secondary structures were avoided; the 3' ends of the primers were terminated with G or C if possible, to achieve greater stability. The products of primers formed during qPCR were around 80-120 bp. The primers used in this thesis are listed in **Appendix 3**.

2.2.2 Polymerase chain reaction (PCR)

PCR reactions were conducted using G-Storm Thermal Cyclers (Kapa Biosystems). Different DNA polymerases and the corresponding set-ups were applied according to purpose. For normal PCR reactions (such as for genotyping), in each 10 µL reaction, 10-20 ng of DNA template was mixed with 1 µL of forward primer (10 µM), 1 µL of reverse primer (10 µM), water (DNase and RNase free), 5 µL of 2X GoTaq® 2G Green master mix containing DNA polymerase, proprietary G2 reaction buffer, 1.6 mM total dNTPs and loading dyes (yellow and blue) (Promega). For the PCR reactions with high-fidelity requirements, which were used for plasmid constructions and sequencing, Phusion® High-Fidelity DNA Polymerase and its corresponding buffer (New England Biolabs) were used in a mixture composed of DNA template, forward and reverse primers, dNTPs and water in a total volume of 20 µL. The standard PCR protocol is shown in **Table 2.3**.

Table 2.3 Standard PCR protocol

Step		Temperature	Duration
Initial denaturation		94 °C	4 min
25-35 cycles[a]	denaturation	94 °C	30 sec
	annealing[b]	60 °C	30 sec
	extension[c]	72 °C	30 sec - 2 min
Final extension		72 °C	10 min

a. The number of cycles in the PCR for genotyping was set no more than 30 cycles, to avoid unspecific amplification.

b. The annealing temperature for specific primers are listed in Appendix 3.

c. The extension time was decided by the sizes of the PCR products as well as the amplification speeds of the specific DNA polymerases. The DNA polymerases used in this thesis are: *Taq* DNA polymerase (New England Biolabs®), 1 min per kb; Phusion High-Fidelity DNA polymerase (Thermo Scientific®), 15-30s per kb.

2.2.3 Purification of DNA from PCR reactions or agarose gels

PCR products or other DNA products were separated using 1% w/v agarose gels in 1 x TBE buffer and purified using NucleoSpin® Gel and PCR Clean-up kits (MACHEREY-NAGEL), following the manufacturer's instructions.

2.2.4 Preparation of competent cells of *E. coli* for heat shock transformation

A single colony of *E. coli* was picked for overnight growth in 5 mL LB liquid medium at 37 °C, 200-220 rpm. Overnight culture (1mL) was added to 100 mL LB liquid medium, and grown at 37 °C, 200-220 rpm, until the OD_{600} reached 0.35-0.4. The culture was cooled down on ice for 15 min, and then centrifuged at 1800 x g, 4 °C for 10 min after transferring to pre-chilled 50 mL falcon tubes. The supernatant was discarded, and the pellet was resuspended in 30 mL pre-cooled 0.1 M $CaCl_2$, followed by incubation on ice for 30 min and centrifugation

for 10 min at 1800 x g, 4 °C. The supernatant was discarded, and the pellet was resuspended in 2 mL pre-cooled 0.1 M $CaCl_2$. The resuspended culture was mixed with an equal volume of 30% glycerol. The competent cells were aliquoted into 1.5 mL Eppendorf tubes, which could be stored at -80 °C for several months.

2.2.5 *E. coli* heat shock transformation

10-50 ng of plasmid were mixed with 100 µL of *E. coli* competent cells by pipetting and kept on ice for 30 min. Then the mixture was kept in a 42 °C water bath for 1 min. SOC medium (900 µL) was added, and the mixture was incubated at 37 °C for 30 min. The cells were pelleted at 1000 x g for 3 min and resuspended in 100 µL SOC medium. Aliquots of 20 µL and 80 µL of the cell suspension were spread onto LB agar plates with the appropriate antibiotics. The plates were incubated at 37 °C overnight.

2.2.6 Plasmid DNA isolation

Single colonies were picked from the selection plates and incubated in 10 mL of LB medium with appropriate antibiotics at 220 rpm at 37 °C. Plasmid DNA was isolated from overnight cultures using QIAprep®Miniprep Kit (Qiagen), following the manufacturer's instructions.

2.2.7 Preparation of electrocompetent *Agrobacterium tumefaciens* competent cells (GV3101:pMP90, AGL-1)

Agrobacterium tumefaciens from a single colony was grown in LB medium with appropriate antibiotics at 220 rpm at 28 °C. A sample of the overnight culture (2 mL) was inoculated into 100 mL of LB selection medium and grown at 220 rpm, 28 °C until the OD_{600} reached 0.5-1.0. The culture was cooled on ice and then transferred into 50 mL pre-cooled falcon tubes. The culture was centrifuged at 1800 x g, 4 °C for 15 min. The supernatant was discarded,

and the pellet was resuspended in 50 mL of ice cold 10% glycerol. The cell suspension was centrifuged again at 1800 x g, 4 °C for 15 min and then washed with 25 mL of cold 10% glycerol. The competent cells were finally resuspended with 2 mL of cold 10% glycerol and then aliquoted with 100 µL per tube, which could be kept at -80 °C for several months.

2.2.8 Electroporation

Plasmid (200 µg) was mixed with 100 µL *Agrobacterium* electrocompetent cells which were stored at -80°C. The mixture was transferred into a pre-cooled electroporation cuvette. The competent cells were electrocuted using a BioRad Pulser (BioRad Laboratories) set as 400 Ω, 25 µFD and 2.5 KVolts. SOC medium (1mL) was added and the *Agrobacterium* cells recovered at 28 °C for 1 hour. The cells were pelleted at 1000 x g for 3 min and resuspended in 100 µL SOC medium. Aliquots of 20 µL and 80 µL of the cell suspension were spread onto LB agar plates with appropriate antibiotics. The plates were incubated at 28 °C and the colonies were checked after two days.

2.2.9 Plant DNA isolation

Plant DNA was isolated from the finely ground powder of plant tissues (leaves, fruits, flowers or roots) using DNeasy® Plant Mini Kit (Qiagen), following the manufacturer's instructions.

2.2.10 Plant RNA isolation and DNase treatment

Tomato fruit RNA was isolated according to an optimised Trizol extraction method. This method is based on the property of an acid guanidinium thiocyanate-phenol-chloroform mixture to separate RNA/DNA (in the aqueous phase) from protein partitions (in the organic phase). Pericarps of tomato fruits were deprived of seeds and ground into fine powder using a blender and liquid nitrogen. About 200 mg homogenised tissues were placed in 2 mL

tubes. 1.5 mL Tri Reagent® (Sigma) was added to each tube and mixed by vigorous vortexing. The samples were incubated in the dark for 5 min at room temperature while shaking. 1-Bromo 3-chloropropane (BCP, Sigma) (150 µL) was added in the mixture was vortexed vigorously for 15 sec. Samples were incubated in the dark at room temperate for 10 min while shaking and then centrifuged at 20,000 x g, 4 °C for 10 min. The aqueous upper phase (around 750 µL) was transferred to fresh 1.5 mL Eppendorf tubes. An equal volume of isopropanol was added and mixed by vortexing at a moderate speed. The samples were incubated for 5 min at room temperature and pelleted at 20,000 x g, 4 °C for 10 min. The supernatant was discarded and the RNA pellet was washed with 500 µL isopropanol. RNA was re-pelleted by centrifugation for 5 min at 6,300 g, 4 °C. The supernatant was discarded. The pellet was washed twice with 1 mL 75% ethanol. The supernatant was removed, and the RNA pellet was air-dried at room temperature for 5 min in a fume hood. The RNA was dissolved in 40 µL RNase- and DNase-free water (Sigma). The RNA solution was centrifuged for 5 min at 20,000 x g, 4 °C, and 32 mL of the supernatant was transferred into an RNase- and DNase-free Eppendorf tube.

10× DNase1 buffer (4µL) and DNase1 (4µL, 4 units) (Roche) were added for DNase treatment. After a 45-min incubation at room temperature, 25 mM EDTA (4 µL) was added and incubated for 10 min at 65 °C to stop the reaction.

Alternatively, plant RNA was isolated using an RNeasy® Mini Kit (Qiagen), following the manufacturer's instructions.

2.2.11 Quantification of DNA/RNA

The concentrations of DNA and RNA were quantified using a NanoDrop 2000C UV-Vis Spectrophotometer (Thermo) following the manufacturer's instructions.

2.2.12 First-strand cDNA synthesis

First strand cDNA was synthesized using SuperScript™ III (Invitrogen). Up to 3 µg total RNA was used for reverse transcription, along with 1 µL of primer mix (mixed with equal amounts of 10 µM oligo dT (Sigma) and 10 µM random primer (Invitrogen)) and 1 µL of 10 mM dNTP was added and the mixture was incubated at 65 °C for 5 min. For reverse transcription, primer-annealed RNA from the last step was mixed with 6 µL 5X First-strand buffer, 2 µL 0.1M DTT, 1 µL RNase OUT (40 u/µL, Invitrogen) and 1 µL SuperScript™ III (Invitrogen). The reaction was incubated at 50 °C for 60 min followed by 70 °C for 15 min. The cDNA was diluted to 10 ng/µL according to the initial amount of RNA used for the reaction, 5 µL of which was used for the following RT-qPCR analysis.

2.2.13 Real-time quantitative PCR (RT-qPCR)

SYBR® Green JumpStart™ Taq ReadyMix™ (Sigma) was used to perform all the RT-qPCR reactions by using the X96 Touch™ Real-Time PCR Detection System (Biorad). The standard RT-qPCR protocol is shown in **Table 2.4**.

Table 2.4 Standard RT-qPCR protocol

Step		Temperature	Duration
Initial denaturation		94 °C	4 min
40-50 cycles	denaturation	94 °C	10 secs
	annealing	60 °C	10 secs
	extension	72 °C	15 secs
Final extension		72 °C	1 min

The data were analysed using CFX Maestro Software. *Actin* (*Solyc03g078400*) was selected as the house-keeping reference gene due to constitutive expression among variable tomato

tissues under different experimental treatments. The data were analysed according to CFX Maestro Software.

2.2.14 Gateway Cloning

*att*B PCR primers were designed using VectorNTI™ (Invitrogen), to incorporate *att*B sites into PCR products, which were used in Gateway® BP recombination reactions with a donor vector (pDONR207™ or pDONR221™, Invitrogen). In 10 µL reaction mix, 1 µL pDONR™ vector, 1-8 µL *att*B-PCR product, 1 µL BP Clonase™ and TE buffer (pH 8.0) were added. After a 2-hour incubation at 25°C, 1 µL Proteinase K solution was added and the reaction mix was incubated at 37°C for 10 min to stop the reaction. 2µL BP reaction then was transformed into suitable competent cells (ccdB sensitive).

In 10 µL mix for LR recombination reactions, 1 µL destination vector, 1-8 µL (100ng/µL) entry clone, 1 µL (1 unit) LR Clonase™ (Invitrogen) and TE buffer (pH 8.0) were added. After a 2-hour incubation at 25°C, 1 µL (1 unit) Proteinase K solution was added and the reaction mix was incubated at 37°C for 10 min to stop the reaction. 2µL LR reaction were then transformed into suitable competent cells with appropriate antibiotic selection.

2.2.15 Stable tomato transformation

Seed germination

Tomato seeds were treated with 70% ethanol for two minutes to loosen the gelatinous seed coat. Ethanol was removed, and the seeds were rinsed with sterile water. The seeds were sterilised in 10% Domestos in water (v/v) for three hours while shaking. After sterilisation, the seeds were washed with sterile water four times.

75-100 seeds were used for each transformation. Approximatly 20-30 seeds were placed in tubs with germination medium and incubated at 4°C for at least three weeks to synchronise their germination time. Seeds could be stored at 4°C up to three months. After the 4°C treatment, seeds were incubated in the culture room (16-hour photoperiod, supplemented with Gro-Lux incandescent light, which was especially important for the regeneration stage). Seedlings were grown for 7-10 days until the cotyledons were expanded but no true leaves were visible.

Transformation procedure

Day 1 *Agrobacterium tumefaciens AGL1* was inoculated in 10 mL of L medium with antibiotics (Rif µg/mL, Carb 50 µg/mL) and grown at 28°C, 200 rpm.

On each plate, containing the cell suspension medium with 0.6% agarose or MS medium supplemented with vitamins and 3% sucrose (either medium was added with 0.5 mg/L 2,4-D), 1 mL of fine tobacco suspension culture was spread to form an even layer as the feeder layer. The plates were left unsealed and stacked in the culture room under low light.

Day 2 Sterilised Whatman No.1 filter paper was placed on the top of the feeder plates. All the air bubbles were carefully removed to make sure the filter paper was completely soaked with the tobacco cells.

1 mL overnight culture of *AGL1* was inoculated into a new flask with 100 mL L medium and grown at 28°C, 200 rpm for three to four hours, until the OD_{600} reached 0.4-0.5. The pelleted AGL1 culture then was resuspended in an equal volume of MS+vitamins medium supplemented with 3% sucrose (pH=5.8).

Cotyledons from the seedlings were used for transformation because hypocotyls can give rise to a high number of tetraploids. Cotyledons were cut under cold, sterile water with a

rolling action of a rounded scalpel blade to minimise damage to the cuttings. The tips of the cotyledon were cut off, and the rest of the cotyledon was cut transversely to generate two explants about 0.5 cm long. The explants were kept in a petri dish with cold water to prevent damage during the cutting.

All the explants were scooped up using rounded forceps and transferred to a new petri dish where they were immersed in the *AGL1* suspension. After incubation, the explants were blotted on sterile filter paper to remove excess Agrobacteria and then placed on the feeder plate with the abaxial surface uppermost (upside down). Each feeder plate could hold 30-40 explants. The explants and Agrobacteria were co-cultivated in the culture room for 48 hours under low light.

Day 4 After 2 days of co-cultivation, the pieces of tomato cotyledons were transferred onto tomato regeneration plates harbouring Timentin (320 mg/L), Zeatin Riboside (2mg/L) and the appropriate antibiotic according to the T-DNA selectable transformation marker. 12-16 pieces of explants were placed right side upwards on each plate to keep good contact with the medium during the growth. Agargel was used as the setting agent because it produces soft medium into which the cuttings can be pushed gently. Unsealed plates were kept in the culture room for about two weeks.

Week 2 or 3 Explants were transferred to fresh plates containing Timentin as well as Cefotaxime (250mg/L), every two to three weeks. Deep Petri dishes and tubs were used when the regenerating materials were too large for the original plates.

Shoots were cut from the explants and put into rooting medium with appropriate antibiotics at reduced concentrations. Recutting was undertaken if the explants did not root initially. Any shoots showing no root generation were likely to be escapes. The real transformants were further confirmed by a simple PCR assay for *npt II* in genomic DNA, wherever the kanamycin resistance gene was used as the selectable marker.

Before being transferred to soil, the roots of the regenerated plants were washed gently under running water to remove as much of the medium as possible. Plants were grown in hydrated and autoclaved Jiffy pots (peat pots). The pots were kept enclosed and covered with tissue paper to ensure high humidity and avoid direct glare during the recovery stage in the growth room. The humidity was gradually reduced. The plants were ready to be moved to the glasshouse once the roots could be seen growing through the peat pots.

(All the recipes for media used for stable tomato transformation can be found in **Appendix 2**).

2.2.16 Statistics

Unless specifically stated, two-tailed Student's t-tests (paired or unpaired) were utilized to compare group differences throughout this thesis. Differences with *p* values less than 0.05, were considered as significant.

Chapter 3

Investigation of genetic foundation and regulatory basis of carotenoid biosynthetic pathway in tomato fruits

3. 1 Introduction

3.1.1 *S. lycopersicum* x *S. pennellii* introgression population

Tomato is an excellent system for the analysis of complex genetic traits, as its wild relatives in the same clade with very similar genetic constitution are divergent morphologically and anatomically (Stevens and Rick, 1986; Moyle, 2008). *Solanum pennellii*, one of the most distant interfertile relatives of cultivated tomato, possesses striking differences in leaf morphology, fruit characteristics, drought resistance and disease resistance compared to *S. lycopercicum* (Moyle, 2008; Bolger et al., 2014; Koenig et al., 2013). The interfertility between these two species, was exploited for the establishment of an introgression population containing 76 introgression lines (ILs), which was the first interspecific population used to introduce novel QTL variation into breeding superior varieties of tomato

(Eshed and Zamir, 1995; Brooks et al., 2014; Lippman et al., 2007). The introgressed regions in ILs cover the entire genome of the wild species as overlapping segments in the genetic background of domesticated tomato *S. lycoersicum* cv. *M82* and consist of 76 distinct lines(Figure 3.1; Eshed and Zamir, 1995). The physical map with Restriction Fragment Length Polymorphism (RFLP) markers of this introgression population is available on SOL Genomics Network (https://solgenomics.net/cview/map.pl?map_version_id=il6). The size of theintrogressed region in each IL ranges from a few genes to more than a thousand gens. The generation of these tomato ILs led to the discovery of more than 3069 QTLs (Alseekh et al., 2013), that affect morphology and yield (Semel et al., 2006), fruit coloration (Liu et al., 2003), metabolite levels (Schauer et al., 2008, 2006), volatile metabolites (Tieman et al., 2006; Liu et al., 2016), antioxidants (Rousseaux et al., 2005) and biotic and abiotic resistance (Sharlach et al., 2013; Uozumi et al., 2012; Bolger et al., 2014). The availability of the genome sequence of *S. pennellii* and *S. lycopersicum* further facilitates the mapping and cloning of the functional genes underlying the QTLs, some of which have been confirmed by reverse genetic studies (Zanor et al., 2009), while others remain to be further validated.

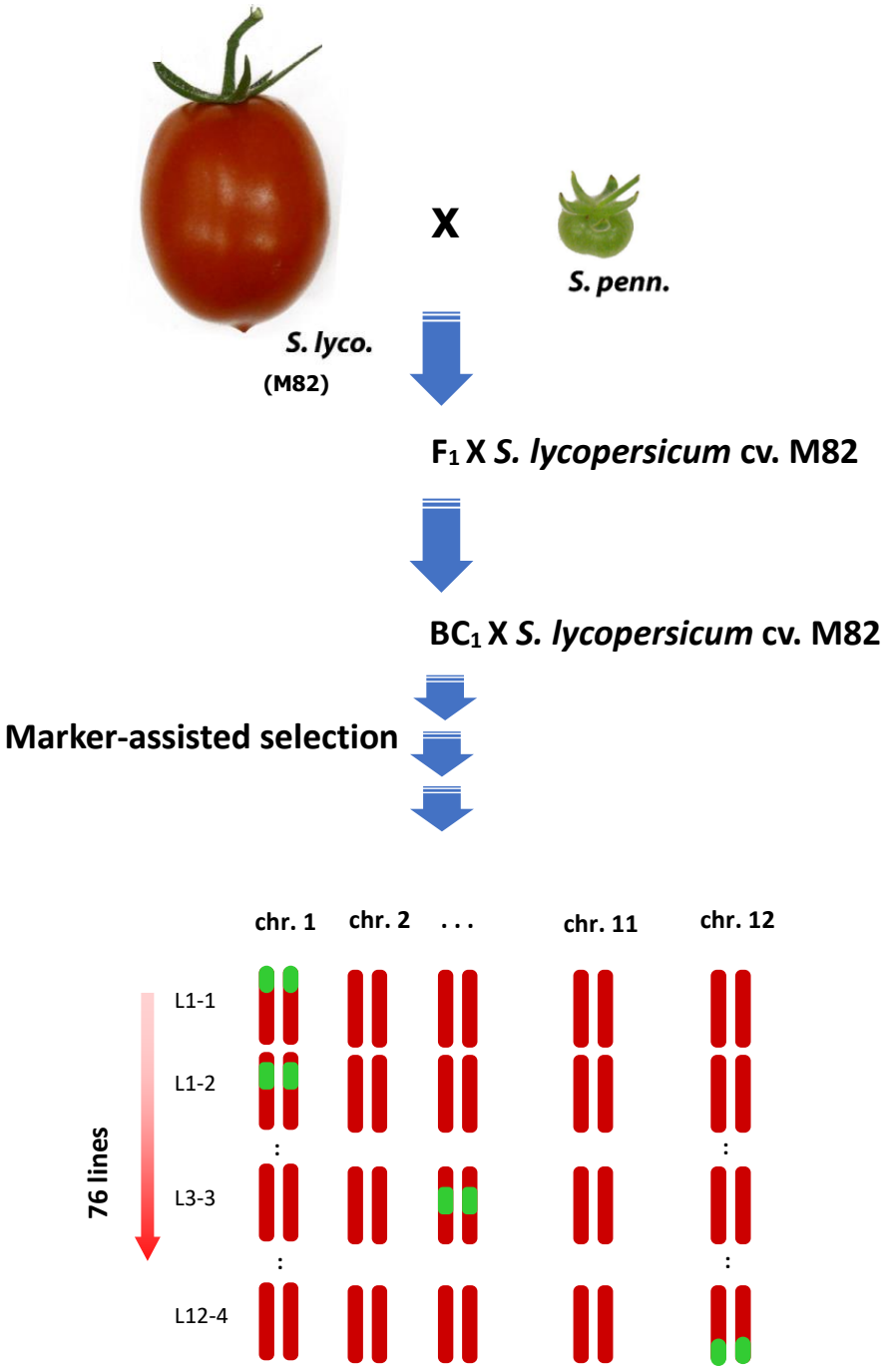


Figure 3.1 Crossing scheme for the derivation of *S. pennellii* introgression lines (ILs). The *S. pennellii* introgression population was developed by Dani Zamir's group (Hebre University, Isreal) in the background of *S. lycopercisum* cv. M82, including single introgressed regions derived from the green-fruited species, *S. pennellii*, which was generate by successive introgression backcrossing, faciliated by marker-assisted selection. Red fragments represent the genomic regions from *S. lycopersicum* cv. M82; green fragments represent the genomic regions from *S. pennellii*. Redrawn from Eshed and Zamir, 1995, *Genetics* (Eshed and Zamir, 1994 and 1995).

The red colour of the fruits of cultivated tomato, *S. lycopersicum*, is closely related to the production and accumulation of carotenoid pigments during the ripening process, particularly lycopene. This population is suitable for establishing functionality in regulating carotenoid biosynthesis is because that *S. pennellii* does not make lycopene in its fruit and produces green fruit when ripe (Figure 3.1). Several studies on carotenoid biosynthesis in tomato have been done using this introgression population. For example, *Delta* mutant, bearing a single dominant gene, *Del*, results in orange fruit when ripe, due to the significantly increased accumulation of δ-carotene at the expense of lycopene. *Del* was first roughly mapped to IL 12-2, containing a homozygous substitution of 35cM segment from *S. pennellii* (Eshed and Zamir, 1995). With the establishment of a backcrossed population between IL 12-2 and M82, it was finally determined located between RFLP markers CT-79 and TG-263 (Ronen et al., 1999). 16 QTLs were mapped based on the fruit colour phenotypes, most of which were relevant to carotenoid accumulation (Liu et al., 2003). As the genome sequence of *S. pennellii* has been available since 2013, with the well-characterised catalytic steps along carotenoid biosynthesis, an *S. pennellii* introgression population is suitable for the study of the transcriptional regulation of carotenoid biosynthesis in tomato fruit by eQTL analysis.

3.1.2 Carotenoid biosynthesis in tomato

The massive accumulation of carotenoids in tomato fruit is associated with a very active endogenous isoprenoid biosynthetic pathway in the plastid. The red colour of lycopene begins to be seen from breaker stage when the ripening process is triggered by ethylene. Several mutants have been found which cause abnormal carotenoid accumulation with impaired ripening, such as *yellow flesh* (*r*), *old gold* (*og*), *tangerine* (*t*), *Never-ripe* (*Nr*), *Green-ripe* (*Gr*), *Colour-less non-ripening* (*Cnr*) and *ripening-inhibitor* (*rin*) mutants (Fray and Grierson, 1993; Ronen et al., 2000a; Isaacson, 2002a; Barry and Giovannoni, 2006; Lanahan, 1994; Liu et al., 2004; Manning et al., 2006; Mustilli, 1999). The carotenoid biosynthesis pathway begins with the formation of phytoene from geranylgeranyl diphosphate from the central isoprenoid pathway, called 2-C-methyl-erythritol 4-phosphate (MEP) pathway, (Figure 1.3) rather than the mevalonic acid pathway.

Whilst both isoprenoid pathways produce isopentenyl pyrophosphate (IPP), the MEP pathway is plastic in nature and leads to the formation of carotenoids, phytols, plastoquinone-9, and diterpenes (Lichtenthaler, 1999). IPP is converted to geranylgeranyl pyrophosphate (GGPP) which is the precursor for the formation of carotenoids. Two molecules of GGPP are condensed in a head to tail manner to from a colourless compound phytoene, and this is a key step in the carotenoid biosynthesis pathway (Cunningham and Gantt, 1998). This two-step reaction is catalysed by the enzyme phytoene synthase (PSY). Tomato contains two paralogs of the *PSY* gene, *PSY1* and *PSY2*. The *PSY1* gene encodes a fruit-ripening-specific isoform, whilst *PSY2* encodes an isoform that predominates in green tissues, including mature green fruit and has no role in carotenogenesis in ripening fruit (Fraser et al., 1999). Phytoene then undergoes a series of four desaturation reactions leading to phytofluene, ζ-carotene, neurosporene and finally lycopene. The four sequential desaturations are catalysed by two related enzymes in plants; phytoene desaturase (PDS) and ζ-carotene desaturase (ZDS). A carotenoid isomerase (CRTISO) activity is additionally required to transform the poly *cis* lycopene (pro-lycopene) to the all trans-isomer. Cyclization of lycopene marks the branch point in the plant carotenoid pathway. Lycopene β-cyclase (LCYB) catalyses the formation of β-carotene from lycopene in a two-step reaction that creates one β-ionone ring at each end of the lycopene molecule. In another branch, δ-carotene is produced by the addition of one ε-ring to lycopene in the presence of lycopene ε-cyclase (LCYE) (Ronen et al., 2000a; Cunningham and Gantt, 2001). Xanthophylls are formed by the oxygenation of carotenes, typically by the addition of hydroxyl-, epoxy- or keto- groups.

In this Chapter, I applied an eQTL approach to study carotenoid biosynthesis in tomato fruit, to find *cis*- and *trans*-eQTL candidates involved in the regulation of this pathway. The analysis was conducted using the *S. lycopercisum* x *S. pennellii* introgession population containing 76 lines in total, as well as the two parental lines *S. lycopercisum* cv. *M82* and *S. pennellii* (LA0716). The high complexity of this pathway due to the multiple roles of its intermediates in plant growth and development, meant that the entire pathway was dissected into three parts and these were analysed independently.

3.2 Materials and Methods

3.2.1 Plant materials

Transcriptome profiling of ripe fruits from *S. pennellii* introgression lines, as well as the parental lines, *S. lycopersicum* (M82) and *S. pennellii* was generated and kindly provided by our collaborators, Dr Je Min Lee and Dr Jim Giovannoni, USDA Robert W Holley Centre, Cornell University. The RNA-Seq data was deposited in Tomato Functional Genomics Database (http://ted.bti.cornell.edu/). The fruits were harvested when 80% to 100% of one of the parental lines, M82 were ripe. The RNA was extracted from fruit flesh only, and the skin/peel of the fruit having been removed prior to collecting tissue for RNA-Seq on Illumina HiSeq 2000 platform. The strand-specific RNA-Seq library construction protocol and analysis methods were applied (Zhong et al., 2011, 2013) The RNA-seq Raw counts for each gene were normalised to Reads Per Kilobase of exon model per Million mapped reads (RPKM).

3.2.2 Heat maps

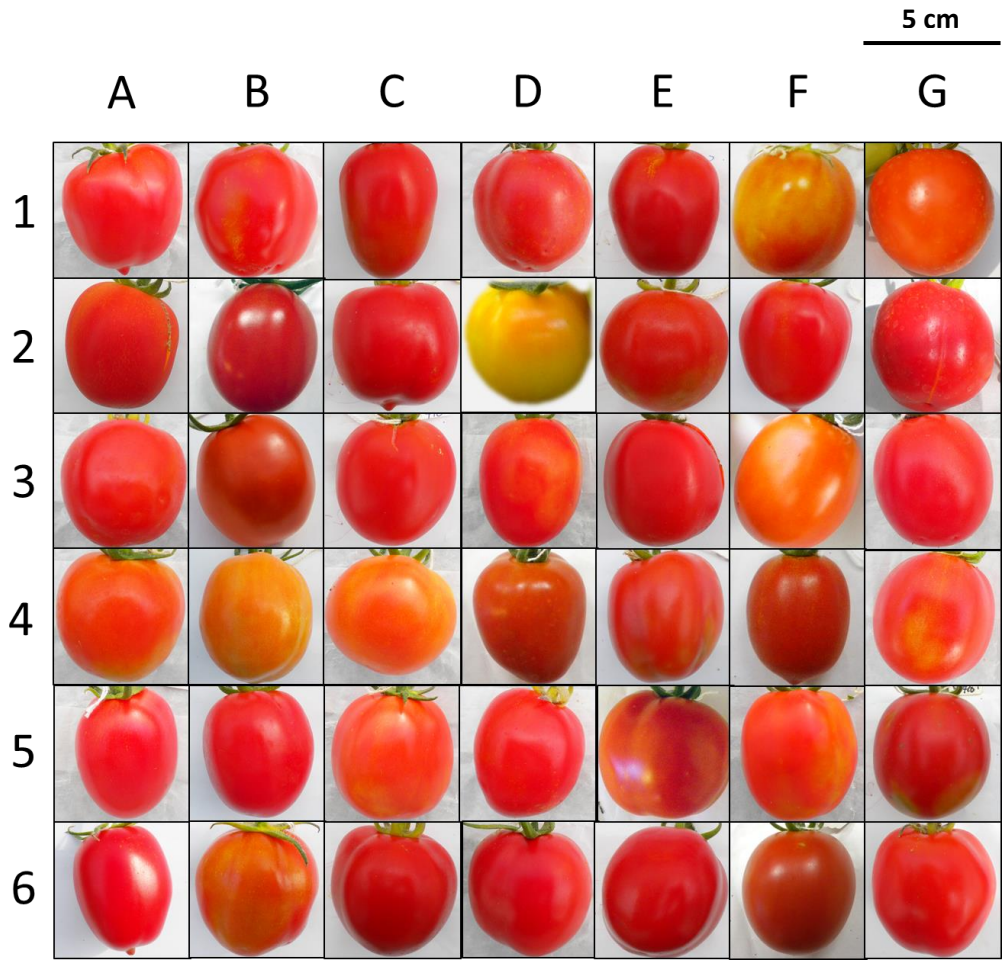
Heat maps were produced using MultiExperiment 4.0 applying the RNA-seq data which was kindly provided by Je Min Lee and Jim Giovannoni, USDA Robert W Holley Centre, Cornell University (http://ted.bti.cornell.edu/). False colour imaging was generated based on relative expression data, which was generated by the fold change of RPKM value of each gene compared to M82. The expression value of M82 was set as 1. For one set of analyses, the published RNA-seq data from leaves of the IL population were used (Lockhart, 2013). The genes selected for analysis were consistent with the annotation of the carotenoid biosynthetic pathway in the tomato genome consortium paper, as well as the abbreviated names for these genes (Tomato Genome Consortium, 2012).

3.3 Results

3.3.1 Phenotypic QTLs related to carotenoid biosynthesis were identified in tomatoes from the introgression lines.

Due to the differential fruit characteristics between two parental lines, *S. pennellii* and *S. lycopersicum*, the fruits displayed divergent phenotypes across the 76 lines, such as size, shape and colour, which were recorded by photography of the fruits representative of 49 lines in the population (Figure 3.2). The major variable phenotype indicating different levels of carotenoids in the ILs is fruit colour, which was observed in the different ILs of the population (Figure 3.2). The colour of ripening tomato fruit is determined largely by the accumulation of two major carotenoids, lycopene and β-carotene, and fruit colour is closely related to the developmental stage of fruit. QTLs associated with the contents of carotenoids and fruit colour have been identified using different introgression lines, including *S. pennellii* and *S. peruvianum*, which prove the association between these two factors (Ronen et al., 1999; Rousseaux et al., 2005; Fulton et al., 1997). However, all the measurements in these studies were undertaken at the 'red' stage, and no specific time point were set to capture the dynamic process of the carotenoid accumulation. Therefore, to make sure the colour phenotypes were representative of carotenoid production, a time point 10 days after breaker was selected for photography of the IL lines to capture the diversity of colour phenotypes.

Compared to M82 (G6), IL2-1 (F1), IL3-2 (D2) and IL6-3 (F3) were obviously less red-coloured (Figure 3.2). Among all the ILs, IL3-2 stood out with a yellow-fruit phenotype even when fruit were fully ripe. Similarly, IL6-3 had orange fruit because it carries a *cis*-eQTL of *SlCYCB* encoding lycopene-β-cyclase from *S. pennellii*, which is the step converting red-coloured lycopene to orange-coloured β-carotene (Rousseaux et al., 2005). This gene is switched off in cultivated tomato during ripening while remaining very active in fruit of *S.pennellii*. Both IL3-2 and IL6-3 have been reported to be associated with carotenoid accumulation (Rousseaux et al., 2005), but IL2-1 has not been identified before as associated with



	A	B	C	D	E	F	G
1	IL1-1	IL1-1-2	IL1-1-3	IL1-3	IL1-4	IL2-1	IL2-1-1
2	IL2-3	IL2-4	IL2-6-5	IL3-2	IL3-3	IL3-4	IL4-1-1
3	IL4-3	IL5-1	IL5-3	IL5-4	IL6-1	IL6-3	IL7-1
4	IL7-4	IL7-4-1	IL7-5	IL8-2	IL8-2-1	IL8-3	IL8-3-1
5	IL9-2	IL9-2-5	IL9-3-1	IL9-3-2	IL10-2	IL10-2-2	IL10-3
6	IL11-2	IL11-3	IL11-4	L12-1	IL12-1-1	IL12-3	M82

Figure 3.2 Photographs of the fruits of 41 selected ILs and M82 at 10 days after breaker. The picture table (above) and the IL name table (below) were corresponding with each other in position. Colums and rows were represented by letters (A to G) and numbers (1 to 6) separately. The pictures were taken at 10 days after breaker. The tomato plants were grown in the greenhouse at John Innes Centre.

carotenoid biosynthesis. At 10 days after breaker, fruit of IL2-1 (F1) showed a yellowish phenotype when the fruit of most other lines had reached the red stage, including its subline IL 2-1-1 (G1), marking IL 2-1 as carrying a phenotypic QTL for fruit colour (Figure 3.2). This observation suggested that the difference in fruit colour in IL2-1 compared to M82 might be associated with lower content of lycopene, an observation that warranted further investigation.

3.3.2 Transcriptome profiling and global eQTL analysis of the carotenoid biosynthetic pathway.

There are more than 40 genes annotated as belonging to the carotenoid biosynthetic pathway in the tomato genome (Table 3.1), including the paralogs of some biosynthetic genes, reflecting the complexity of the genetic basis of this pathway (Tomato Genome Consortium, 2012). All the carotenoid biosynthetic genes listed in the Table 3.1 were consistent with the annotation in the tomato genome, as well as the corresponding abbreviated gene names (Tomato Genome Consortium, 2012). The genome of *S. pennellii* was fully annotated by *de novo* identification followed by whole-genome annotation (Bolger et al., 2014). With the resequenced genome of *S. lycopersicum* cv. M82 using *S. lycopersicum* cv. Heinz as a reference (Bolger et al., 2014), orthologs of the structural genes in carotenoid biosynthetic pathway between *S. pennellii* and M82 were clearly distinguished and mapped to the corresponding IL line (s) (Table 3.1) . Some of the annotated genes had been characterised by identification of specific mutants, whereas the functions of the other genes remain to be fully elucidated. To gain a comprehensive understanding of the genetic basis of the entire pathway, genome-wide eQTL analysis was undertaken to observe the distribution of the structural genes among the 76 lines (Figure 3.3). The number of carotenoid biosynthetic genes in each interval defining each IL is not determined by the size of the introgressed region. For example, IL12-3, containing the highest number of genes in the replaced segment, contains only one structural gene associated with carotenoid biosynthesis, *SlZISO*. IL1-1, the IL with the largest introgressed fragment, is one of the lines harbouring the most carotenoid biosynthetic genes, most likely because of the tandem duplication of *CCD* genes, *SlCCD1A* (*Solyc01g087250*) and *SlCCD1B* (*Solyc01g087260*),

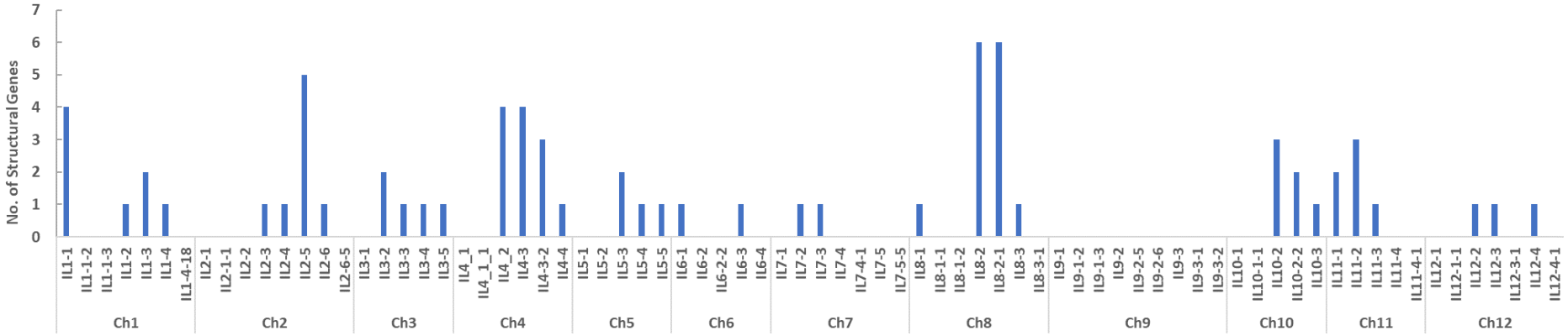


Figure 3.3 The distribution of carotenoid biosynthetic genes annotated in the tomato genome throughout *S. lycopersicum* x *S. pennellii* introgression lines. The *y* axis shows the number of the structural genes in the carotenoid biosynthetic pathway located in the introgressed region of each IL and the *x* axis represents the individual ILs grouped by chromosomes.

similar to IL8-2, which contains two *CCD* genes, and two *NCED* genes from *S. pennellii*. The numbers of structural genes in carotenoid biosynthesis in each IL were not evenly distributed. There were several regions which lacked biosynthetic genes, defined as structural gene-free regions, for example, the first half of chromosome 2, the second half of chromosome 7 and the entirety of chromosome 9 (Figure 3.3). These regions and the respective IL lines that carried them were the best candidates to identify *trans*-eQTL hotspots controlling carotenoid production, because effects caused by variation in the activity/expression of the biosynthetic genes, could be discounted.

3.3.3 Genetic regulation of transcriptional responses associated with isoprenoid biosynthesis in fruit.

The carotenoid biosynthetic pathway produces a wide range of compounds serving diverse biological functions participating in many physiological processes, and linked to several other metabolic pathways by sharing precursors produced via the MEP pathway, such as tocopherol biosynthesis. Even though the catalytic steps in the pathway have been very well-characterised in tomato, it would be difficult to develop any understanding of the underlying regulatory networks if the pathway were analysed in its entirety. The situation may be complicated even more by the presence of genes encoding isoforms catalysing specific biosynthetic steps. There are more than 40 genes encoding proteins involved in carotenoid biosynthesis, including isoforms, in tomato and there is much more redundancy in these metabolic pathways in tomato than in *Arabidopsis*. Two enzymatic steps create bottlenecks in the carotenoid biosynthetic pathway. The first rate-limiting step involves the condensation of two molecules of GGPP to form phytoene, catalysed by phytoene synthase, which is regarded as a key control point for the whole pathway. The second key step is the branch point involving the formation of β-carotene or δ-carotene from lycopene catalysed by lycopene β-cyclase and lycopene ε-cyclase respectively. These steps (or these enzymes) are switched off during ripening of domesticated tomato fruit, resulting in the massive accumulation of lycopene which typifies red-fruited tomato species, whereas both enzymes

are highly active in green-fruited wild relatives, including *S.pennellii*, which has undetectable levels of lycopene in its fruit. Therefore, to better understand the

Table 3.1 Genes involved carotenoid biosynthetic pathway annotated in Tomato genome

Gene ID	Gene name abbreviated	Metabolic Pathway	IL Location		
Solyc01g067890	DXS1	Isoprenoid biosynthesis	IL1-1		
Solyc08g066950	DXS2	Isoprenoid biosynthesis	IL11-1	IL11-2	
Solyc11g010850	DXS3	Isoprenoid biosynthesis	IL8-2-1	IL8-2	
Solyc03g114340	DXR	Isoprenoid biosynthesis	IL3-3	IL3-4.B	
Solyc01g102820	ISPD or CMS	Isoprenoid biosynthesis	IL1-3		
Solyc01g009010	ISPE or CMK	Isoprenoid biosynthesis	IL1-1		
Solyc08g081570	ISPF or MCS	Isoprenoid biosynthesis	IL8-3		
Solyc11g069380	ISPG or HDS	Isoprenoid biosynthesis	IL11-3		
Solyc01g109300	HDR	Isoprenoid biosynthesis	IL1-4		
Solyc04g056390	IPP1	Isoprenoid biosynthesis	IL4-2	IL4-3	IL4-3-2
Solyc05g055760	IPP2	Isoprenoid biosynthesis	IL5-5		
Solyc08g075390	IPI	Isoprenoid biosynthesis	IL8-2-1	IL8-2	
Solyc11g011240	GGPPS1	Isoprenoid biosynthesis	IL11-1	IL11-2	
Solyc04g079960	GGPPS2	Isoprenoid biosynthesis	IL4-3	IL4-4	
Solyc02g085700	GGPPS3	Isoprenoid biosynthesis	IL2-5		
Solyc02g085710	GGPPS4	Isoprenoid biosynthesis	IL2-5		
Solyc02g085720	GGPPS5	Isoprenoid biosynthesis	IL2-5		
Solyc03g031860	PSY1	Lycopene biosynthesis	IL3-2.B		
Solyc02g081330	PSY2	Lycopene biosynthesis	IL2-3.B	IL2-4	IL2-5
Solyc03g123760	PDS	Lycopene biosynthesis	IL3-5		
Solyc12g098710	ZISO	Lycopene biosynthesis	IL12-3	IL12-4	
Solyc01g097810	ZDS	Lycopene biosynthesis	IL1-3		
Solyc10g081650	CrtISO	Lycopene biosynthesis	IL10-2	IL10-2-2	
Solyc05g010180	CrtISO-like	Lycopene biosynthesis	gap		
Solyc11g011990	PTOX	Lycopene catabolism	IL11-2		
Solyc06g074240	CYCB	Lycopene catabolism	IL6-3		
Solyc04g040190	LCYB1	Lycopene catabolism	IL4-2	IL4-3	IL4-3-2
Solyc10g079480	LCYB2	Lycopene catabolism	IL10-2	IL10-2-2	
Solyc12g008980	LCYE	Lycopene catabolism	IL12-2.A		
Solyc04g051190	HYDB or CYP97A29	Lycopene catabolism	IL4-2		
Solyc10g083790	HYDE or CYP97C11	Lycopene catabolism	IL10-2	IL10-3	
Solyc05g016330	CYP97B2	Lycopene catabolism	IL5-3		

Solyc02g090890	ZEP	Lycopene catabolism	IL2-5	IL2-6	
Solyc04g050930	VDE	Lycopene catabolism	IL4-2	IL4-3	IL4-3-2
Solyc06g036260	CHY1	Lycopene catabolism	IL6-1		

(Continued from last page)

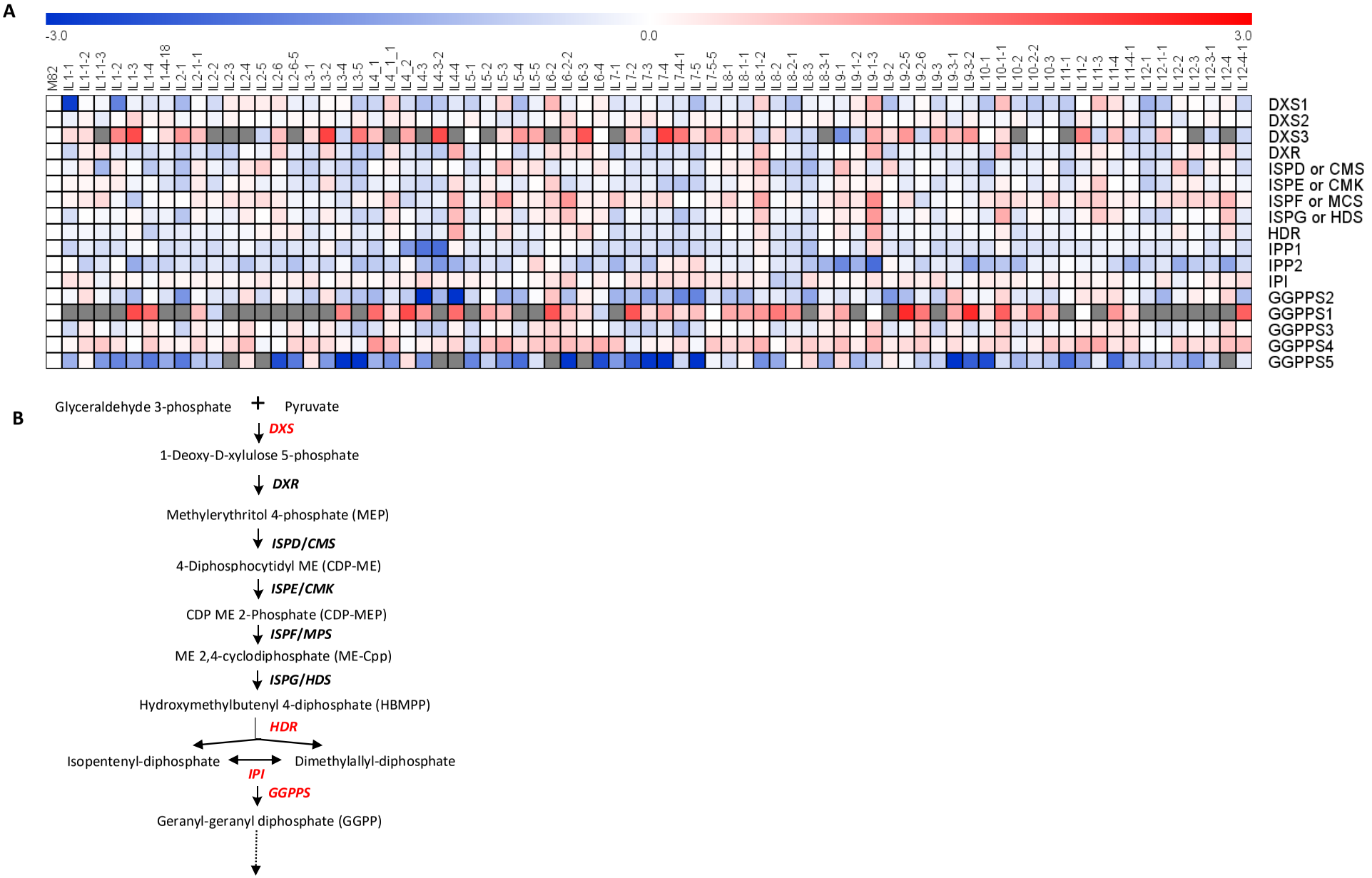
Gene ID	Gene name abbreviated	Metabolic Pathway	IL Location		
Solyc03g007960	CHY2	Lycopene catabolism	IL3-2.A		
Solyc01g087250	CCD1A	Lycopene catabolism	IL1-1		
Solyc01g087260	CCD1B	Lycopene catabolism	IL1-1		
Solyc08g066650	CCD8	Lycopene catabolism	IL8-2-1	IL8-2	
Solyc08g066720	CCD-like	Lycopene catabolism	IL8-2-1	IL8-2	
Solyc07g056570	NCED	Lycopene catabolism	IL7-2	IL7-3	
Solyc08g016720	NCED2	Lycopene catabolism	IL8-1		
So					

Table 3.1 Genes involved carotenoid biosynthetic pathway annotated in Tomato genome. DXS, 1-deoxyxylulose-5-phosphate synthase; DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; CMS, 4-(cytidine 5'-diphospho)-2-C-methyl-D synthase; CMK= 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase; MCS, ME-cPP synthase; HDS, hydroxymethylbutenyl 4-diphosphate [HMBPP] synthase; HDR= 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase; IPI, isopentenyl diphosphate isomerases; GGPPS, geranyl-geranyl diphosphate (GGPP) synthase; PSY, phytoene synthase; PDS, phytoene desaturase; ZDS, ζ -carotene desaturase; Z-ISO, ζ-carotene isomerase; CRTISO, carotenoid isomerase; LYCB, lycopene β-cyclase; LYCE, lycopene ε-cyclase; HYD-B, β-carotene hydroxylases; HYD-E, ε-carotene hydroxylases; CYP, Cytochrome P450 hydroxylases; CHY, non-heme hydroxylases; ZEP, zeaxanthin epoxidase; VDE, violaxanthin deepoxidase; PTOX, alternative oxidase; NCED, CCD, carotenoid cleavage enzymes.

transcriptional associations between the genes and to identify the *cis*- and *trans*-eQTLs governing the entire pathway or parts of the pathway, the genes encoding enzymes involved in the carotenoid biosynthetic pathway were classified into three subgroups: isoprenoid biosynthesis, lycopene biosynthesis and lycopene degradation and analysed separately.

Three heatmaps were generated corresponding to subgroups of genes encoding enzymes of the separate pathways described above, based on the fold change of the transcript abundance in fruit of each IL compared to transcript levels in fruit of cultivated M82 tomatoes (Figure 3.4, 3.5 and 3.6).

The rapid accumulation of lycopene in tomato fruit is closely associated with a very active isoprenoid biosynthetic pathway which supplies building blocks for the synthesis of lycopene, a linear C40 molecule. The isoprenoid biosynthetic pathway supplies precursors not only for lycopene biosynthesis, but also for the synthesis of gibberellins, tocopherols, isoprenes and diterpenes, reflecting its central role in the terpenoid secondary metabolic network. This explains some of the redundancy observed in genes encoding enzymes in isoprenoid biosynthetic pathway. There are three *SlDXS* genes (*Solyc01g067890*, *Solyc08g066950*, *Solyc11g010850*), three *SlIPI* genes (*Solyc04g056390*, *Solyc05g055760*, *Solyc08g075390*) and five *SlGGPPS* genes (*Solyc02g085700*, *Solyc02g085710*, *Solyc02g085720*, *Solyc04g079960*, *Solyc11g011240*) annotated in the tomato genome (Tomato Genome Consortium, 2012). However, it is not difficult to identify the different functional genes active in tomato fruit from expression profiling combined with eQTL mapping. As discussed before, in this structured introgression line population, differential 'local' expression of any particular gene is the result either of polymorphism(s) within the gene itself between *S. lycopersicum* and *S. pennellii* (*cis*-eQTL), or because other genes in the same replacement region influence the transcript levels of the specific gene. Among the three genes encoding isoforms of SlDXS, 'local' expression of *Solyc01g067890*, also named as *SlDXS1*, was dramatically decreased in IL1-1 and IL1-2, which was not observed for the other two *SlDXS* genes (Figure 3.4). The detection of this *cis*-eQTL is consistent with studies that showed that expression of *SlDXS1* was positively correlated with the formation of


A
-3.0
0.0
3.0
M82
IL1-1
IL1-1-2
IL1-1-3
IL1-2
IL1-3
IL1-4
IL1-4-18
IL2-1
IL2-1-1
IL2-2
IL2-3
IL2-4
IL2-5
IL2-6
IL2-6-5
IL3-1
IL3-2
IL3-4
IL3-5
IL4_1
IL4_1_1
IL4_2
IL4-3
IL4-3-2
IL4-4
IL5-1
IL5-2
IL5-3
IL5-4
IL5-5
IL6-2
IL6-2-2
IL6-3
IL6-4
IL7-1
IL7-2
IL7-3
IL7-4
IL7-4-1
IL7-5
IL7-5-5
IL8-1
IL8-1-1
IL8-1-2
IL8-2
IL8-2-1
IL8-3
IL8-3-1
IL9-1
IL9-1-2
IL9-1-3
IL9-2
IL9-2-5
IL9-2-6
IL9-3
IL9-3-1
IL9-3-2
IL10-1
IL10-1-1
IL10-2
IL10-2-2
IL10-3
IL11-1
IL11-2
IL11-3
IL11-4
IL11-4-1
IL12-1
IL12-1-1
IL12-2
IL12-3
IL12-3-1
IL12-4
IL12-4-1
DXS1
DXS2
DXS3
DXR
ISPD or CMS
ISPE or CMK
ISPF or MCS
ISPG or HDS
HDR
IPP1
IPP2
IPI
GGPPS2
GGPPS1
GGPPS3
GGPPS4
GGPPS5
B
Glyceraldehyde 3-phosphate
+
Pyruvate
DXS
1-Deoxy-D-xylulose 5-phosphate
DXR
Methylerythritol 4-phosphate (MEP)
ISPD/CMS
4-Diphosphocytidyl ME (CDP-ME)
ISPE/CMK
CDP ME 2-Phosphate (CDP-MEP)
ISPF/MPS
ME 2,4-cyclodiphosphate (ME-Cpp)
ISPG/HDS
Hydroxymethylbutenyl 4-diphosphate (HBMPP)
HDR
Isopentenyl-diphosphate
Dimethylallyl-diphosphate
IPI
GGPPS
Geranyl-geranyl diphosphate (GGPP)

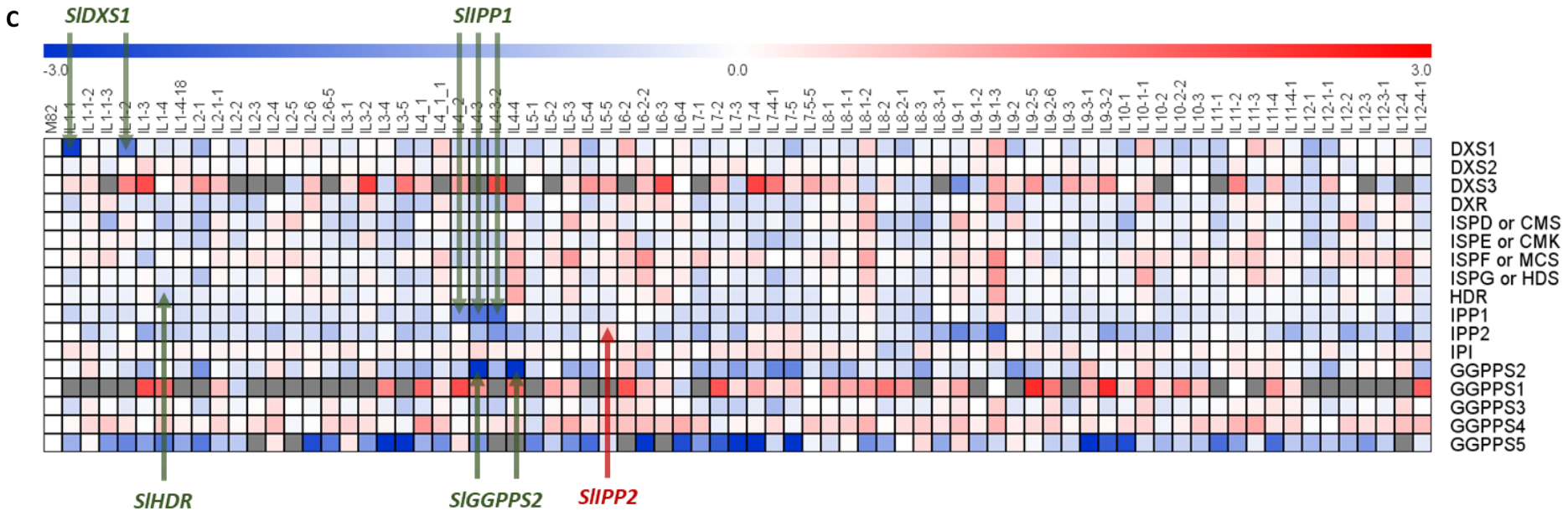
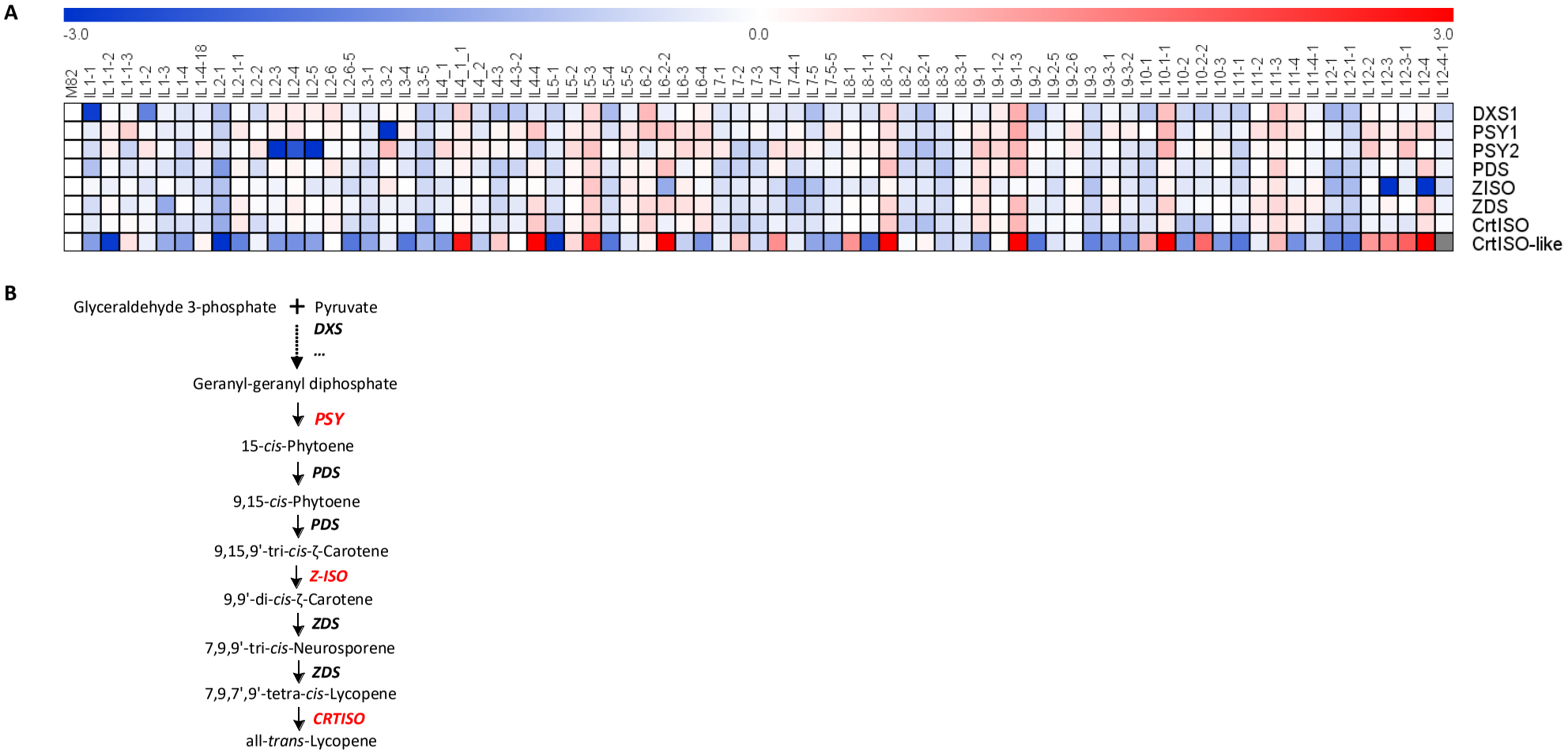


Figure 3.4 Identification of *cis*-eQTLs for isoprenoid biosynthetic pathway in tomato fruit. A, the heatmap of transcript abundance of isoprenoid biosynthetic genes in fruit compared to *S. lycopercisum* cv. *M82*. The expression values of genes in M82 was set as 1 (in log2). B, the isoprenoid biosynthetic pathway in tomato fruit. The genes marked with red colour were those identified with *cis*-eQTL (s) in the analysis. C, structural genes in isoprenoid biosynthetic pathway, as well as their *cis*-eQTL (s) indicated by arrows. Red arrow, *cis*-eQTL with increased expression of the biosynthetic gene; Green arrow (s), *cis*-eQTL (s) with reduced expression of the biosynthetic gene (s). See Table 3.1 for abbreviations. Gray means no reads of this gene were detected in the RNA-seq profiling.

lycopene during fruit ripening (Lois et al., 2000), whereas the transcripts of *Solyc11g010850* (*SlDXS2)*, were abundant in young leaves, trichomes and petals but not in fruit (Paetzold et al., 2010, Tomato eFP Browser (http://bar.utoronto.ca/efp_tomato/cgi-bin/efpWeb.cgi). Therefore, the identification of the region overlapping in IL1-1 and IL1-2 as the *cis*-eQTL of *SlDXS1* (based on the transcriptome data generated from fruit) further validated the specific function of DXS1 during fruit ripening. Similarly, IL4-3-2 was identified as the *cis*-eQTL of *SlIPP1* (*Solyc04g056390*) and the overlapping region between IL4-3 and IL4-4 was characterised as the *cis*-eQTL of *SlGGPPS2* (*Soly04g079960*). Unlike *SlIPP1*, the *cis*-eQTL which was identified in IL4-3 and IL4-4 as having the lowest expression levels of *SlIPP1* in fruit, IL5-5 was characterised as the *cis*-eQTL of *SlIPP2* (*Solyc05g055760*) with the most abundant transcript levels. According to the Tomato eFP Browser, *SlIPP1* is mainly expressed in ripening fruit, whereas the highest expression level of *SlIPP2* is found in root, which partially explains the differences in expression between the *cis*-eQTLs of these two isogenes. These data identified DXS1, IPP1 and GGPPS2 as the functional isoforms in fruit, consistent with the results of earlier studies (Ament et al., 2006). The lowest transcript abundance for *SlHDR*, encoding a hydroxymethylbutenyl diphosphate reductase, which catalyses the last step of isoprenoid biosynthesis, coincided with the physical position of the gene, in IL1-4 (approximately 40 cM). HDR contributes to the formation of the isoprenoid precursor of the carotenoid biosynthesis, geranylgeranyl diphosphate (GGPP), by condensing isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), and has been proved to play a major role in controlling the production of DXOP/MEP-derived precursors for plastid isoprenoid biosynthesis (Botella-Pavía et al., 2004). Overall five *cis*-eQTLs were found for genes in the isoprenoid biosynthetic pathway (Figure 3.4).


A
-3.0
0.0
3.0
M82
IL1-1
IL1-1-2
IL1-1-3
IL1-2
IL1-3
IL1-4
IL1-4-18
IL2-1
IL2-1-1
IL2-2
IL2-3
IL2-4
IL2-5
IL2-6
IL2-6-5
IL3-1
IL3-2
IL3-4
IL3-5
IL4_1
IL4_1_1
IL4_2
IL4-3
IL4-3-2
IL4-4
IL5-1
IL5-2
IL5-3
IL5-4
IL5-5
IL6-2
IL6-2-2
IL6-3
IL6-4
IL7-1
IL7-2
IL7-3
IL7-4
IL7-4-1
IL7-5
IL7-5-5
IL8-1
IL8-1-1
IL8-1-2
IL8-2
IL8-2-1
IL8-3
IL8-3-1
IL9-1
IL9-1-2
IL9-1-3
IL9-2
IL9-2-5
IL9-2-6
IL9-3
IL9-3-1
IL9-3-2
IL10-1
IL10-1-1
IL10-2
IL10-2-2
IL10-3
IL11-1
IL11-2
IL11-3
IL11-4
IL11-4-1
IL12-1
IL12-1-1
IL12-2
IL12-3
IL12-3-1
IL12-4
IL12-4-1
DXS1
PSY1
PSY2
PDS
ZISO
ZDS
CrtISO
CrtISO-like
B
Glyceraldehyde 3-phosphate + Pyruvate
DXS
...
Geranyl-geranyl diphosphate
PSY
15-cis-Phytoene
PDS
9,15-cis-Phytoene
PDS
9,15,9'-tri-cis-ζ-Carotene
Z-ISO
9,9'-di-cis-ζ-Carotene
ZDS
7,9,9'-tri-cis-Neurosporene
ZDS
7,9,7',9'-tetra-cis-Lycopene
CRTISO
all-trans-Lycopene

C

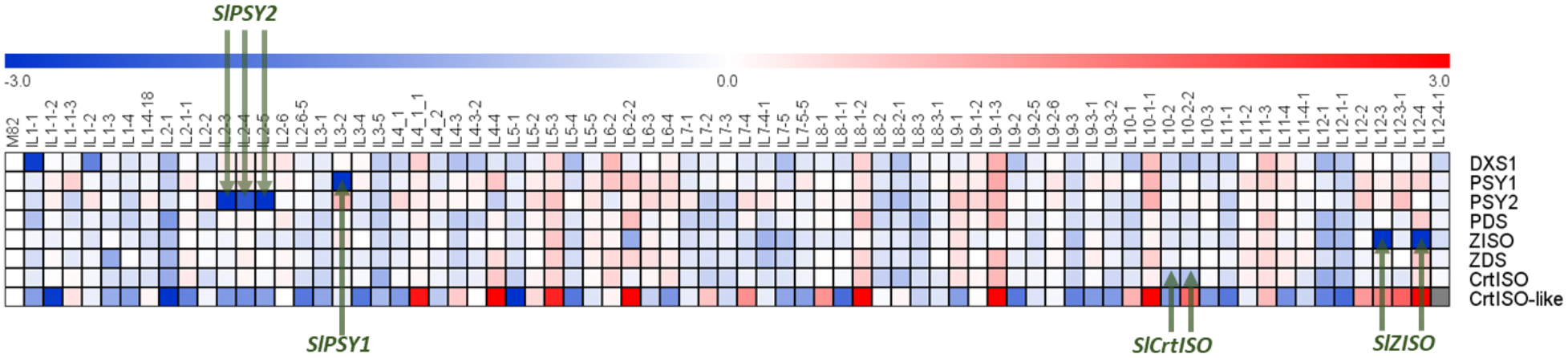


Figure 3.5 Identification of *cis*-eQTLs for lycopene biosynthetic pathway in tomato fruit. A, the heatmap of transcript abundance of lycopene biosynthetic genes in fruit compared to *S. lycopercisum* cv. *M82*. The expression values of genes in M82 was set as 1 (in log2). B, the lycopene biosynthetic pathway in tomato fruit. The genes marked with red colour were those identified with *cis*-eQTL (s) in the analysis. C, structural genes in lycopene biosynthetic pathway, as well as their *cis*-eQTL (s) indicated by arrows. Green arrow (s), *cis*-eQTL (s) with reduced expression of the biosynthetic gene (s)**.** See Table 3.1 for abbreviations.

3.3.4 Genetic regulation of transcriptional responses associated with lycopene biosynthesis in fruit.

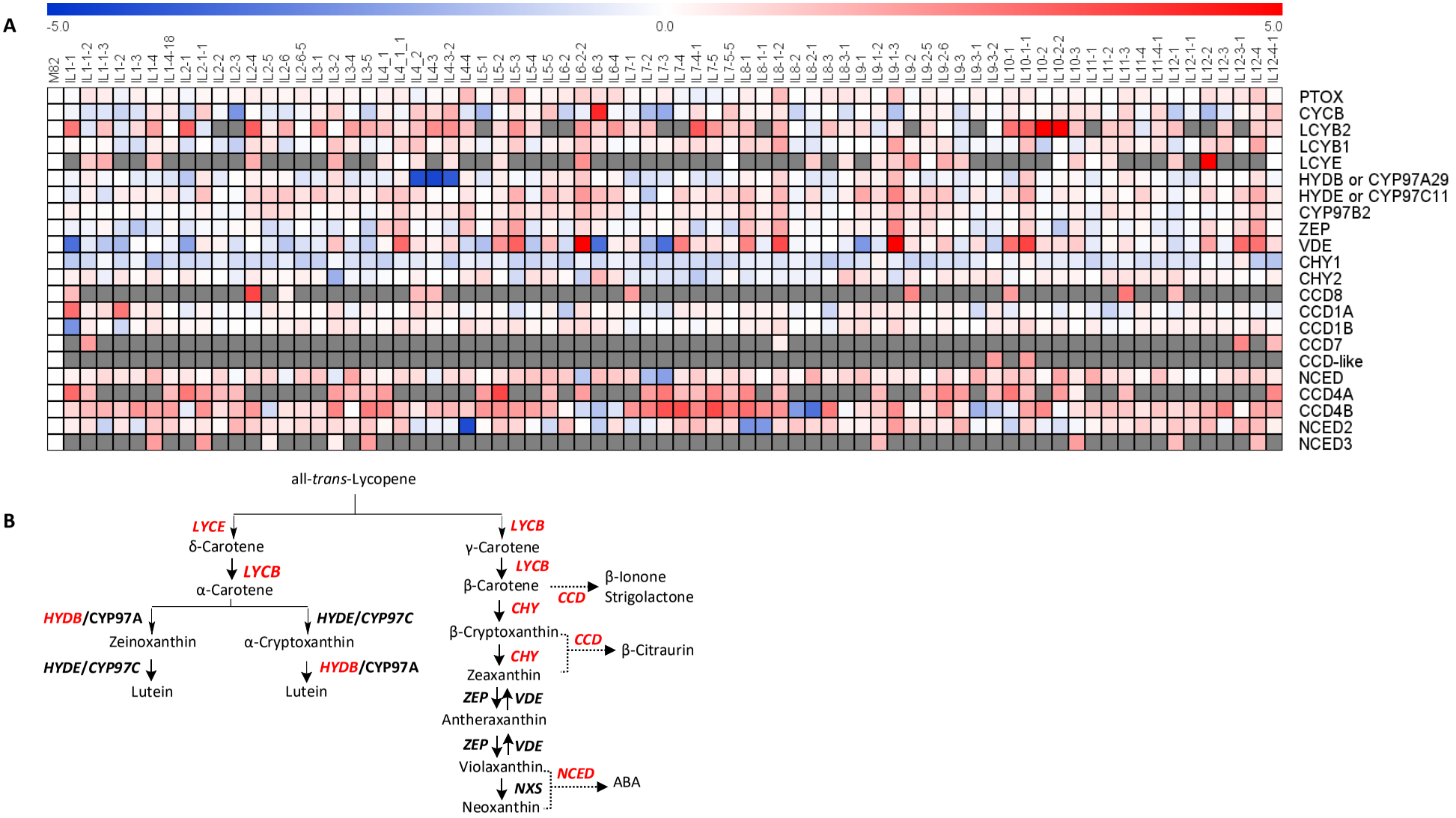
According to the heatmap, which represented the fold changes of the expression level of the carotenoid biosynthetic genes in each IL compared to M82 (Figure 3.5), three *cis*-eQTLs could be mapped for the genes of the lycopene biosynthetic pathway. Phytoene synthase (PSY) has been reported to be the key step in lycopene biosynthesis, controlling carbon flux leading to carotenoid production (Bou-Torrent et al., 2015). PSY has three isoforms in tomato. Two *cis*-eQTLs were identified based on the significantly reduced expression of two paralogs of *PSY* in fruit. Transcript abundance of *SpPSY1*, located in IL3-2, was less than 2% of that in M82 or other ILs mainly because the polymorphisms of *PSY1* in *S.pennellii* result in a non-functional PSY protein (Kachanovsky et al., 2012). *SlPSY2*, a second gene encoding an isoform of PSY, is expressed in all tissues of tomato (Fraser et al., 1999). *cis*-eQTLs for *PSY2* were identified in IL2-3, IL2-4 and IL2-5 (Figure 3.5, Table 3.1). No *cis*-eQTL for the IL harbouring *SpPSY3* was identified in fruit, but *SlPSY3* has been reported to be expressed mainly in roots contributing to ABA and strigolactone production, in response to environmental stresses (Li et al., 2008a). Finally, the overlapping region between IL12-3 and IL12-4 was identified as a *cis*-eQTL due to the significantly reduced expression of *SlZISO* (Table 3.1).

In tomato, the lycopene biosynthetic pathway is tightly regulated through tissue-specific and developmental stage-specific expression of structural genes. The relatively low level of functional redundancy between structural genes and the differential expression of the genes encoding isoforms (such as the *SlPSYs*) has allowed the characterisation of the enzymatic steps active in specific tissues based mainly on mutant identification. The roles of *SlPSY1* and *SlCRTISO* in lycopene production in fruit were defined by the discovery of mutants; *yellow flesh* (locus *r)* and *tangerine* (locus *t*) respectively (Isaacson, 2002; Fray and Grierson, 1993).

IL3-2 was identified as a phenotypic QTL related to lycopene accumulation as shown in Figure 3.2 (D2). IL3-2 presents a very similar phenotype to that observed in *yellow flesh* with a very low red colouration in ripe fruit compared to M82 and other IL lines. The association between the large-phenotypic effects and the massive alterations in gene expression make it possible to use such *cis*-eQTLs to identify the polymorphisms underlying the phenotypic changes in natural mutants. Such large-effect *cis*-eQTLs have been identified in many studies, and several of the genes underpinning such QTLs have been cloned, prior the emergence of global eQTL analysis (Jiang et al., 2000; Wang et al., 2005; Asíns, 2002). The associations between *cis*-eQTLs and phenotypic QTLs suggest that identification of potential large-effect *cis*-eQTLs is a good starting point for characterising the genetic basis of phenotypic QTLs. The genes underpinning *cis*-eQTLs in fruit comprise the first list of gene candidates for further analysis with respect to understanding the regulation of lycopene biosynthesis in tomato fruit.

3.3.5 Genetic regulation of transcriptional responses associated with lycopene catabolism in fruit.

Lycopene accumulation in tomato fruit is the outcome of both its increased biosynthesis and reduced degradation. Cyclization of lycopene serves as a branch point in the carotenoid biosynthetic pathway, but is almost completely switched off in ripe fruit of *S.lycopersicum*, while it remains very active in *S.pennellii* fruit. There are two pathways which branch from lycopene, of which one leads to the production of β-carotene, xanthophylls and carotenoid derivatives, and the other is responsible for the production of α-carotene and lutein. Unlike the *cis*-eQTLs identified for the isoprenoid biosynthetic pathway and the lycopene biosynthetic pathway, where structural genes are expressed at significantly lower levels, the *cis*-eQTLs containing the lycopene cyclase genes were characterised by their much higher transcript levels compared to M82. *SlCYCB* (*Solyc06g074240*), *SlLCYB2* (*Solyc10g079480*) and *SlLCYE* (*Solyc12g008980*) were 19, 384 and 4218 times more highly expressed in their corresponding *cis*-eQTLs, IL6-3, IL10-2 and IL12-2 than M82 (Figure 3.6). *SlCYCB* (*Solyc06g074240*) was characterised by mapped-based cloning in tomato colour mutants, *old-gold* (*og*) and *Beta* (Ronen et al., 2000a). Similarly, *SlLCYE* (*Solyc12g008980*)


A
-5.0
0.0
5.0
M82
IL1-1
IL1-1-2
IL1-1-3
IL1-2
IL1-3
IL1-4
IL1-4-18
IL2-1
IL2-1-1
IL2-2
IL2-3
IL2-4
IL2-5
IL2-6
IL2-6-5
IL3-1
IL3-2
IL3-4
IL3-5
IL4-1
IL4-1-1
IL4-2
IL4-3
IL4-3-2
IL4-4
IL5-1
IL5-2
IL5-3
IL5-4
IL5-5
IL6-2
IL6-2-2
IL6-3
IL6-4
IL7-1
IL7-2
IL7-3
IL7-4
IL7-4-1
IL7-5
IL7-5-5
IL8-1
IL8-1-1
IL8-1-2
IL8-2
IL8-2-1
IL8-3
IL8-3-1
IL9-1
IL9-1-2
IL9-1-3
IL9-2
IL9-2-5
IL9-2-6
IL9-3
IL9-3-1
IL9-3-2
IL10-1
IL10-1-1
IL10-2
IL10-2-2
IL10-3
IL11-1
IL11-2
IL11-3
IL11-4
IL11-4-1
IL12-1
IL12-1-1
IL12-2
IL12-3
IL12-3-1
IL12-4
IL12-4-1
PTOX
CYCB
LCYB2
LCYB1
LCYE
HYDB or CYP97A29
HYDE or CYP97C11
CYP97B2
ZEP
VDE
CHY1
CHY2
CCD8
CCD1A
CCD1B
CCD7
CCD-like
NCED
CCD4A
CCD4B
NCED2
NCED3
B
all-trans-Lycopene
LYCE
δ-Carotene
LYCB
α-Carotene
HYDB/CYP97A
HYDE/CYP97C
Zeinoxanthin
α-Cryptoxanthin
HYDE/CYP97C
HYDB/CYP97A
Lutein
Lutein
LYCB
γ-Carotene
LYCB
β-Carotene
CCD
β-Ionone
Strigolactone
CHY
β-Cryptoxanthin
CCD
β-Citraurin
CHY
Zeaxanthin
ZEP
VDE
Antheraxanthin
ZEP
VDE
Violaxanthin
NCED
NXS
ABA
Neoxanthin

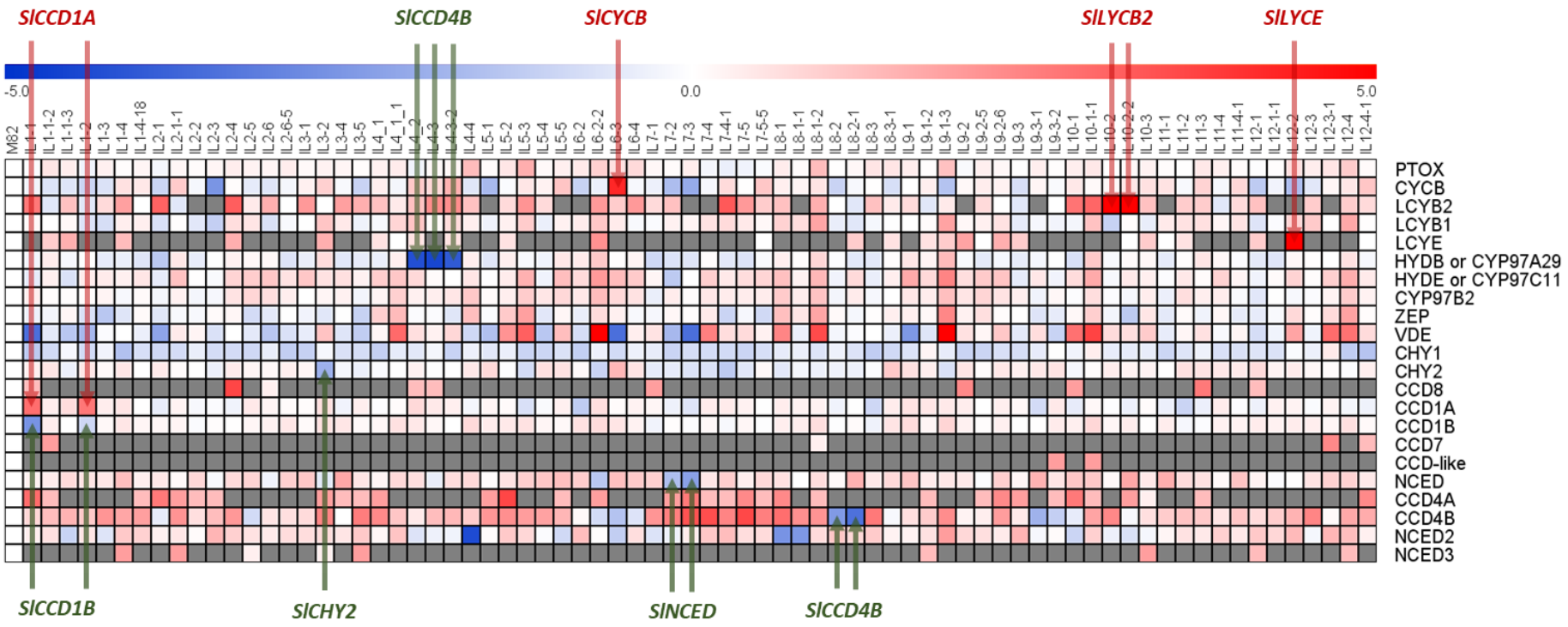


Figure 3.6 Identification of *cis*-eQTLs for lycopene catabolism in tomato fruit. A, the heatmap of transcript abundance of genes involved in lycopene catabolism in fruit compared to *S. lycopercisum* cv. *M82*. The expression values of genes in M82 was set as 1 (in log2). B, the lycopene catabolic pathway in tomato fruit. The genes marked with red colour were those identified with *cis*-eQTL (s) in the analysis. C, structural genes in lycopene catabolism, as well as their *cis*-eQTL (s) indicated by arrows. Red arrow, *cis*-eQTL with increased expression of the biosynthetic gene; Green arrow (s), *cis*-eQTL (s) with reduced expression of the biosynthetic gene (s)**.** See Table 3.1 for abbreviations. Gray means no reads of this gene deteted in the RNA-seq profiling.

was identified by a dominant mutant *Delta*, in which the expression level of *SlLCYE* was more than 30-fold increased accompanied by elevated levels of δ-carotene (Ronen et al., 1999). Previously, the study of the genetic and molecular basis of repressed metabolic pathways has relied largely on the identification of dominant mutants, like *Beta* and *Delta*, containing the mutations in CYCB and LCYE respectively, which are less common naturally than recessive ones. What is more, it is very difficult to identify several unclustered isogenes at the same time, because the phenotypic effect caused by the weak isoform is very likely to be masked by that caused by strong ones. Genome-wide eQTL analysis offers an alternative solution in such situations. The identification of IL6-3, IL10-2 and IL12-2 in one structured intregression population by eQTL mapping showed the potential of ILs for characterising the genetic basis of metabolic pathways, bypassing the requirement for dominant mutants necessary in previous studies. *cis*-eQTLs, IL6-3 and L10-2, containing two lycopene β-cyclase genes were identified in my study (Figure 3.6), showing the advantage of the eQTL approach for understanding the individual contributions of genes encoding different isoforms of enzymes (genetic redundancy) to specific catalytic reactions.

Several other *cis*-eQTLs of lycopene degradation are highlighted in Figure 3.6. IL1-1 and IL1-2 have been identified as containing carotenoid cleavage dioxygenase (CCD) genes arranged in tandem, *SlCCD1A* (*Solyc01g087250*) and *SlCCD1B* (*Solyc01g087260*) (Figure 3.6). It has been reported the transcript level of *CCD1* is correlated with the production of lutein in strawberry (García-Limones et al., 2008). The enzyme activities of these two genes of *SlCCD1* have been analysed in vitro, and they are involved in the degradation of carotenoids, with relaxed substrate specificity targeting carotenoids, acyclic carotenes and apocarotenoids (Ilg et al., 2014). Interestingly this pair of genes showed opposite expression patterns in the same ILs. *SlCCD1A* was upregulated in fruit in those introgressions where the transcript level of *SlCCD1B* was decreased (Figure 3.6). Wei *et al*. have reported that the highest transcript levels of *SlCCD1A* were observed in vegetative tissues and *SlCCD1B* transcripts were most abundant in ripe fruit, but both of the genes are expressed in all tissues of tomato plant (Wei et al., 2016). The opposite changes of expression within this pair of homologs may be associated with potential tissue-specific functions, as *SlCCD1B* is more active compared to *SlCCD1A* in the fruit of cultivated tomato according to Tomato eFP Browser (http://bar.utoronto.ca/efp_tomato/cgi-bin/efpWeb.cgi). Another reason might

be because that the activity of CDD1A of *S. pennellii* is higher than its ortholog in *S.lycopersicum*. The *cis*-eQTL for another CCD gene, *SlCCD4B* (*Solyc08g075490*), IL8-2-1/IL8-2 was identified with less than 20% transcript abundance compared to its expression level in M82 (Figure 3.6). Similar to *SlCCD1A* and *SlCCD1B*, *SlCCD4A* (Solyc08g075480) and *SlCCD4B* are in tandem alignment. However, *SlCCD4A* is exclusively expressed in fully opened flowers, and its expression was not detectable in fruit. CCD4 has been extensively studied due to its function in saffron crocetin biosynthesis from the carotenoid zeaxanthin in the determination of colour of flowers (Frusciante et al., 2014; Brandi et al., 2011; Rodrigo et al., 2013). The distinct identification of *cis*-eQTL clearly revealed the differential expression patterns of *SlCCD4A* and *SlCCD4B* in tomato. β-carotene hydroxylase, HYDB, takes part in the formation of lutein and apocarotenoids, such as zeaxanthin (Galpaz, 2006). The expression level of *HYDB* (*Solyc04g0511980*) was more than 90% reduced in IL4-2, IL4-3 and IL 4-3-2 compared to M82, meaning these three overlapping ILs include strong *cis*-eQTLs of *SlHYDB* (Figure 3.6).

Other *cis*-eQTLs in the lycopene catabolic pathway were identified associated with 9-cisepoxycarotenoid dioxygenase (NCED), which are IL7-2/IL7-3, as well as with β-carotene hydroxylases 2 (CHY2), which is IL3-2 (Figure 3.6).

3.3.6 Six *trans*-eQTL hotspots for lycopene biosynthesis were identified

Many *cis*-eQTLs are easy to identify based on the physical location of structural genes of carotenoid biosynthesis combined with the changes of gene transcript levels in the corresponding introgression line. I searched for *trans*-eQTL hotspots affecting the transcript levels of all, or most structural genes in particular sections of carotenoid metabolism. Such hotspots are likely to identify specific regulators, controlling the expression of a group of genes involved in the same metabolic process. Unlike the *cis*-eQTLs which have large effects on the expression of single genes, *trans*-eQTL hotspots affect a more extensive range of genes but generally have smaller effects, with 10-20% changes in transcript abundance typically (West et al., 2007). Most genes functioning in the same metabolic pathway show

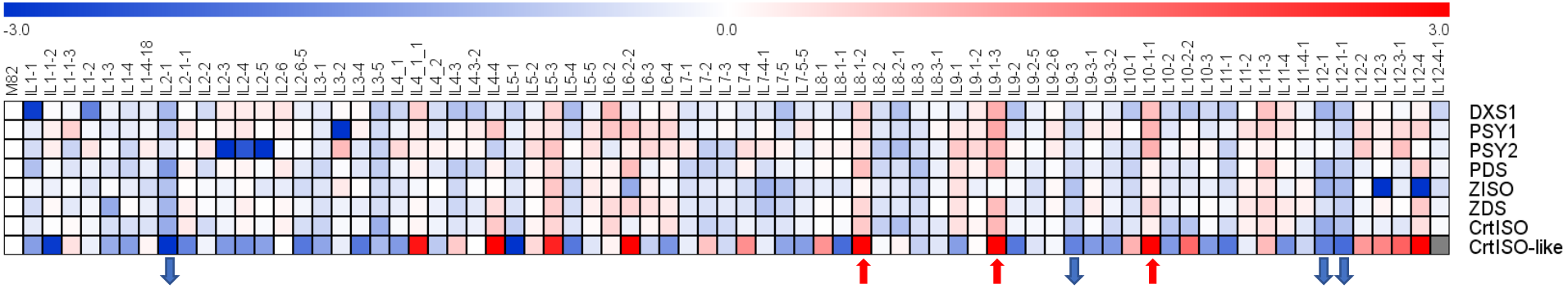


Figure 3.7 The heatmap of transcript abundance of the genes in lycopene biosynthetic pathway compared to *S. lycopercisum* cv. *M82*. The expression values of genes in M82 was set as 1. *Trans*-eQTL hotspots were indicated by arrows. Red arrows, *trans*-eQTL hotspots with activated expression of the genes of lycopene biosynthesis; blue arrows, *trans*-eQTL hotspots with reduced expression of the genes of lycopene biosynthesis

Figure 3.8 Photographs of fruit of IL2-1, compared to M82 and its subline IL2-1-1, at three days after breaker (B+3). Lower red colouration was observed in IL2-1 compared to IL2-1-1 and M82 at the same developmental stage. The scale bar indicates 1 cm.

Figure 3.9 During fruit ripening, red colouration in IL2-1 developed much more slowly than that in IL2-1-1, and fruits were much less coloured at the red-ripe stage (B+10 for IL2-1-1). Time-course photography of IL2-1 and IL2-1-1, at the stages of breaker, B+3, B+7, B+10 and B+15. Upper lanes, whole fruit; lower lanes, cross-sections of the same fruit. The scale bars indicate 1 cm.

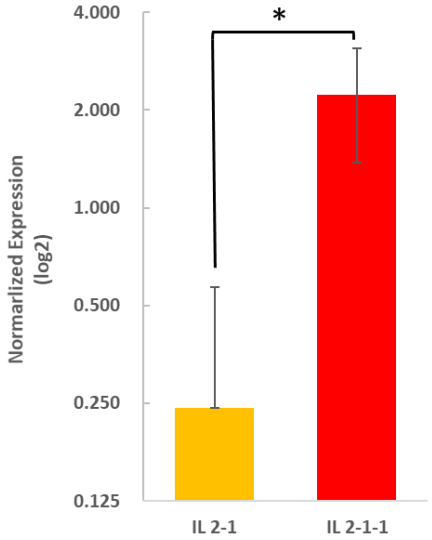


Figure 3.10 Normalized expression of *SlPSY1* in IL2-1 and IL2-1-1 at the stage of B+3. The expression of *SlPSY11* was analysed using quantitative PCR. Expression levels shown is relative to that for *ACTIN* (Solyc03g078400). Statistical analysis was performed using a two-tailed *t* test. * $P < 0.05$. Error bars, s.e.m.; three biological replicates, each with three technical replicates.

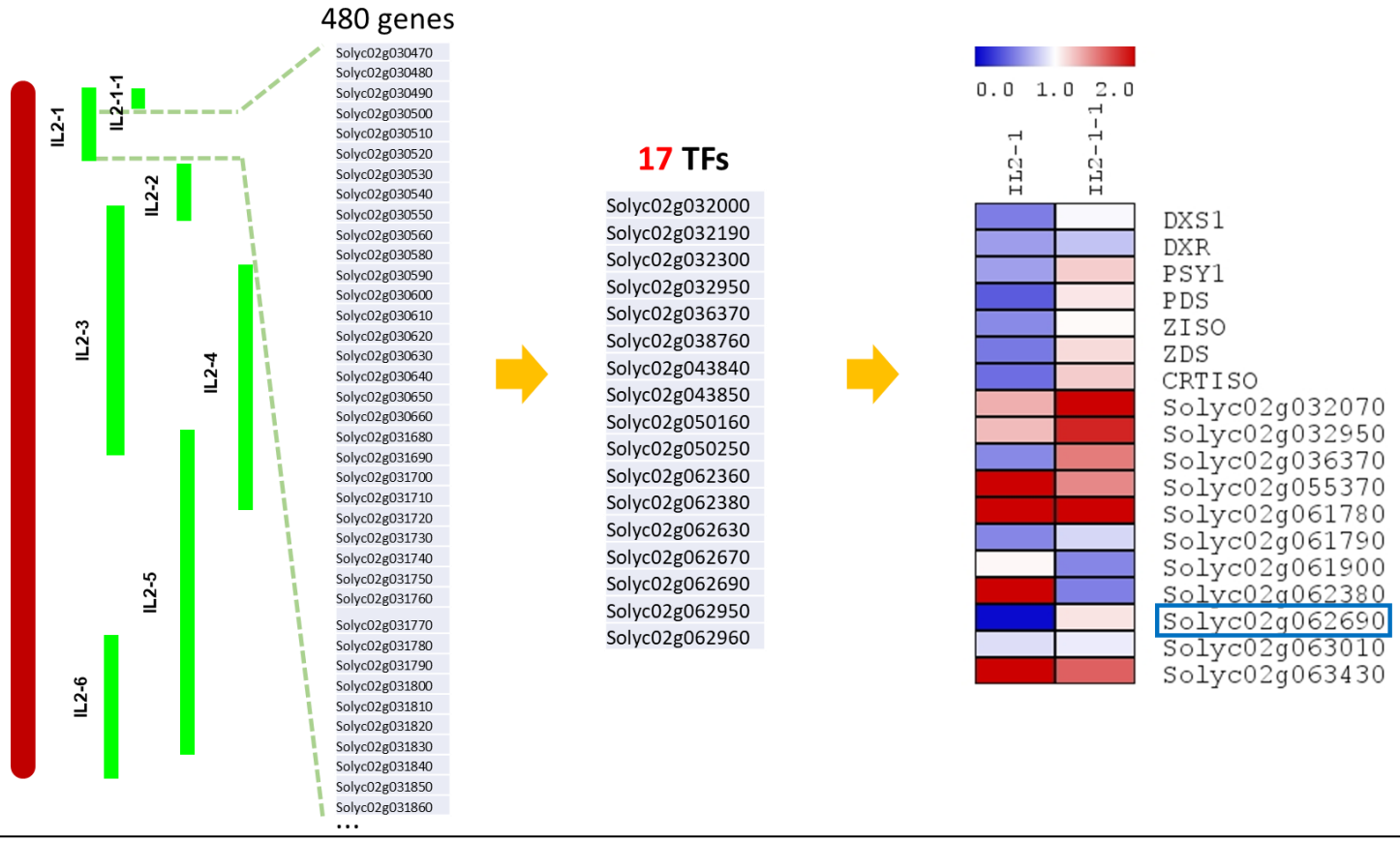


Figure 3.11 Schematic view of identification of *SlHONG* (*Solyc02g062690*) as encoding a transcription factor candidate controlling the lycopene biosynthetic pathway in tomato fruit. Based on the transcriptome profiling and global eQTL analysis of the carotenoid biosynthetic pathway, the non-overlapping region between IL2-1 and IL2-1-1 was identified as one the *trans*-eQTL hotspots harbouring the transcriptional regulator of lycopene biosynthesis in tomato fruit. There were 480 genes in this region, among which 17 genes were annotated as transcription factors. A heat map was drawn according to the transcript levels of the 17 genes, as well as the major genes in the lycopene biosynthesis pathway, which were *DXS1*, *DXR*, *PSY1*, *PDS*, *ZISO*, *ZDS* and *CRTISO*. *Solyc02g062690*, encoding a basic helix-loop-helix transcription factor, was outstanding out of the 17 genes, due to its most similar expression pattern compared to the structural genes. Therefore, Solyc02g062690 was identified as a candidate positive regulator of lycopene biosynthesis.

strong transcriptional associations (which forms the theoretical basis for gene discovery through co-expression analysis), which allows sets of genes to work in a coordinated way in response to different regulatory signals, internal or external. Thus, it is possible to identify the genomic locations of pathway-specific regulators by global eQTL studies.

A well-characterised metabolic pathway is essential for successful identification of regulators based on polymorphisms in transcript levels. As described previously, the entire carotenoid biosynthetic pathway in tomato fruit can be divided into three parts, isoprenoid biosynthesis, lycopene biosynthesis and lycopene degradation. Usually, *cis*-acting eQTLs have more significant effects on transcript levels than *trans*-eQTLs. Introgressions which showed co-ordinated changes in expression of the structural genes in each pathway sector, but which were free from structural genes (as shown in Figure 3.3, the ILs containing zero carotenoid biosynthetic genes), were prioritised to provide the first list of candidates for identification of *trans*-eQTL hotspots, eliminating possible noise caused by the presence of *cis*-acting eQTLs. Due to the central role of the isoprenoid biosynthetic pathway and the complexity of the branches of lycopene catabolism, as well as the high gene redundancy observed in both of these two pathways, no obvious *trans*-eQTL hotspots, in which most of the structural genes along the pathway were co-upregulated or co-downregulated, were identified as potentially regulating either of these two pathways.

Expression of five structural genes, *SlPSY1*, *SlZDS*, *SlPDS*, *SlZISO* and *SlCRTISO*, together with *SlDXS1*, encoding the enzyme catalysing the first step of the carotenoid biosynthetic pathway, were considered for identification of *trans*-eQTLs of lycopene production. Six *trans*-eQTLs hotspots for lycopene biosynthesis were identified (Figure 3.7). In IL8-1-2, IL9-1-3 and IL10-1-1, the expression levels of more than 80% of target genes were increased, indicating that there might be negative regulators in *S. lycopersicum*, which have been replaced by the corresponding genome of *S. pennellii*, leading to the upregulation of the structural genes. I was more interested in *trans*-eQTL hotspots which showed down-regulation of the structural genes, where lycopene biosynthesis was down-regulated likely due to the presence of an introgression from *S. pennellii*. IL2-1, IL9-3 and IL12-1-1 were identified as having co-downregulation of the biosynthetic genes (Figure 3.7), suggesting

that positive regulatory loci might be located in the introgressed regions in *S. lycopersicum*. In particular, IL2-1 had also been identified as one of the phenotypic QTLs with significantly less red colouration during fruit ripening (Figure 3.2, F1). It has been reported that the content of lycopene in IL2-1 is significantly lower than that in M82 (Tieman et al., 2006). Thus, IL2-1 was selected as a *trans*-eQTL hotspot for further investigation.

3.3.7 The region in IL2-1 non-overlapping with IL2-1-1 was identified as an active *trans*-eQTL.

Variation in fruit colour was observed at the whole population level of the introgression lines (Figure 3.2). At 10 days after breaker, IL2-1 (Figure 3.2 F1) was less coloured than M82 (Figure 3.2 G6), as well as its control, subline IL2-1-1 (Figure 3.2 G1). The difference in colour phenotype between IL2-1, and IL2-1-1 and M82 was clearer during the early stages of ripening (Figure 3.8). Lycopene accumulation in tomato fruit is continuous during fruit ripening, reflected as increasing red colouration, starting at breaker. Time-course photography was undertaken to observe any differences between IL2-1 and IL2-1-1. IL2-1 showed a much slower red pigmentation compared to IL2-1-1 (Figure 3.9), and was much less coloured when the fruit were fully ripe (Breaker + 15 days; B+15).

Transcriptome profiling of lycopene biosynthetic genes showed that the transcript abundance of all structural genes was decreased in IL2-1, which was not observed in its subline IL2-1-1 (Figure 3.7), consistent with the differences in lycopene accumulation. Furthermore, I showed that expression of *SlPSY1* was significantly reduced in IL2-1, three days after breaker compared to IL2-1-1 by qRT-PCR (Figure 3.10). The different performance of IL2-1 and its adjacent IL enabled a further narrowing down of the *trans*-eQTL hotspot harbouring regulator candidates to the region of IL2-1 that was non-overlapping with IL2-1-1 (Figure 3.11).

3.3.8 A gene encoding a basic helix-loop-helix transcription factor, SlHONG, (Solyc02g062690) was identified as a candidate positive regulator of lycopene biosynthesis in tomato fruit.

Each *trans*-eQTL hotspot contains a relatively small region replaced by the *S.pennellii* genome, which is about 16 cM of the non-overlapping region between IL2-1 and IL2-1-1, allowing the exploration of potential regulatory candidates at the molecular level. There are 480 genes located in the region of IL2-1 non-overlapping with IL2-1-1, among which, 17 were annotated as transcription factors (Figure 3.11). The best candidate within the interval was the gene whose expression correlated with multiple structural genes whose transcript levels were affected in IL2-1 (Figure 3.11). *Solyc02g062690*, encoding a basic helix-loop-helix transcription factor, was significantly less expressed in its *cis*-eQTL, IL2-1, with less than 5% transcript abundance of that in M82 fruit. No significant difference in the expression level of *Solyc02g062690* was observed between IL2-1-1 and M82 (Figure 3.11). Therefore, Solyc02g062690 was identified as a candidate positive regulator of lycopene biosynthesis. *Solyc02g062690* was named *HONG*, which means 'red' in Chinese, after the colour of lycopene, due to its potential role as a positive regulator in the production of this abundant red compound in tomato fruit.

3.4 Discussion

3.4.1. Genome-wide eQTL analysis provides a comprehensive understanding of the genetic regulation of carotenoid biosynthesis in tomato fruit.

The genetic basis of the regulatory mechanisms underlying carotenoid biosynthesis is likely a multifaceted process that contains many network connections and regulatory factors. Due to the complexity of the carotenoid biosynthetic pathway in tomato and the important roles of the compounds produced, contributing to a wide range of physiological process, it was very difficult to analyse the entire pathway in a single study, and also unlikely to be scientifically informative since different parts of the pathway are probably regulated

independently. Through the genome-wide eQTL mapping, 31 *cis*-eQTLs were identified covering genes encoding 18 different enzymes, annotated in tomato genome, among which some have been shown to be active in fruit by the characterisation of mutants, while most of them remain to be characterised further. Apart from complex connections linked to carotenoid biosynthesis, the redundancy in the genes encoding enzymes in the pathway further increases the difficulty in identification of the genes underlying the corresponding phenotypes. Expression profiling combined with eQTL mapping greatly facilitated the identification of functional paralogs operating in a tissue-specific manner. For example, the eQTL mapping revealed that *SlDXS1*, *SlIPP1*, *SlGGPP2*, *SlPSY1* and *SlCCD4B* were the genes functional in tomato fruit out of the multiple genes encoding different isoforms of these enzymes.

This analysis was completed in a single study, which would not be possible by mutant identification. SlPSY1 and SlCRTISO were the first two enzymes identified in the lycopene biosynthetic pathway using natural mutants. The function of *SlZISO* was not confirmed until 2013, because it works together with *SlPDS* as a single unit in the lycopene biosynthetic pathway, catalysing the synthesis of di-*cis*-ζ-carotene. The single mutant of *SlZISO* does not show significant changes in visible phenotype (Fantini et al., 2013). The identification of an eQTL containing *SlZISO* suggested that eQTL mapping has the potential for functional validation of candidate genes, specifically in two extreme situations, when naturally occurring knockouts are not phenotypically identifiable or are lethal.

Apart from applications in the characterisation of the genetic basis of metabolic pathways, eQTL analysis offers a powerful method for the elucidation of transcriptional regulation by identification of *trans*-eQTLs. Six *trans*-eQTL hotspots for lycopene biosynthesis in fruit were identified in this study, which defined candidate regions harbouring genes encoding regulators that affect the expression of all or most of the genes in the lycopene biosynthetic pathway, in either a positive or negative manner. The identification of genes encoding transcriptional regulators of a specific pathway remains difficult, because any phenotypic effects caused by variation in *trans*-acting genes may be masked by variation in *cis*-acting elements of structural genes, observed at the transcript level. Combining understanding of

the well-elucidated genes encoding enzymes of the carotenoid biosynthetic pathway, eQTL analysis offered an alternative experimental approach to identify regulators by focusing on the expression levels of the structural genes of the target pathway to identify transcriptional master regulators through the identification of *trans*-acting eQTL hotspots.

3.4.2 eQTL analysis is a sensitive method for the identification of genes encoding transcriptional regulators through genome-wide analysis

Usually, when eQTL analysis is conducted in different populations using similar tissues or treatments, the eQTLs identified remain consistent between populations. However, eQTL studies performed in the same population but using different tissues or treatments may yield distinct or complementary results, reflecting the dynamic nature of the transcriptome and transcriptional regulation in multicellular organisms. For instance, a similar genome-wide eQTL analysis was conducted using the published RNA-seq data of the same *S.lycopersicum* x *S. pennellii* introgression population, but in leaves, instead of in fruit (Lockhart, 2013). I applied the same principles to investigate carotenoid biosynthesis in leaves using this except, due to a lack of expression data for M82, heatmaps were generated based on the relative expression of each IL normalised to the average transcript level of the total 76 lines, which was set as 1 (Figure 3.12, Figure 3.13 and Figure 3.14). Generally, variations in transcript abundance of the carotenoid biosynthetic genes in leaves between ILs was much smaller than in fruit, which might be because carotenoid production in leaves of the two parental lines was substantially lower than in fruit. Amongst the *cis*-eQTLs identified, some of which were consistent with those observed in fruit, there were some results of the eQTL analysis in leaves that were complementary to the fruit *cis*-eQTLs, especially where there was gene redundancy. For example, there are five *SlGGPPS* genes annotated in the tomato genome, and *SlGGPP2* was characterised as the functional isoform in tomato fruit by identification of IL4-3/IL4-4 as a *cis*-eQTL. However, based on the eQTL analysis in leaves, *cis*-eQTLs for *SlGGPP1*, *SlGGPP4* and *SlGGPP5* were identified, which reside in IL11-1/IL11-2, IL2-5 and also in IL2-5 respectively (Figure 3.12). IL6-1 and IL3-2 were identified as *cis*-eQTLs for *SlCHY1* in leaves and *SlCHY2* in fruit respectively according to the comparative analyses conducted (Figure 3.6 and Figure 3.14). It has been shown that the

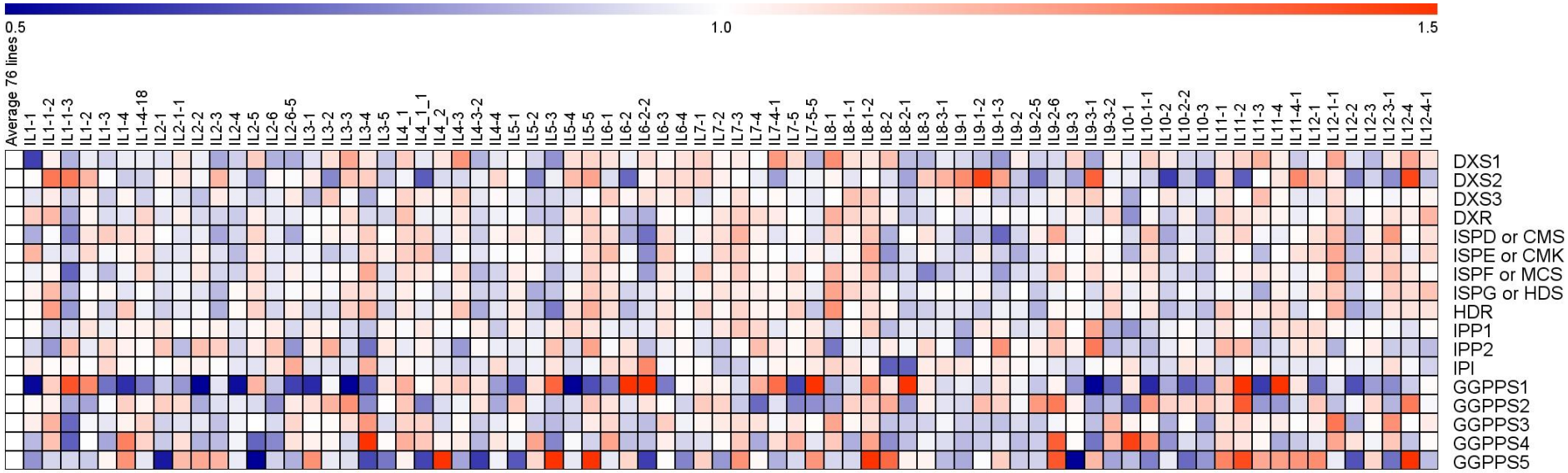


Figure 3.12 The heatmap of transcript abundance of isoprenoid biosynthetic genes in leaves compared to the average value of 76 lines, which was set as 1. The data used to generate the heatmap was from Lockhart *et al*, 2013. The heatmap was based on the relative expression of each IL normalised to the average transcript level of the total 76 lines, as the expression data for M82 was not available.

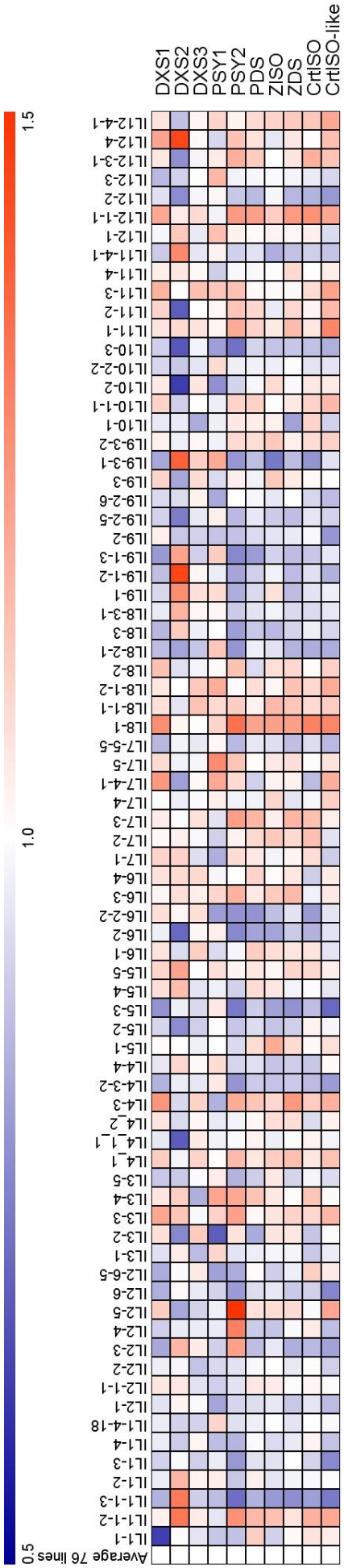


Figure 3.13 The heatmap of transcript abundance of the genes in lycopene biosynthetic pathway in leaves compared to the average value of 76 lines, which was set as 1. The data used to generate the heatmap was from Lockhart *et al*, 2013. The heatmap was based on the relative expression of each IL normalised to the average transcript level of the total 76 lines, as the expression data for M82 was not available.

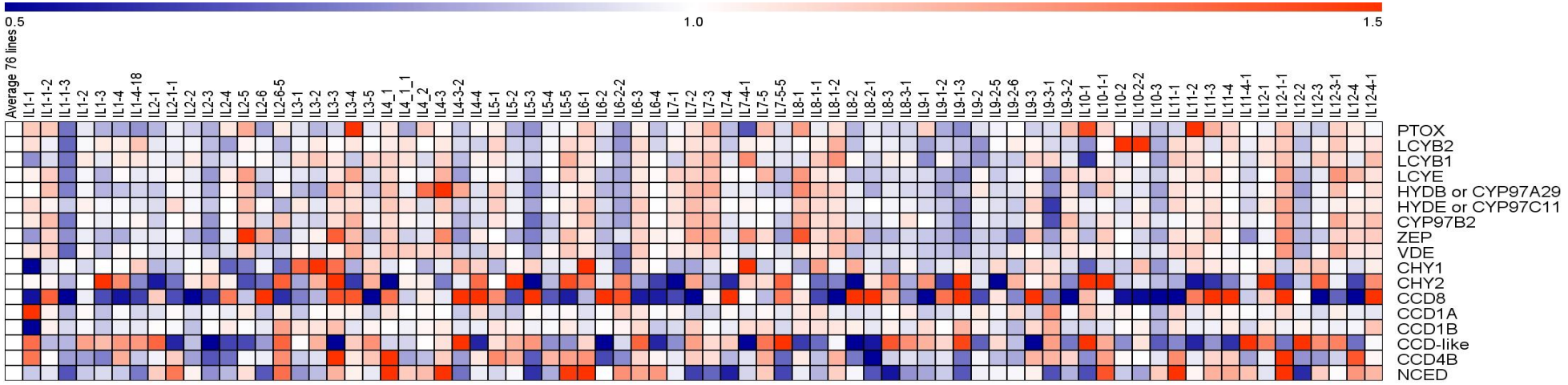


Figure 3.14 The heatmap of transcript abundance of the genes in lycopene metabolism in leaves compared to the average value of 76 lines, which was set as 1. The data used to generate the heatmap was from Lockhart *et al*, 2013. The heatmap was based on the relative expression of each IL normalised to the average transcript level of the total 76 lines, as the expression data for M82 was not available.

tomato *wf* mutant, which maps to the *CHY2* gene, is sufficient to severely impair flower β-xanthophyll biosynthesis, while leaf β-xanthophyll levels remain similar to those found in wild-type plants (Galpaz, 2006). In keeping with this result, the tomato *CHY2* transcript is expressed preferentially in flowers, while the *CHY1* transcript is expressed preferentially in leaves (Galpaz, 2006), which is consistent with the results of eQTL analysis. These data indicate that, in different tissues, the hydroxylation of β-carotene is preferentially performed by different isoforms encoded by different genes. The distinct results obtained from the eQTL analysis of different tissues from the same population, further established

the sensitivity of the eQTL method, which is useful for studying transcriptional regulation of tissue-specific pathways, such as lycopene biosynthesis.

3.4.3 eQTL mapping may be more informative than mQTL mapping in identifying the candidate regions containing the genes responsible for specific expression phenotypes.

Two major carotenoids, lycopene and β-carotene, have been measured across the whole IL population in three continuous growing seasons (Rousseaux et al., 2005). Three IL lines were reported with undetectable lycopene contents: IL3-2, IL12-2 and IL12-3, and it was claimed that it was because the replaced regions harboured the genes encoding phytoene synthase 1 (PSY1) and lycopene-ε-cyclase (LCYE) from *S. pennellii* respectively (Rousseaux et al., 2005). Although IL3-2 was confirmed as the *cis*-eQTL of *SlPSY1*, the expression of *SlLCYE* was more than 4000 times higher in IL12-2 (compared to M82), whereas no expression of *SlLCYE* was detected in IL12-3 (Figure 3.6). This suggests that the lack of lycopene in IL12-3 must be due to another gene than *SlLCYE*. Similarly, IL6-2 and IL6-3 both have higher β-carotene levels than M82 although only IL6-3 contains the *SlCYCB* allele from *S. pennellii,* which has been characterised in the previous eQTL mapping (Figure 3.6). This suggests that IL6-2 might, in fact, involve a *trans*-eQTL. These inconsistencies between mQTL and the genetic determinants of metabolic control, show the disadvantages of identification of structural genes for certain metabolic pathways purely on mQTL mapping, an approach which could well be complemented by eQTL mapping. A combination of RNA-seq and metabolic profiling could facilitate characterisation of the key points of metabolic regulation and the identification of the corresponding regulatory genes.

Chapter 4

Identification of SlHONG as positive regulator candidate controlling the lycopene biosynthetic pathway in tomato fruits

4.1 Introduction

Carotenoids serve critical roles in plant growth and development, particularly in responses to environmental stresses. For instance, carotenoids absorb light energy for use in photosynthesis, and protect chlorophyll from photodamage. In addition, they serve as precursors for the biosynthesis of the phytohormone abscisic acid (ABA). Many carotenoids confer colour phenotypes, and serve as pigments in flowers to attract animal pollinators and in fruit to attract animal dispersors. They may be important in regulating the rate of post-harvest fruit softening because of their antioxidant capacity. Carotenoids are produced in both vegetative and reproductive tissues, such as leaves, stems, root, flowers and fruit. The enzymatic steps of the carotenoid biosynthesis pathway have been described in Chapter 1 and 3, and due to the integral roles of carotenoids in plant growth and development, carotenoid biosynthesis is regulated by multiple factors including environmental factors (Zhang et al., 2014; Quail, 2002; Cocaliadis et al., 2014), phytohormones (Marty et al., 2005; Jones et al., 2002; Barickman et al., 2014; Park et al., 2009; Zhu et al., 2015), and plastid number (Galpaz et al., 2008). Among these, ethylene plays a crucial role in the regulation of fruit ripening and carotenoid accumulation in tomato fruit. Rapid accumulation of lycopene in tomato is triggered by ethylene production during fruit ripening, and expression of *SlPSY1* and *SlPDS* is closely associated with ethylene levels. It has been shown that application of an ethylene precursor, aminocyclopropane carboxylic acid (ACC), accelerates carotenoid accumulation and, consequently, pigmentation in tomato fruit (Su et al., 2015). Several studies have described the regulation of carotenoid biosynthesis at the molecular level in plants (Hirschberg, 2001; Cunningham and Gantt, 1998), and some transcription factors have been found that contribute to controlling carotenoid levels in fruit (Manning et al., 2006; Lee et al., 2012; Toledo-Ortiz et al., 2010; Vrebalov et al., 2002). However, all of these transcription factors have broad effects on ripening and none regulate single metabolic pathways, specifically in fruit. One of the reasons why mutants of regulatory genes may not have been identified is that these genes may have pleiotropic effects and therefore pleiotropic phenotypes, making it very difficult to establish the links between phenotypes and the specific regulator. Since most of the accumulation of carotenoids in tomato fruit is closely related to the process of fruit ripening, which, in turn, can be influenced by a wide range of internal and external factors, it is critical

to investigate the expression of genes encoding carotenoid biosynthetic enzymes during the different developmental stages of fruit and relate these to the expression of any transcriptional regulator potentially directly controlling the induction of the carotenoid biosynthetic pathway.

Transient silencing of genes, particularly in a specific physiological context, can provide evidence of their molecular functions, although compelling proof of function normally requires stable, loss-of-function alleles. Virus-Induced Gene Silencing (VIGS) is an efficient method for rapid characterisation of gene function in plants, which has been successfully applied in tomato (Baulcombe, 1999; Orzaez, 2005).The VIGS system has been further modified to a visual system using high-anthocyanin tomato, *Del*/*Ros1* purple tomato(Orzaez et al., 2009). Silencing of target genes-of-interest, together with *Del* and *Ros1* transcription factors leading to red sectors, which can be distinguished easily from the non-silenced sectors which remain purple on the same fruit, allows the quantitative assessment of gene function based on the colour phenotype of the silenced red sectors (Orzaez et al., 2009; Figure 1.5).

Despite numerous mutant screens searching for transcriptional regulators of carotenoid biosynthesis in tomato, none have been identified by forward genetic screening although many mutations in structural genes encoding biosynthetic enzymes have been identified. Consequently, construction of a stable, knock-out line of *SlHONG* was considered essential to establish its role in the regulation of carotenoid biosynthesis in tomato fruit. An engineered form of the clustered, regularly interspaced, short palindromic repeat (CRISPR) system of *Streptococcus pyogenes* has been shown to function in plants (Nekrasov et al., 2013; Belhaj et al., 2013; Jinek et al., 2013; Upadhyay et al., 2013; Xie and Yang, 2013). CRISPR systems provide a unique prokaryotic defence against invading DNAs, such as plasmids and viruses (Barrangou et al., 2007). The CRISPR-Cas system has been demonstrated to work efficiently for genome editing in bacterial, yeast, and animal systems (Cong et al., 2013; Gasiunas et al., 2012; Jinek et al., 2013) and has been applied to different plants, such as *Arabidopsis thaliana*, *Nicotiana benthamiana*, tomato, rice and wheat (Jiang

et al., 2013b; Feng et al., 2014; Upadhyay et al., 2013). This method is well-suited for de novo generation of knock-out mutations of candidate regulatory genes controlling carotenoid biosynthesis in tomato.

According to differential results obtained from eQTL analysis of fruit and leaves separately in Chapter 3, expression of genes encoding enzymes of the carotenoid biosynthetic pathway appeared to be regulated in a tissue-specific pattern. Due to the complexity of carotenoid biosynthesis and the important roles of the compounds it produces, which are involved in widely diverse biological processes, the entire pathway was sub-grouped into three sections. As there were no clear *trans*-eQTL hotspots for isoprenoid biosynthesis or for lycopene catabolism (as described in Chapter 3), the study of the molecular basis of transcriptional regulation of carotenoid production in tomato fruit was focused down on lycopene biosynthesis, for which six *trans*-eQTL hotspots were identified, including IL2-1. In this chapter, expression profiles of genes involved lycopene biosynthesis were characterised in different tissues of tomato plants, as well as in seven fruit developmental stages, through which the key stages of transcriptional regulation of the lycopene biosynthetic pathway were identified. Correspondingly, the expression levels of the candidate *bHLH* transcription factor found in Chapter 3, *SlHONG*, were analysed. Transient and stable transformation technologies were used to assess the regulatory function of *SlHONG* in controlling the transcription of genes in the lycopene biosynthetic pathway in fruit.

4.2 Materials and Methods

4.2.1 Plasmid construction

A full-length cDNA of *SlHONG* (*Solyc02g062690*) was amplified with Gateway compatible primers and then recombined into pDONR207™ (Invitrogen) by BP reaction to generate the entry clone pENTR207-*SlHONG*-CDS. The entry clone was recombined with pBIN19-p35S-GW and pBIN19-pE8-GW using LR clonase™ (Invitrogen) to make expression vectors,

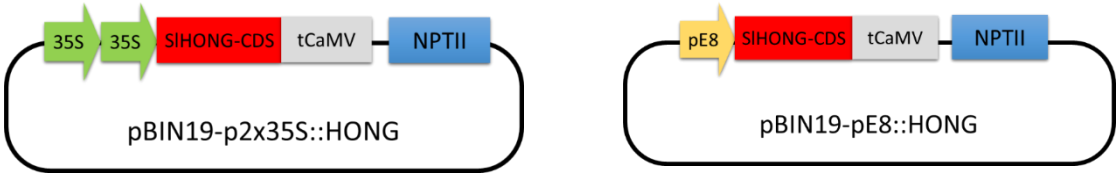


Figure 4.1 Schematic view of constructs for *SlHONG*-overexpression driven by double CaMV 35S and E8 promoters in stable transformations. Each colour represents different transcript elements. Coding sequence of *SlHONG* (red rectangles) was driven by a double 35S promoter (green arrows) and a E8 promoter (yellow arrow) respectively. Two constructs were used for Agrobacterium-mediated stable transformation of *Solanum lycopersicum.* Gray rectangle, CaMV terminator. Blue rectangle, NPTII, transcript unit of kanamycin resistance gene.

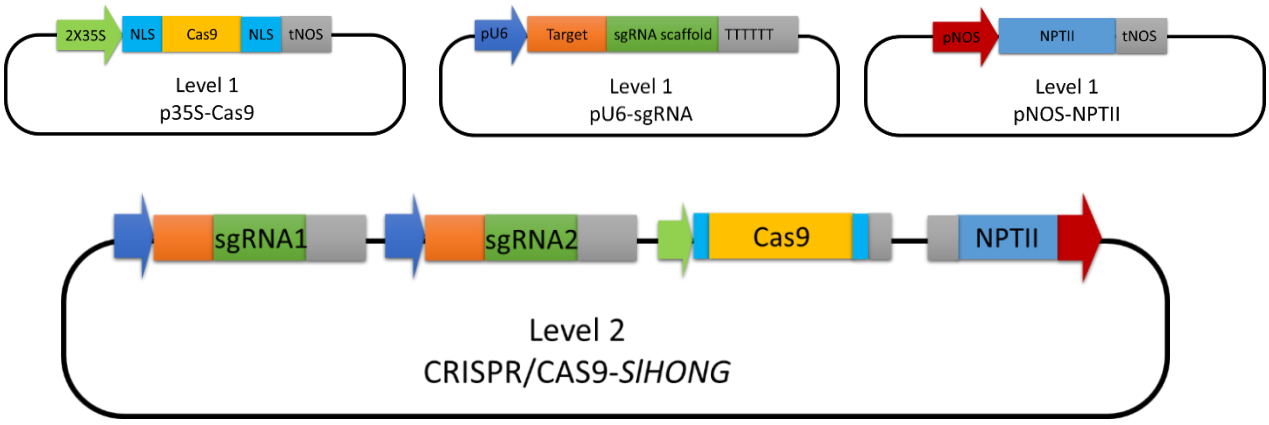


Figure 4.2 Schematic view of CRISPR constructs used for generation of *ΔSlHONG* lines. Each colour represents different transcript elements. Transcript units of sgRNA1, sgRNA2, Cas9 and NPTII were cloned into level 1 vectors ec41722-ele1, ec41744-ELE2, ec41766-ELE3 and ec41780-ELE4 respectively, by Golden Gate cloning. *Cas9* was driven by double CaMV 35S promoters and flanked with nuclear localization sequences. sgRNAs and *NPTII* gene were under the control of Arabidopsis U6 promoter, and Nos promoter separately. Four level 1 constructs were cloned into a binary level 2 vector, pAGM4723-pL2B for Agrobacterium-mediated stable transformation of *Solanum lycopersicum.* The level 1 and level 2 vectors were obtained from the TSL SynBio group (http://synbio.tsl.ac.uk/).

pBIN19-p35S::*HONG* and pBIN19-pE8::*HONG* used for stable overexpression of *SlHONG* in tomato (Figure 4.1). 262 bp of *SlHONG* (*Solyc02g062690*) coding sequence near the 3' UTR was amplified with Gateway compatible primers and then cloned into pDONR207™ (Invitrogen) by using BP clonase™ (Invitrogen) to make the entry clone pENTR207-*SlHONG*-f. The entry clone was recombined with pTRV2-GW and pTRV2-*Del*/*Ros1*-GW, by BP reaction to make expression vectors, which were then used for VIGS experiments conducted in wild type tomato and *Del*/*Ros1* purple tomato respectively.

4.2.2 Application of CRISPR/Cas9 mediated genome editing to generate *SlHONG* knock-out tomato plants.

The vectors and protocol were kindly provided by TSL SynBio group (http://synbio.tsl.ac.uk/).

In order to knock out the entire, or most of the *SlHONG* gene in tomato, I used two sgRNAs to achieve a complete knock out deletion, each of which contained 20 bp target sequence. The following principles were applied for selecting the target sequences: 1) the target sequence should be followed by an 'NGG' PAM (protospacer adjacent motif) sequence, as in 5' NNNNN NNNNN NNNNN NNNNN NGG 3'; 2) the target sequence should sit near the 5' or 3' end of the *SlHONG* gene. Eight target sequences were selected according these two rules.

The chosen target sequences for the gene of interest were introduced into the sgRNA scaffold by PCR. The forward primer harbouring the 20 bp guide sequence was applied to amplify the specific sgRNA, using plasmid Picsl70001 as template, which contained the sgRNA scaffold. The forward primer was 'tgtggtctca ATTG NNNN NNNNN NNNNN NNNNN gttttagagctagaaatagcaag', in which the sequence in lower-case annealed to the sgRNA and N sequence representsed the target 20 bp guide sequence. The same reverse primer was

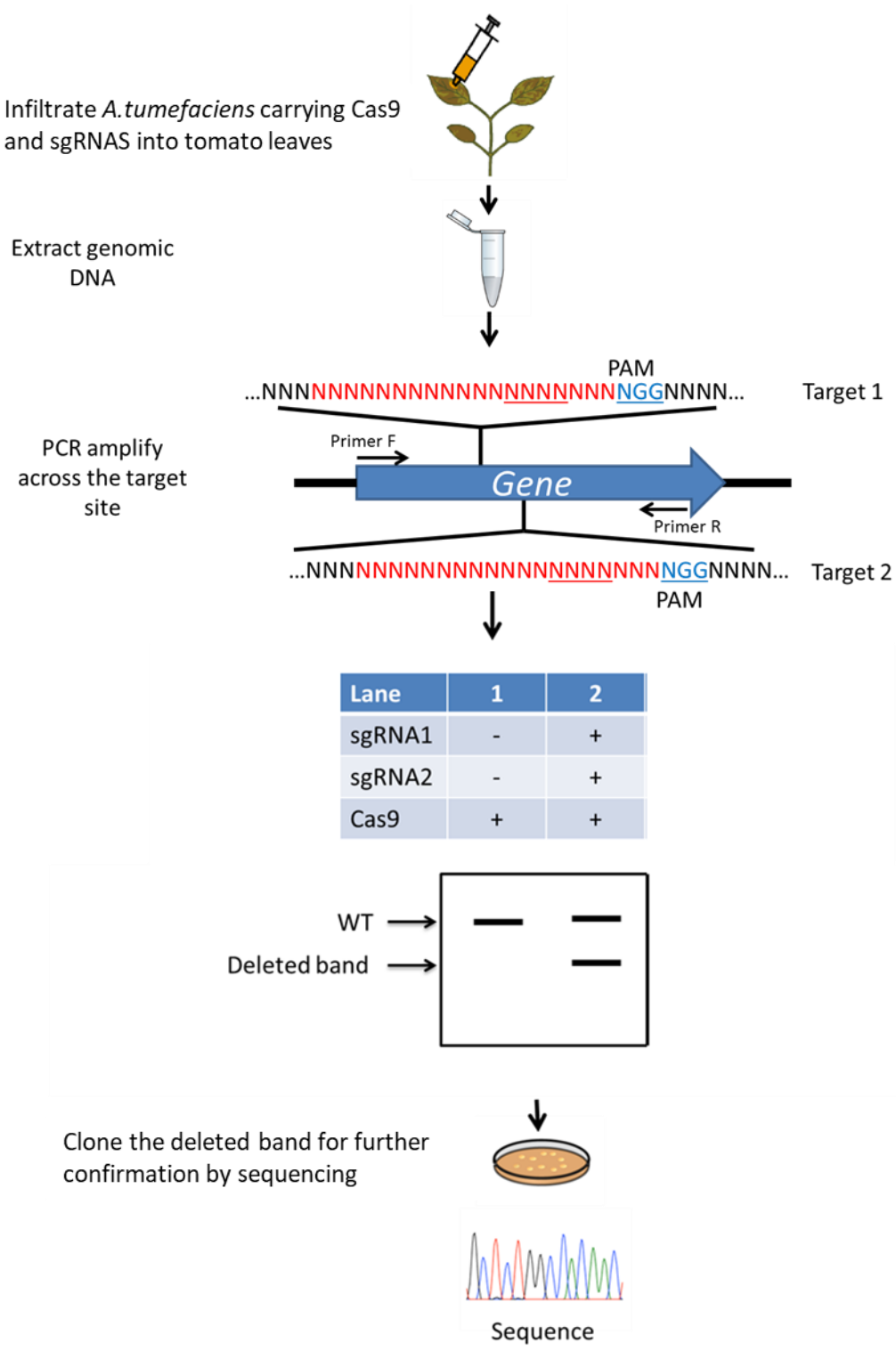


Figure 4.3 Schematic drawing illustrating the strategy of sgRNA efficiency check by Agroinfiltration. The Cas9 nuclease and the sgRNAs matching the gene of interest were co-infiltrated in different plasmids using *Agrobacterium tumefaciens* as a vector in tomato leaves. The genomic DNA is extracted from the leaf tissues and subject to PCR-amplification with primers flanking the deletion region. The deletion bands were further confirmed by sequencing. The figure was redrawn from Nekrasov et al., *Nature Biotechnology*, 2013.

used for amplification of all the sgRNAs as tgtggtctca AGCGTAATGCCAACTTTGTAC. BsaI sites shown in blue, were used for Golden Gate cloning. The final amplicon was like tgtggtctcaATTGNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAA GGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCTAGACCCAGCTTTCT TGTACAAAGTTGGCATTACGC. After confirmation by sequencing, each amplicon was combined with plasmid pICSL90001, which contained a synthesised U6-III promoter, by Golden Gate digestion-ligation reactions to assemble the promoter and sgRNA into a complete transcription unit. The reaction was then transformed into *E. coli* and white colonies were selected on LB agar plates with carbenicillin and X-Gal as selection, leading to a Level 1 sgRNA expression cassette, pU6-sgRNA (Figure 4.5). sgRNA expression cassettes were transformed into *Agrobacterium tumefaciens* strain *AGL1* in pairs (to generate a complete deletion of *SlHONG*) and infiltrated into tomato leaves to check their efficiencies. The sgRNAs pairs which could generate the deletion bands with expected size were selected as the best candidates for stable transformation. A schematic view of the strategy of sgRNA efficiency check was illustrated in Figure 4.6.

Any exact sequence match to the 3′ half of the sgRNA (known as the seed region) followed by a PAM motif, may potentially be a target for off-target activity. If there are mismatches in the 3′ half of the guide RNA, or if the PAM is absent, then the sequence is unlikely to be cleaved. To minimise the possibility of 'off-target' activity of the sgRNAs, each target sequence was checked against the tomato genome. There were two sgRNAs with no off-target possibilities as well as acceptable efficiency, which were GTGGAGATAAGAAGAAACAGTGG (upstream of start codon) and GCGGATTTTGTGTACAGGATAGG (downstream of stop codon).

The two Level 1 sgRNA expression cassettes, the Cas9 cassette and the Kanamycin selection cassette were recombined in into an empty level 2 vector, pAGM4723, by Golden Gate digestion-ligation reaction to make the final expression cassette (Figure 4.6). The final level 2 construct was transformed into *Agrobacterium tumefaciens* (strain GV3101:pMP90) and used for stable transformation.

4.3 Results

4.3.1 *SlHONG* was expressed in most of the tissues across tomato plants.

Although *HONG* was predicted to play a positive role in carotenoid accumulation in tomato fruit, as identified by eQTL mapping and co-expression analysis, *SlHONG* was not specifically expressed in fruit as shown by RT-qPCR. Transcripts of *SlHONG* were detected in all tested tissues across the tomato plant (Figure 4.4). Moreover, very high mRNA abundance of *SlHONG* was observed in unopened buds, as well as fully opened flowers (Figure 4.4). *SlHONG* was also highly expressed in vegetative tissues, such as stems and leaves (Figure 4.4), indicating that *SlHONG* might participate in other physiological processes in addition to be a candidate positive regulator of carotenoid biosynthesis in fruit.

4.3.2 Expression profiles of carotenoid biosynthetic genes in tomato plants.

The accumulation of lycopene during fruit development in tomato follows the induction of expression of genes encoding specific isoforms of enzymes involved in lycopene synthesis. However, carotenoids accumulate not only in fruits, but also play other vital roles in plants, such as transferring electrons to chlorophyll during photosynthesis, protecting chlorophyll from photodamage, as well as serving as precursors for the biosynthesis of the phytohormone abscisic acid (ABA). To capture the expression pattern of the genes activelyinvolved in carotenoid biosynthesis in fruit, RT-qPCR was conducted to analyse the mRNA abundance of these structural genes in tomato plants. *SlPSY1*, encoding phytoene synthase catalysing the rate-limiting step in lycopene biosynthesis, was almost exclusively expressed in reproductive tissues, such as flowers and red-ripe fruits (Figure 4.5A). Transcripts of *PSY1* were hardly detected in vegetative tissues and immature fruit, whereas *SlPSY2*, which encodes the other isoform of SlPSY, was highly expressed in most tissues according to Tomato eFP Browser (http://bar.utoronto.ca/efp_tomato). *SlZISO* showed a very similar tissue-specific pattern of expression to *SlPSY1*, and highest expression was

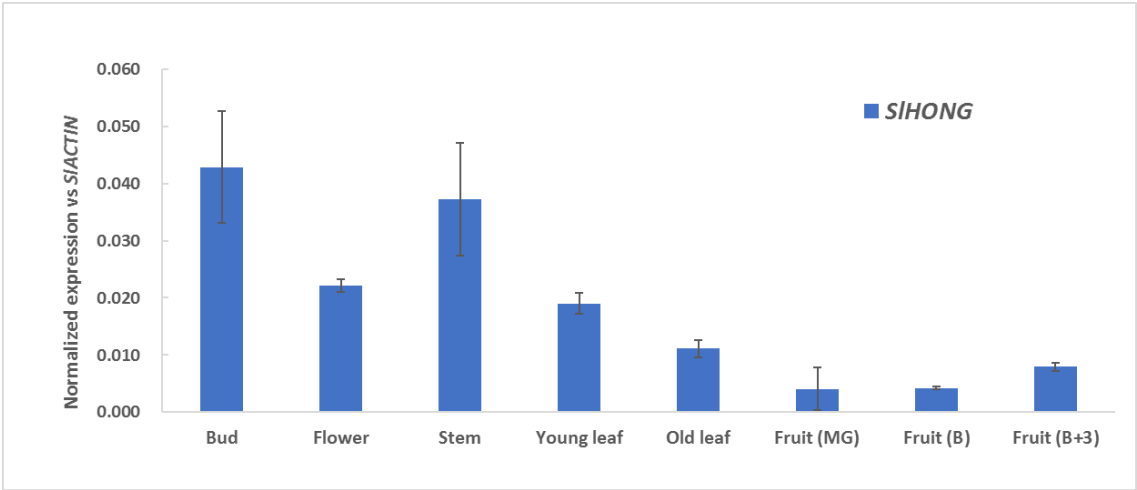
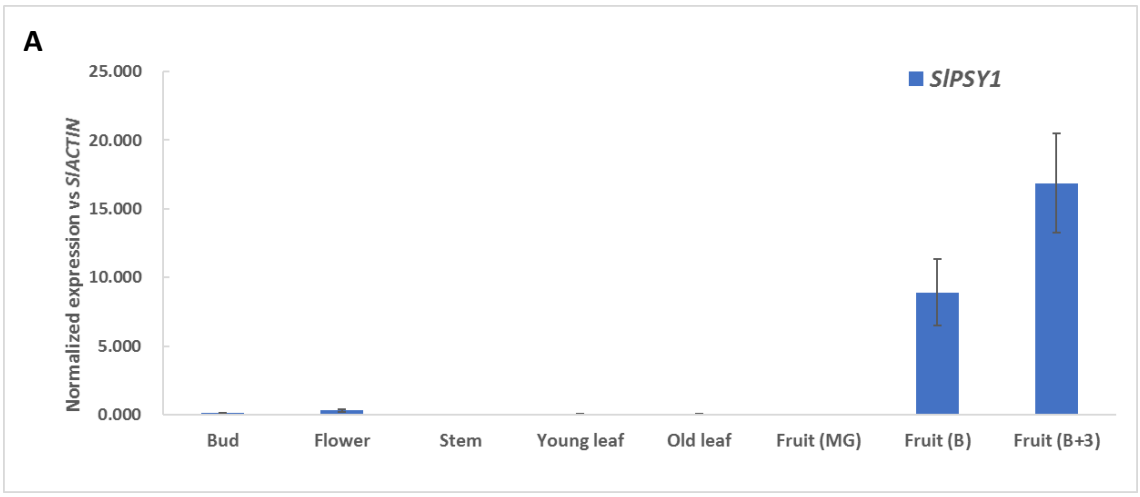
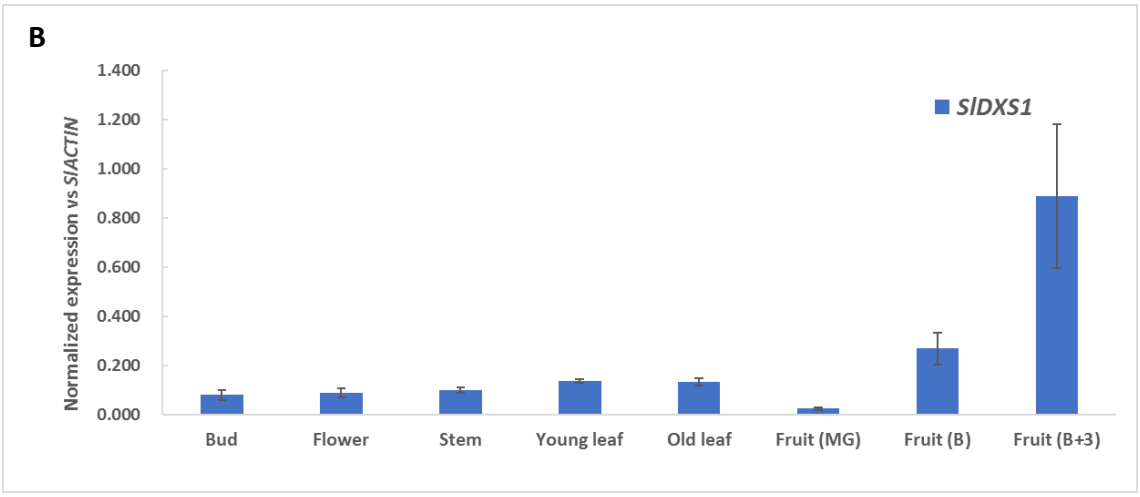


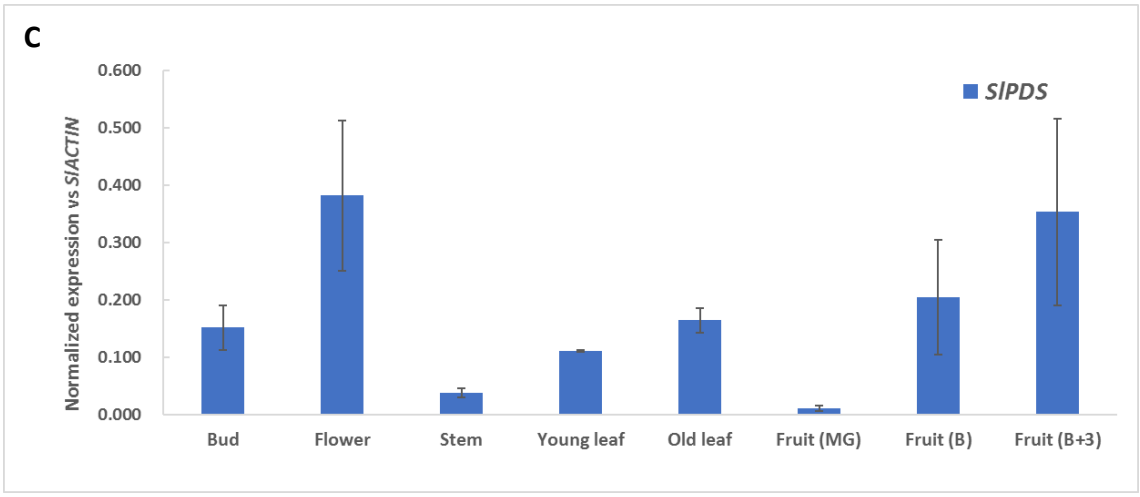
Figure 4.4 Expression of *SlHONG* in different tissues of tomato plants as determined by RT-qPCR. MG, Mature Green; B, Breaker; B+3, three days post breaker; Bud, unopened flower; Flower, fully opened flower (before pollination). Expression levels shown were relative to that for *ACTIN* (*Solyc03g078400*). Data are the average ± standard deviation of at least three biological replicates.

detected in ripening fruit (Figure 4.5E). DXS catalyses the first step of the isoprenoid biosynthetic pathway in plastids, providing precursors for lycopene biosynthesis (Figure 1.3) and its activity is thought to determine carotenoid accumulation in many tissues (Carretero-Paulet et al., 2006; Botella-Pavía et al., 2004). There are three genes encoding isoforms of SlDXS found in tomato. According to the Tomato eFP Browser, the transcripts of one *DXS* isoform (*Solyc11g010850*) are abundant in flowers, but hardly detected in fruits. Another *DXS* isoform (*Solyc08g066950*) is expressed across the plant at relatively low levels. Expression of *SlDXS1 (Solyc01g067890)*, the most active isoform in fruit, was observed in all tested tissues, and was much higher in mature fruit compared to other tissues (Figure 4.5B). RT-qPCR results showed that *SlPDS*, *SlZDS* and *SlCRTISO* were highly expressed in ripe-stage fruit and flowers, but expression of these three genes was also observed in green tissues at various levels (Figure 4.5C, D, F). From the differential expression of the genes involved in lycopene biosynthesis, I observed that all the biosynthetic genes were highly expressed in mature fruit, consistent with the high level of accumulation of lycopene during fruit ripening.

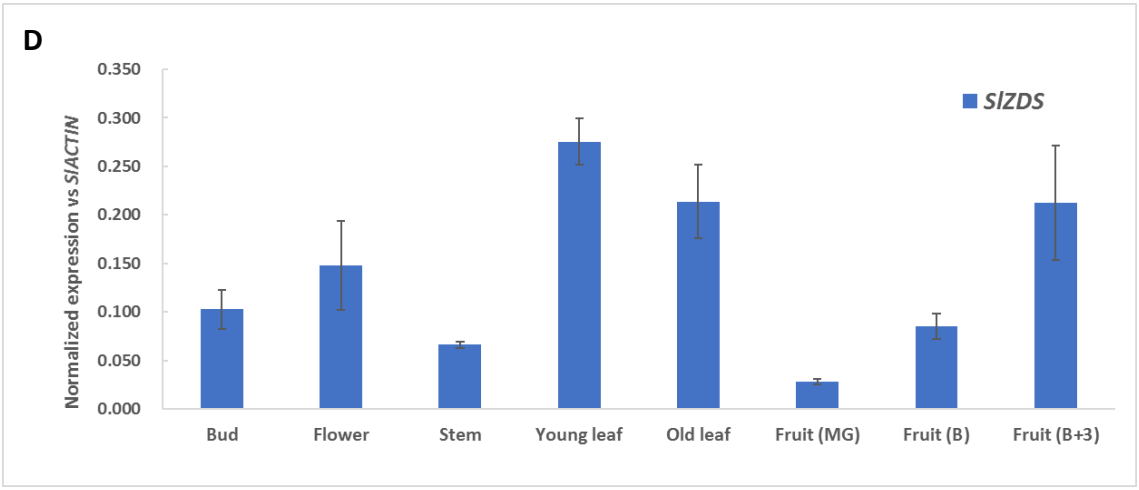
LCYB operates at a branch point of the carotenoid pathway, and catalyses the first step of lycopene catabolism to form other carotenoids, including β-carotene and xanthophylls (Figure 1.3). The highest expression of *SlLCYB* was in flowers (Figure 4.5G), confirming its role in the production of β-carotene, the precursor for synthesis of neoxanthin and violaxanthin, which are the two dominant xanthophylls responsible for the yellow colouration to tomato flowers. Although the transcript level of *SlLCYB* was slightly increased in fruit during ripening, its expression at 3 days after breaker was much lower than for other lycopene biosynthetic genes (Figure 4.5 G). This observation was consistent with the observation that the expression of genes encoding enzymes that convert lycopene to β-carotene and lutein are reduced in their expression in *S. lycopersicum* and other red-fruit species, but not reduced in green-fruited species such as *S. pennellii*, which was confirmed by the eQTL analysis shown in Chapter 3 (Figure 3.6).


A
SlPSY1
Normalized expression vs SlACTIN
25.000
20.000
15.000
10.000
5.000
0.000
Bud
Flower
Stem
Young leaf
Old leaf
Fruit (MG)
Fruit (B)
Fruit (B+3)


B
SlDXS1
Normalized expression vs SlACTIN
1.400
1.200
1.000
0.800
0.600
0.400
0.200
0.000
Bud
Flower
Stem
Young leaf
Old leaf
Fruit (MG)
Fruit (B)
Fruit (B+3)


C
SlPDS
Normalized expression vs SlACTIN
0.600
0.500
0.400
0.300
0.200
0.100
0.000
Bud
Flower
Stem
Young leaf
Old leaf
Fruit (MG)
Fruit (B)
Fruit (B+3)

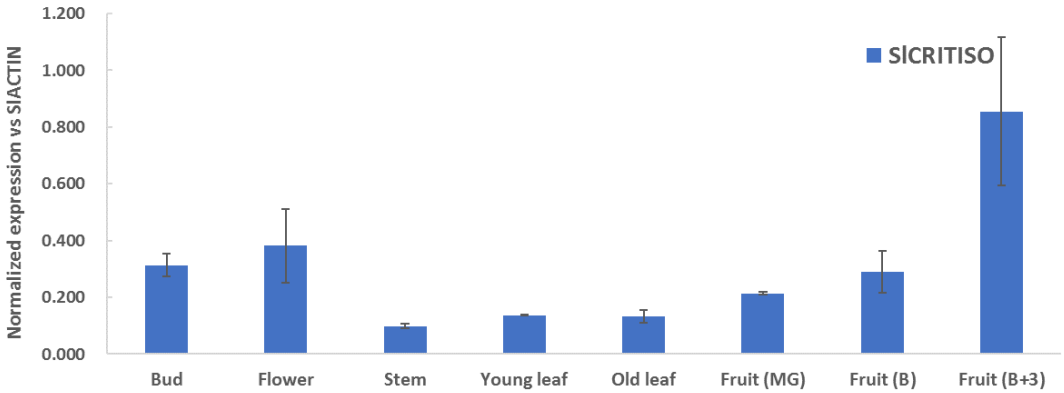
D


Normalized expression vs SlACTIN
0.350
0.300
0.250
0.200
0.150
0.100
0.050
0.000
SlZDS
Bud
Flower
Stem
Young leaf
Old leaf
Fruit (MG)
Fruit (B)
Fruit (B+3)

E


Normalized expression vs SlACTIN
16.000
14.000
12.000
10.000
8.000
6.000
4.000
2.000
0.000
SlZISO
Bud
Flower
Stem
Young leaf
Old leaf
Fruit (MG)
Fruit (B)
Fruit (B+3)

F


Normalized expression vs SlACTIN
1.200
1.000
0.800
0.600
0.400
0.200
0.000
SlCRITISO
Bud
Flower
Stem
Young leaf
Old leaf
Fruit (MG)
Fruit (B)
Fruit (B+3)

G

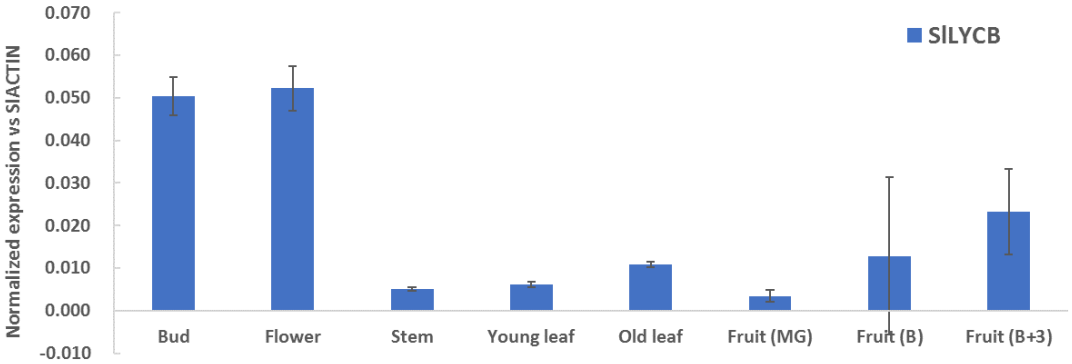
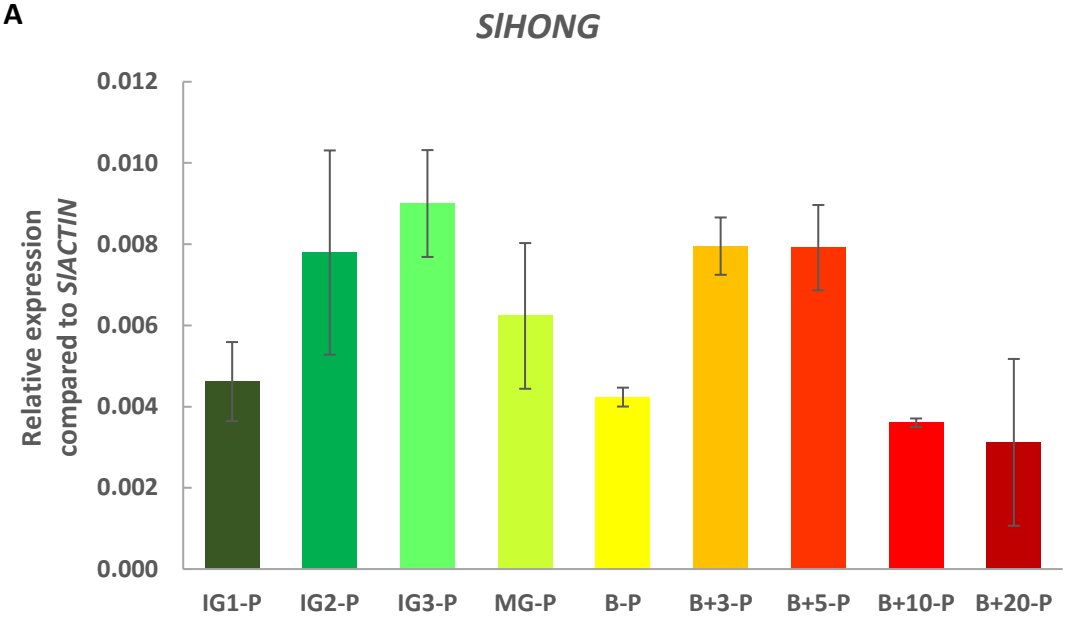


Figure 4.5 Expression of genes involved in carotenoid biosynthesis in different tissues of tomato plants measured by RT-qPCR. MG, Mature Green; B, Breaker; B+3, three days post breaker; Bud, unopened flower; Flower, fully opened flower (before pollination). Expression levels shown are relative to that for ACTIN (Solyc03g078400). Data are the average ± standard deviation of at least three biological replicates.

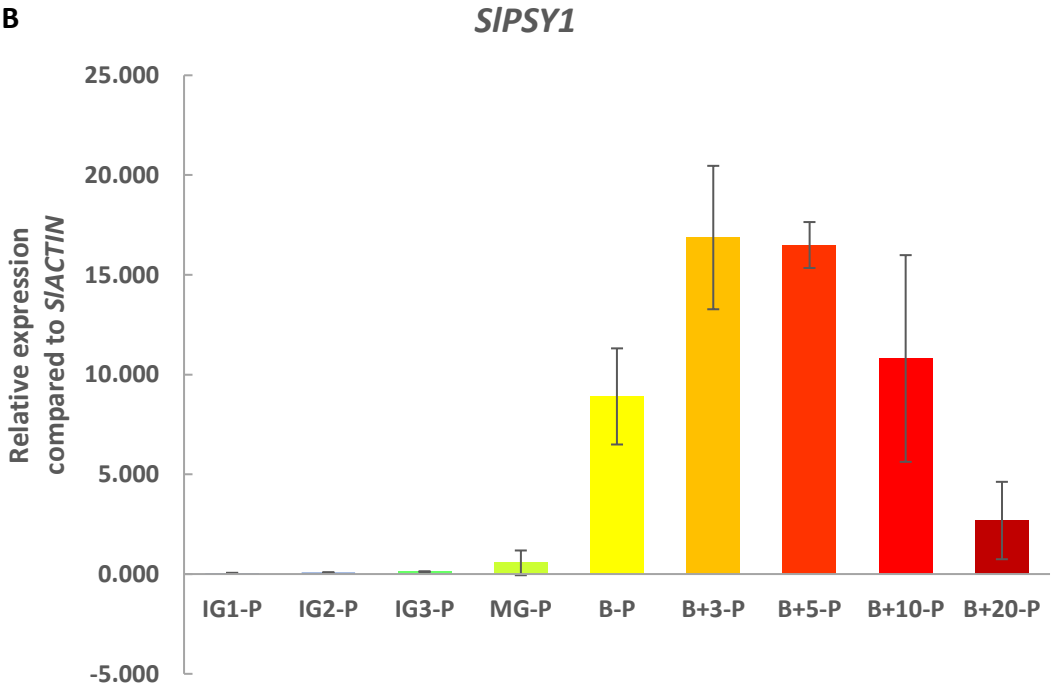
4.3.3 Expression levels of *SlHONG* were correlated with mRNA abundance of carotenoid biosynthetic genes during tomato fruit ripening.

Carotenoid biosynthetic genes are expressed tissue-specifically in tomato plants (Figure 4.5) suggesting that carotenogenesis in photosynthetic tissues and reproductive tissues is likely controlled by distinct regulatory mechanisms. Even within the same tissue/organ, such as in fruit, the expression levels of structural genes varied at different developmental stages. The massive accumulation of carotenoids, especially lycopene, in tomato fruit was associated with induced expression of the genes encoding enzymes of lycopene biosynthesis, which was closely aligned to fruit ripening. The RNA-seq data used for eQTL mapping and co-expression analyses were based on field grown S.*lycopersicum* x *S.pennellii* populations from two seasons, and the fruits used for the transcript analyses were harvested when 80 to 100% of M82 tomatoes, (one of the parental lines) were red, which might lead to differences in the stage of ripening among the ILs at harvest. Therefore, to establish the relationship between *SlHONG* and carotenogenesis in tomato fruit, it was necessary to capture the full expression profiles of the structural genes throughout fruit ripening. The peels were removed during the sample preparation, therefore only the pericarp of the fruit was used for checking the expression levels. Nine stages (Figure 4.6) were chosen during fruit development to establish the expression profiles of *SlHONG* and expression of the six most important biosynthetic genes along the pathway; *SlDXS1*, *SlPSY1*, *SlPDS*, *SlZDS* and *SlCRTISO*, among which *SlPSY1*, *SlPDS*, *SlZDS* and *SlCRTISO* were the structural genes responsible for lycopene biosynthesis in tomato fruit, and *SlDXS1* encoding the enzyme catalysing the first step of the isoprenoid biosynthetic pathway in plastids, providing precursors for lycopene biosynthesis (Figure 1.3). Transcript levels were measured by RT-qPCR using gene-specific primers (Appendix 3). Transcripts of *SlPSY1*, catalysing the key step of carotenoid biosynthesis in tomato fruit, could not be detected in the initial phase of fruit ripening , which was consistent with its extremely low abundance in photosynthetic tissues, but started to increase from the mature green (MG) stage, and its expression rose dramatically over the next few days, reaching a maximum at three days after breaker (Figure 4.6B). This high level of expression lasted for a few days but decreased later as the fruit ripened (Figure 4.6B). *SlZISO*, which is almost exclusively expressed in

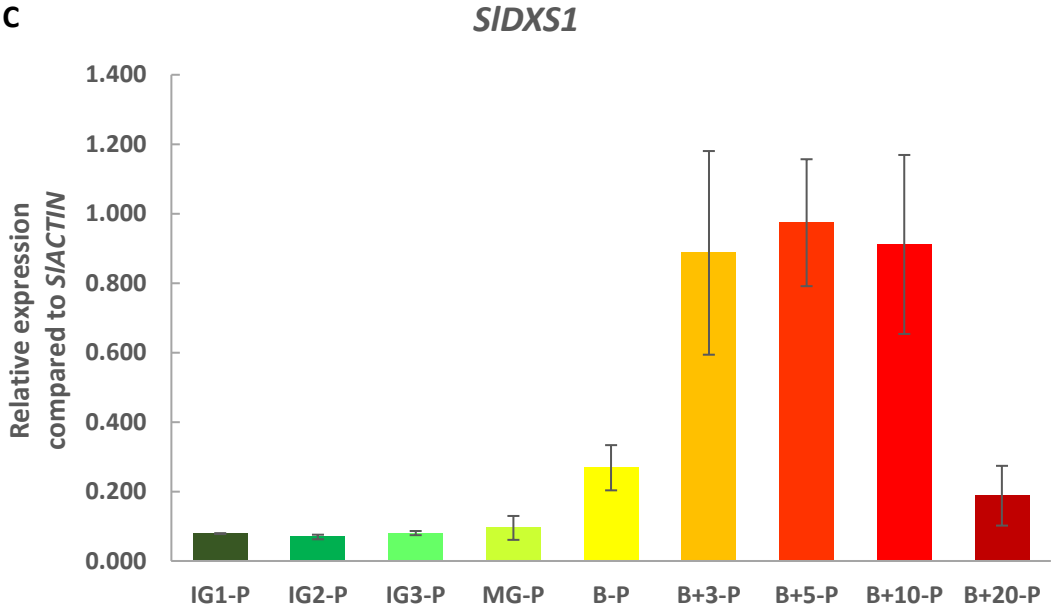
A


SlHONG
Relative expression compared to SlACTIN
0.012
0.010
0.008
0.006
0.004
0.002
0.000
IG1-P
IG2-P
IG3-P
MG-P
B-P
B+3-P
B+5-P
B+10-P
B+20-P

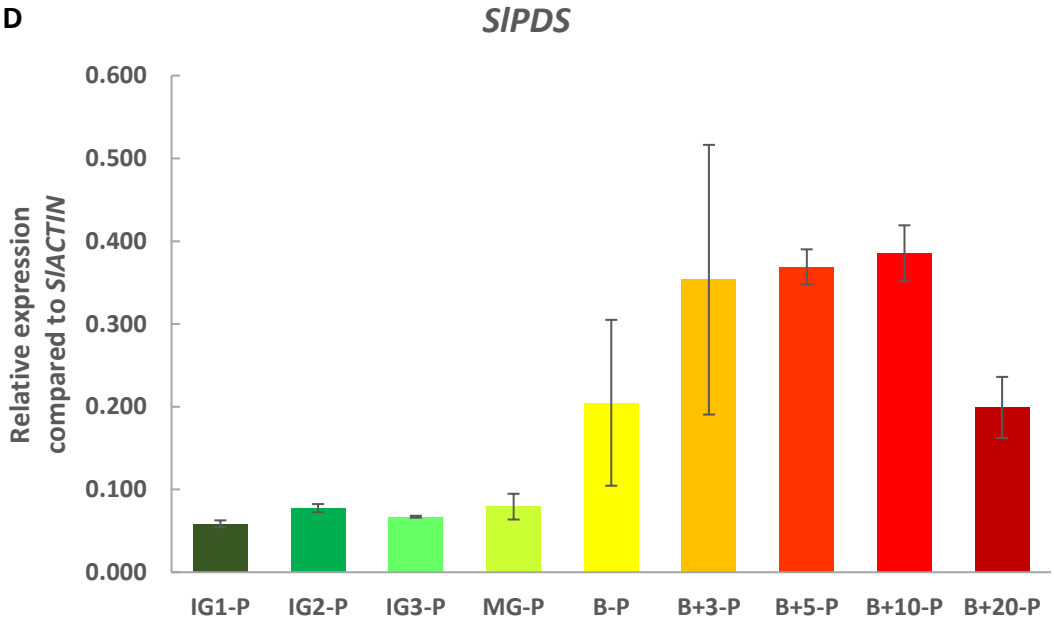
B


SlPSY1
Relative expression compared to SlACTIN
25.000
20.000
15.000
10.000
5.000
0.000
-5.000
IG1-P
IG2-P
IG3-P
MG-P
B-P
B+3-P
B+5-P
B+10-P
B+20-P

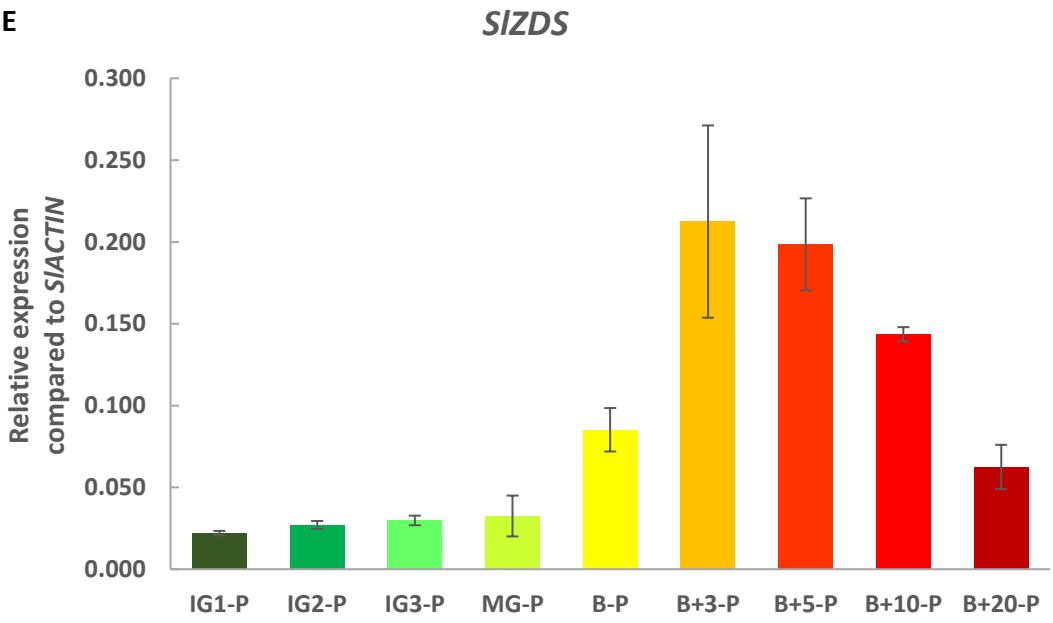
C


SlDXS1
Relative expression compared to SlACTIN
1.400
1.200
1.000
0.800
0.600
0.400
0.200
0.000
IG1-P
IG2-P
IG3-P
MG-P
B-P
B+3-P
B+5-P
B+10-P
B+20-P

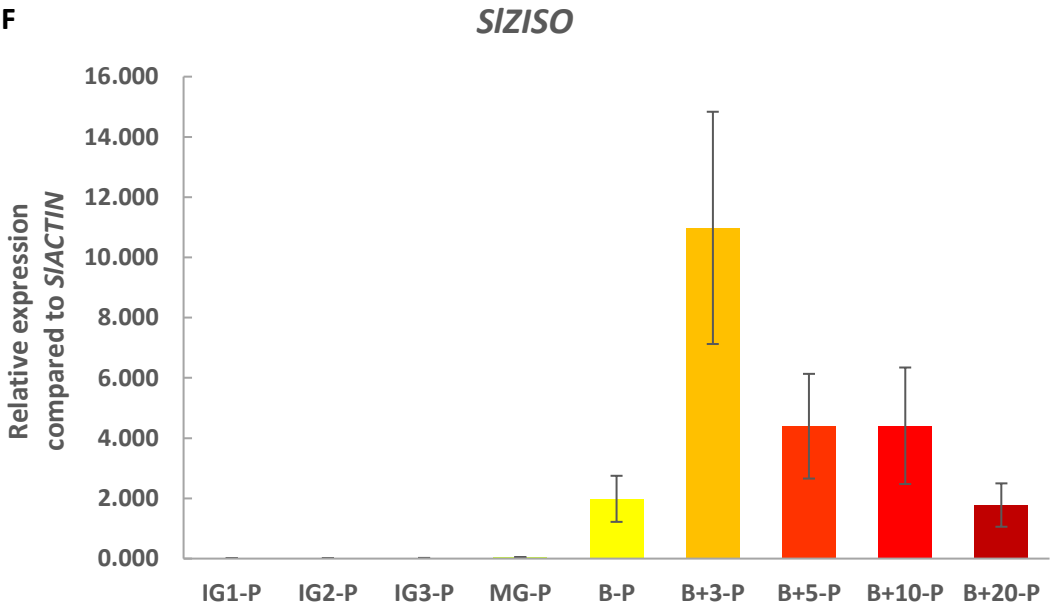
D


SlPDS
Relative expression compared to SlACTIN
0.600
0.500
0.400
0.300
0.200
0.100
0.000
IG1-P
IG2-P
IG3-P
MG-P
B-P
B+3-P
B+5-P
B+10-P
B+20-P

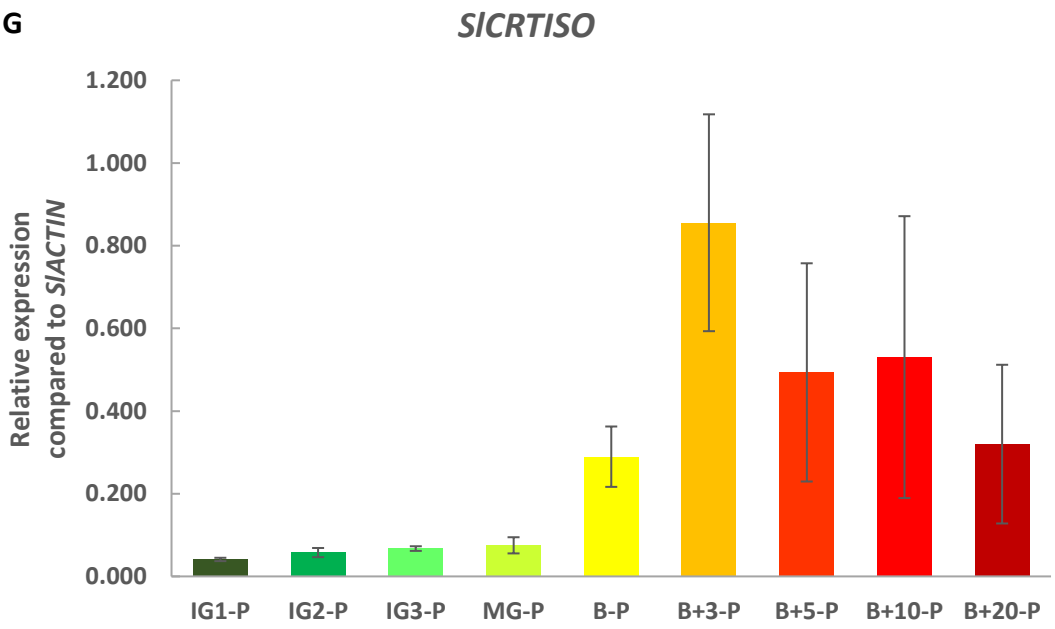
E


SlZDS
Relative expression compared to SlACTIN
0.300
0.250
0.200
0.150
0.100
0.050
0.000
IG1-P
IG2-P
IG3-P
MG-P
B-P
B+3-P
B+5-P
B+10-P
B+20-P

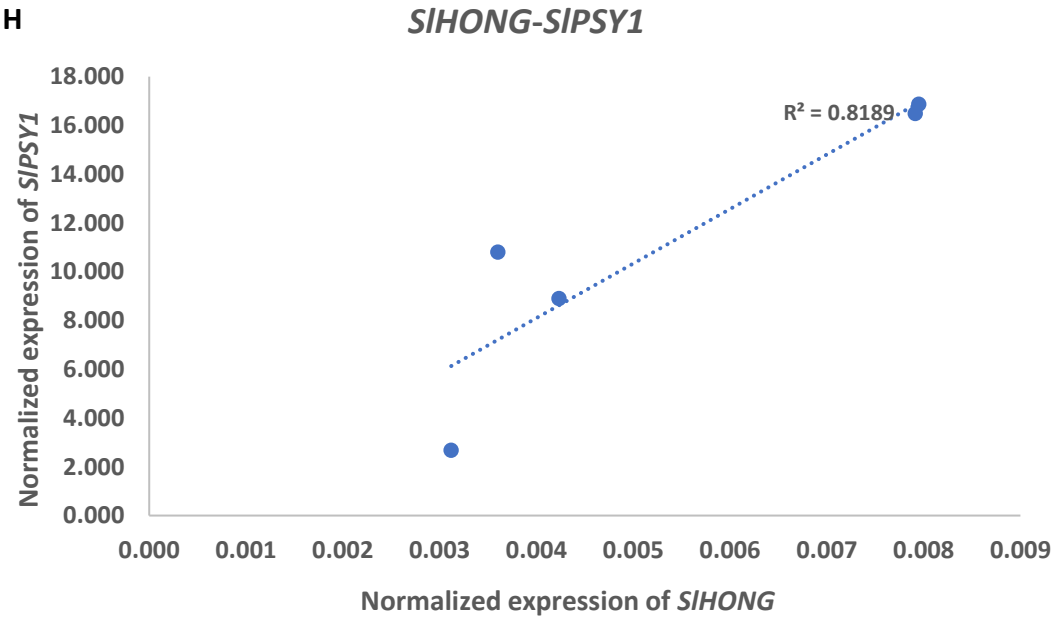
F


SlZISO
Relative expression compared to SlACTIN
16.000
14.000
12.000
10.000
8.000
6.000
4.000
2.000
0.000
IG1-P
IG2-P
IG3-P
MG-P
B-P
B+3-P
B+5-P
B+10-P
B+20-P

G


SlCRTISO
Relative expression compared to SlACTIN
1.200
1.000
0.800
0.600
0.400
0.200
0.000
IG1-P
IG2-P
IG3-P
MG-P
B-P
B+3-P
B+5-P
B+10-P
B+20-P

H


SlHONG-SlPSY1
Normalized expression of SlPSY1
18.000
16.000
14.000
12.000
10.000
8.000
6.000
4.000
2.000
0.000
R² = 0.8189
0.000
0.001
0.002
0.003
0.004
0.005
0.006
0.007
0.008
0.009
Normalized expression of SlHONG

I

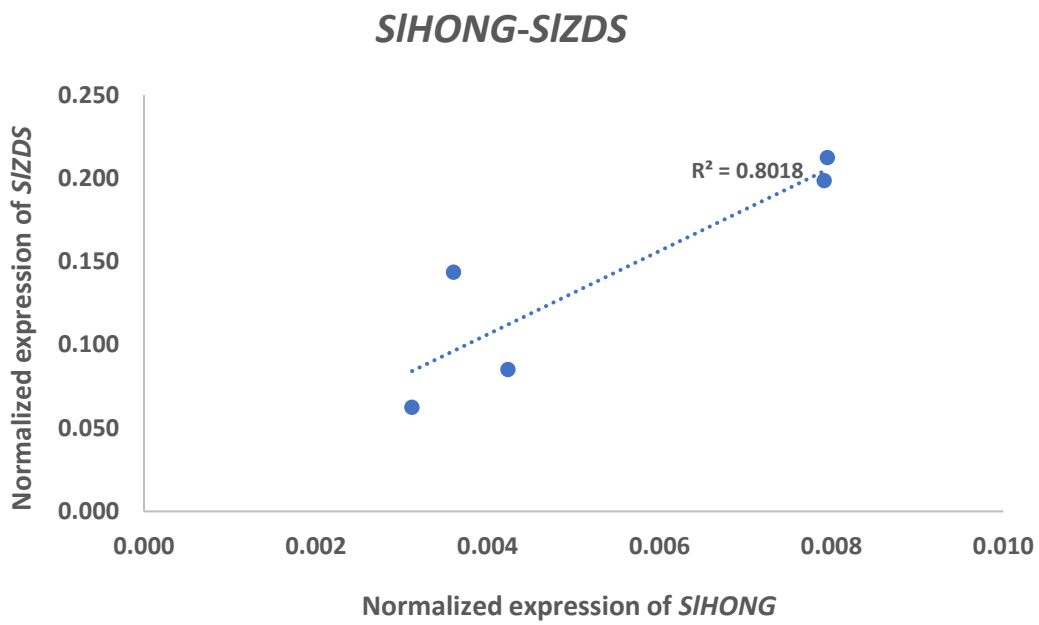


Figure 4.6 Expression of *SlHONG* and lycopene biosynthetic genes in tomato fruit during ripening. A, Expression of *SlHONG* in tomato fruit during ripening; B-G, Expression of lycopene biosynthetic genes in tomato fruit during ripening; H, Linear positive correlation between the expression of *SlHONG* and *SlPSY* after the breaker stage; I, Linear positive correlation between the expression of *SlHONG* and *SlZDS* after the breaker stage. IG1, 7 days after anthesis; IG2, 19 days after anthesis; IG3, 29 days after anthesis; MG, mature green (39 days after anthesis); B, breaker (42 days after anthesis); B+3, 3 days after breaker; B+5, 5 days after breaker; B+10, 10 days after breaker; B+20, 20 days after breaker; P, pericarp. Expression levels shown are relative to that for *ACTIN* (*Solyc03g078400*); error bars, s.e.m.; three biological replicates, each with three technical replicates.

flowers and mature fruit, like *SlPSY1* (Figure 4.5E), showed a very similar expression pattern to *PSY1* during fruit ripening (Figure 4.6F). In contrast, low levels of transcripts of *SlDXS1*, *SlPDS*, *SlZDS* and *SlCRTISO* were detected at green stages of fruit development (Figure 4.6). In general, the six structural genes involved in lycopene biosynthesis showed dynamic changes in their transcript levels during fruit ripening which stayed at very low levels until the breaker stage, increased to a maximum approximately three days later and then decreased along with fruit maturation (Figure 4.6B-G). Unlike the carotenoid biosynthetic genes, the expression of *SlHONG* was maintained at a relatively high level from the beginning of fruit development, achieving an initial maximum at the end of the green stage (42 days after anthesis) (Figure 4.6A). This suggested that *SlHONG* might be involved in the regulation of other physiological processes in tomato fruit apart from controlling lycopene accumulation. During fruit development, the second peak of *SlHONG* expression appeared simultaneously with the expression of the lycopene structural genes and maintained a similar pattern afterwards. Notably, in mature fruit (after the breaker stage), the time-course of expression of *SlHONG* was positively correlated with those of both *SlPSY1* and *SlZDS*, and fitted well to a linear model (Figure 4.6 H and I). Altogether, these data suggested that *SlHONG* might play a positive role in the transcriptional regulation of the carotenoid biosynthetic pathway in tomato fruit, although it is unlikely that this is its exclusive role.

4.3.4 Time-course of lycopene and total carotenoid levels in tomato fruit.

The use of a C18 reverse phase column (Luna®) and high-performance liquid chromatography allowed the detection and separation of different carotenoids using a programme from Ultimate 3000 HPLC systems (Thermo Scientific™), including lycopene. Only traces of lycopene could be detected in fruit at green stages, and levels increased dramatically as the red colour began to develop, from the breaker stage after the ripening process had been triggered by ethylene (Figure 4.7A). Maximum levels of lycopene were reached between five days and ten days after breaker, which followed the maximum expression of genes involved in lycopene synthesis at 3 days after breaker, as well as maximum expression of *SlHONG* (Figure 4.6). Taking the expression data (Figure 4.6) into

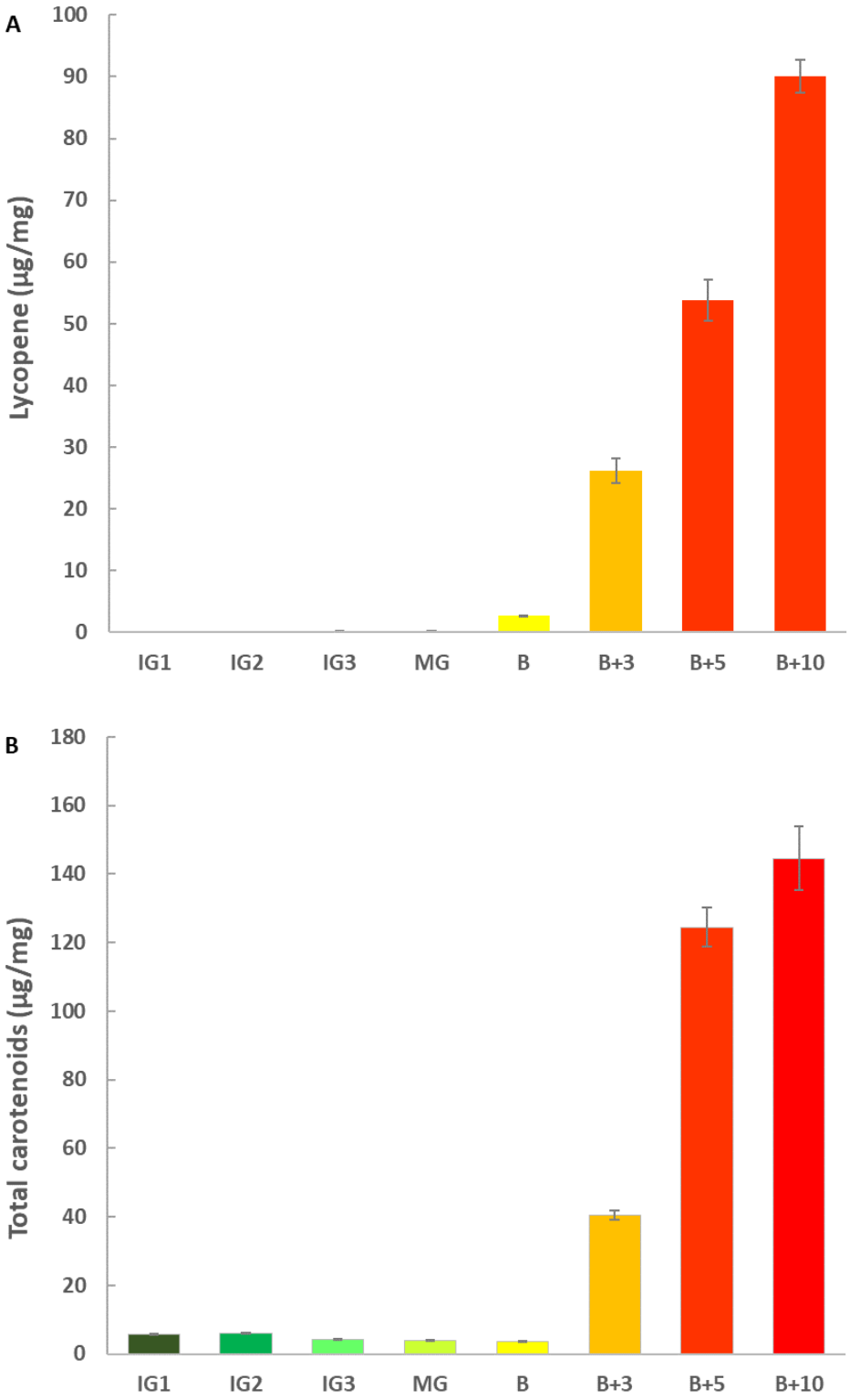


Figure 4.7 Accumulation of lycopene and total carotenoids during ripening of tomato fruit (fresh weight). IG1, 7 days after anthesis; IG2, 19 days after anthesis; IG3, 29 days after anthesis; MG, mature green (39 days after anthesis); B, breaker (42 days after anthesis); B+3, 3 days after breaker; B+5, 5 days after breaker; B+10, 10 days after breaker. Error bars, s.e.m.; five biological replicates, each with three technical replicates.

consideration, the patterns of accumulation of lycopene and total carotenoids (Figure 4.7) were the result of a very active endogenous isoprenoid biosynthetic pathway coupled with the induced expression of genes encoding specific isoforms of lycopene biosynthetic enzymes, which further supported the hypothesis that *SlHONG* is a candidate positive regulator of carotenoid biosynthesis in tomato fruit.

4.3.5 VIGS-silencing in tomato fruit showed that expression of carotenoid biosynthetic genes was positively correlated with the expression of *SlHONG*.

Virus-Induced Gene Silencing (VIGS) has been applied extensively as an efficient method for characterisation of gene function in plants (Baulcombe, 1999). This method has been improved by development of a visual reporter system based on anthocyanin monitoring utilising *Del*/*Ros1* in purple tomatoes (Orzaez et al., 2009). Agroinfiltration of the mature green *Del*/*Ros1* Money Maker fruits with a TRV2 vector harbouring fragments of the *Del* and *Ros1* genes results in a block of anthocyanin production in silenced tissues, leading to red sectors in fruit (Orzaez et al., 2009). It has been demonstrated that silencing of *Del* and *Ros1* in purple tomatoes does not significantly affect lycopene biosynthesis or the production of the pathway intermediates (Fantini et al., 2013). To study the relationship between *SlHONG* and the structural genes of lycopene biosynthesis at the transcriptional level, *Del*/*Ros1* Money Maker tomato fruit at the mature green stage were agroinjected with the TRV1 vector combined with TRV2-*DR* or TRV2-*DR*-*SlHONG* plasmids. According to the time-course expression data (Figure 4.6), *SlHONG* and carotenoid biosynthetic genes reached their highest expression levels around three days after breaker, so fruits showing good VIGS responses (50% silenced surface on average) were harvested and processed at this time point to capture the most significant effects. The effect of *SlHONG* silencing and the corresponding influence on the expression of structural genes were measured by RT-qPCR. *SlHONG* was successfully silenced in red sectors, with less than 10% expression left compared to control sectors (Figure 4.8). All the genes of lycopene biosynthesis were down-regulated to varying extents together with the silencing of *SlHONG*. *SlZDS* and *SlPSY1* were the two genes which showed the most significant down-regulation by silencing of *SlHONG*,

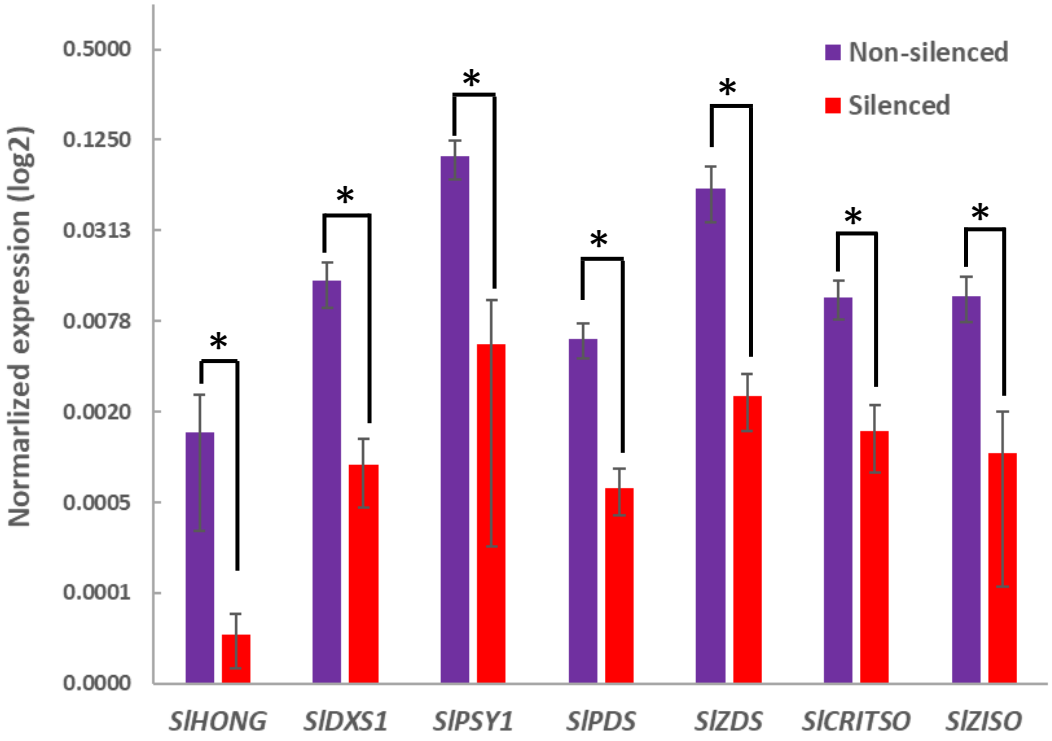


Figure 4.8 Relative expression of *SlHONG* and lycopene biosynthetic genes in the pericarp of *SlHONG* VIGS-silencing Del/Ros1 MoneyMaker tomato. The Agrobacterium-mediated infiltration was conducted at the mature green stage. The samples were taken at three days after breaker. The expression levels were analysed using quantitative PCR, relative to that for *ACTIN* (*Solyc03g078400*). Statistical analysis was performed using a two-tailed t test. * $P < 0.05$. Error bars, s.e.m.; three biological replicates, each with three technical replicates.

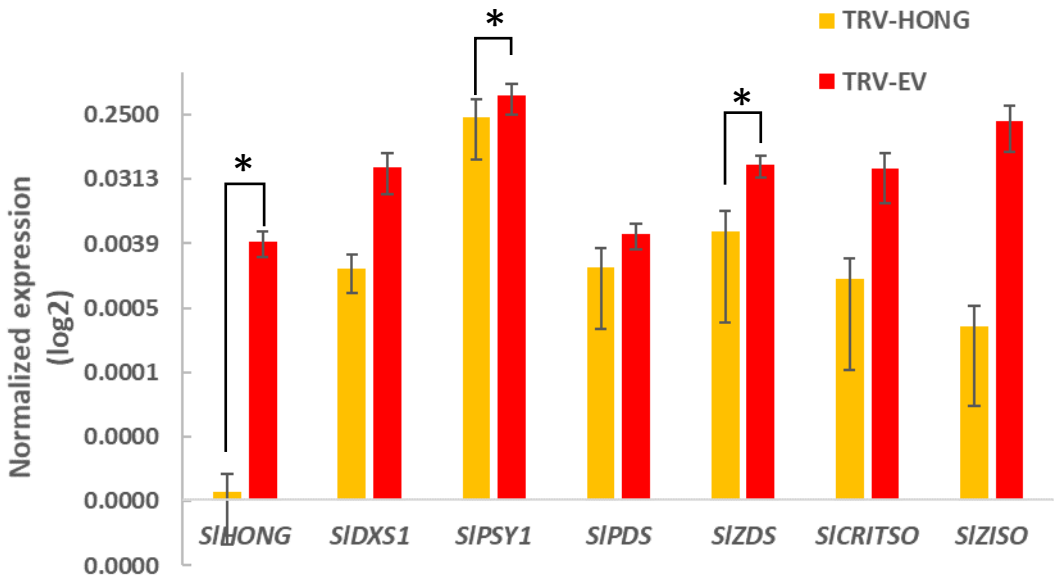


Figure 4.9 Relative expression of *SlHONG* and lycopene biosynthetic genes in the pericarp of *SlHONG* VIGS-silencing WT MoneyMaker tomato fruit. The Agrobacterium-mediated infiltration was conducted at the mature green stage. The samples were taken at three days after breaker. The expression levels of the tested genes were standardised according to *SlACTIN*. Statistical analysis was performed using a two-tailed *t* test. * $P < 0.05$. Error bars, s.e.m.; three biological replicates, each with three technical replicates.

with 24-fold and 17-fold lower expression respectively compared with non-silenced, control sectors (Figure 4.8). These data confirmed the time-course expression profiles which showed linear associations between *SlHONG* and *SlPSY1* expression, and between *SlHONG* and *SlZDS* expression, respectively (Figure 4.6H and I).

My transient analysis of the effects of silencing *SlHONG* during fruit ripening confirmed it to be a regulator of lycopene biosynthesis in *S. lycopersicum*. Although the effect of silencing of *Del* and *Ros1* in purple tomato has been shown to have minimal effects on lycopene accumulation, concerns about interference of the phenylpropanoid metabolic pathway impacting lycopene biosynthesis, remained. Therefore, a similar VIGS experiment was conducted in WT Money Maker tomato fruit without a visual marker for silenced tissues to eliminate the possibility of any potential interference. Whole tomatoes, agroinjected with TRV2-*SlHONG* or TRV2-EV, were harvested separately, three days after breaker. The expression of the carotenoid biosynthetic genes was dramatically decreased with the down-regulation of *SlHONG* in these tomatoes (Figure 4.9), in the same way as in the *Del/Ros1/SlHONG* VIGS experiment, which confirmed that silencing of *SlHONG* in tomato fruit lead to a reduction in the expression of the biosynthetic genes required for lycopene production in tomato fruit. Together, these data suggested strongly that *SlHONG* may be a transcriptional regulator with a positive role in lycopene biosynthesis in tomato fruit.

4.3.6 *SlHONG* knock-out lines were generated by CRISPR/Cas9 genome editing

Cas9, a sequence-specific nuclease guided by RNA, has been developed to generate targeted double-strand breaks in DNA, which are then repaired either by error-prone nonhomologous end joining (NHEJ) or by high-fidelity homologous recombination. The high efficiency of CRISPR/Cas9 system in tomato has already been demonstrated (Brooks et al., 2014), and is more efficient and precise than more traditional methods of gene silencing, such as RNAi. Therefore, CRISPR/Cas9 genome editing was used to generate *SlHONG* knock-out (Δ*SlHONG*) lines.

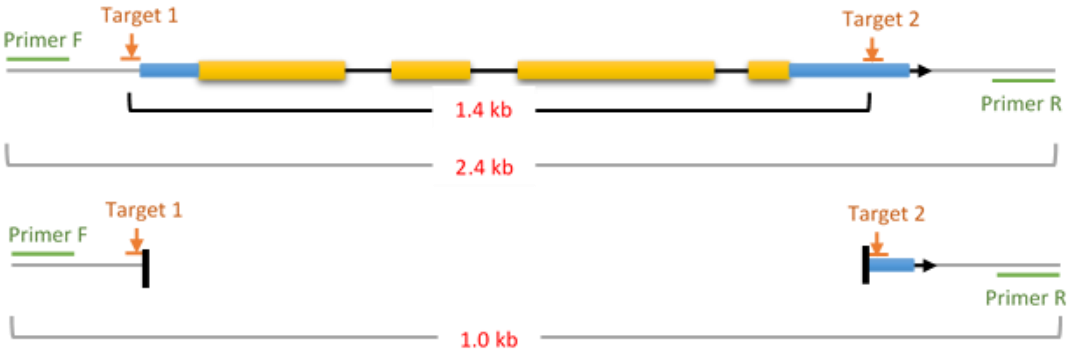


Figure 4.10 Schematic view of CRISPR/Cas9 vector used to generate Δ*SlHONG* lines. Two target sequences (Target 1 and Target 2) adjacent to PAM (NGG) were selected with the intention of creating large, defined deletions, covering the whole genomic sequence of *SlHONG*, in order to ensure a loss of function of *HONG*. Target 1 was in the promoter region, while target 2 was picked from the 3' UTR. The successful mutations driven by the corresponding sgRNAs (Guide 4A and Guide 4B) should result in a deletion about 1.4 kb. Forward and reverse primers, for which binding sites were located outside the deletion region, were designed to select deletion mutants. Yellow boxes, exons; blue boxes, UTRs. Primer F and primer R were used for genotyping.

The CRISPR/Cas9 constructs designed to target *SlHONG* contained two single guide RNAs (sgRNAs) with the intention to create large, defined deletions, covering the whole genomic region to ensure a loss of function of *SlHONG*. Five pairs of sgRNAs harbouring different target sequences adjacent to the PAM motif (NGG) were tested by agroinfiltration in tomato leaves to check their efficiency. One pair of sgRNAs, Guide 4A and Guide 4B was selected for detection of deletions in transient assays in tomato leaves and no additional hits were detected when compared to the tomato genome sequence, which guaranteed the highest efficiency and lowest risk of off-targets. This pair of sgRNAs was designed to target the promoter region (Target 1) and 3'UTR (Target 2) of *SlHONG* respectively to create a deletion of about 1.4 kb encompassing the entire *SlHONG* gene (Figure 4.10).

4.3.7 Tomato lines carrying deletions of the *SlHONG* locus were identified by PCR and validated by sequencing.

45 T0 plants were generated and confirmed by PCR that each carried an integrated transfer DNA (T-DNA) from the introduced CRISPR/Cas9 construct. To detect sgRNA-guided, Cas9-induced deletions of *SlHONG* in the T0 generation, PCR was performed using two primers flanking the sgRNA targets. The wildtype amplicon was predicted to be 2.4 kb while amplicons with the expected deletion of 1.4 kb should be 1 kb (Figure 4.10). Five lines with detected deletions were identified; #4, #16, #28, #33 and #44 (Figure 4.11). To assess further the deletions in these five lines, the shifted bands (indicated by red arrows) were cloned into DONR vectors by Gateway™ Cloning, followed by sequencing (Figure 4.12). The sizes of the deletions in T0 lines #4, #16, #28, #33 and #44 were 1410 bp, 1404 bp, 1412 bp and 1411 bp respectively. *SlHONG* KO lines #4, #16, #28, #44 were edited by sgRNAs near the designed targeted sites, resulting in deletions of around 1.4 kb, whereas a larger deletion (156 bp larger) was detected in line #33, which was because the mutagenesis guided by sgRNA2, occurred about 200 bp downstream of the target site.

The purpose of using CRISPR/Cas9 genome editing was to produce a complete knock-out mutation of *SlHONG*, in case some residual activity of the transcription factor might still

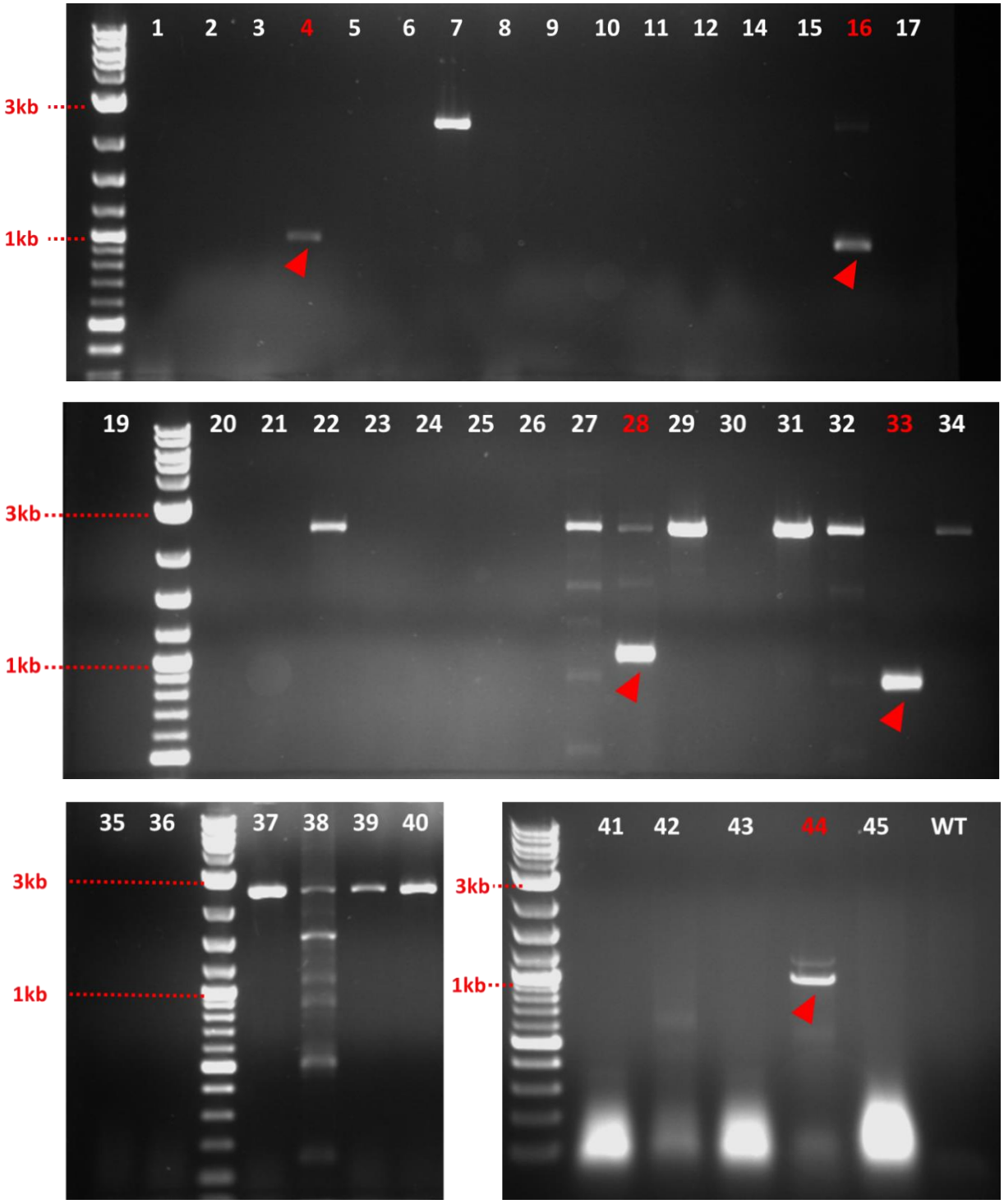
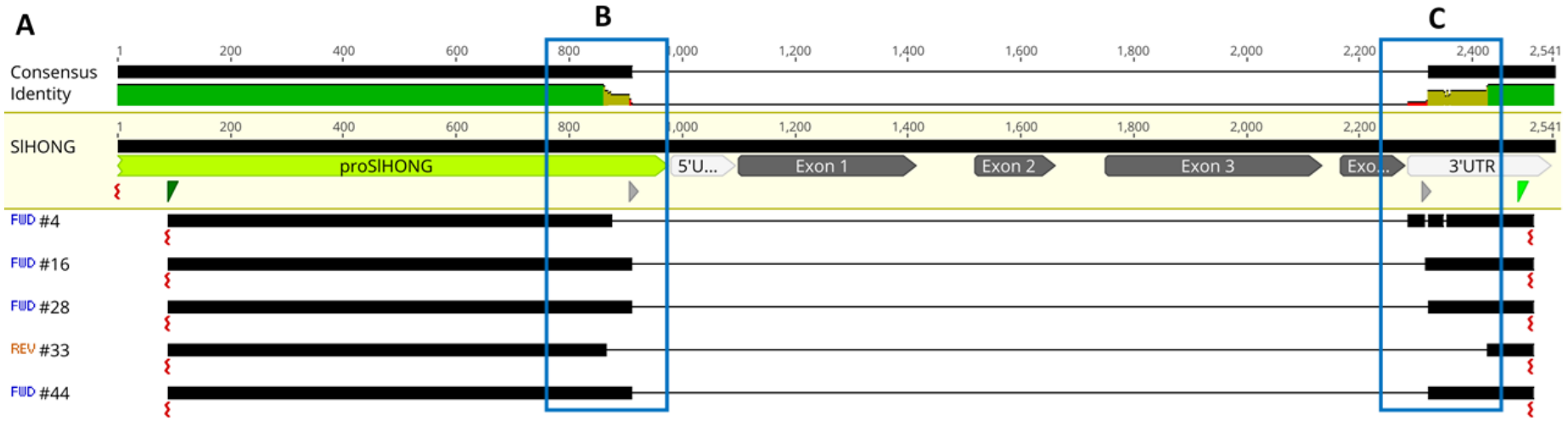
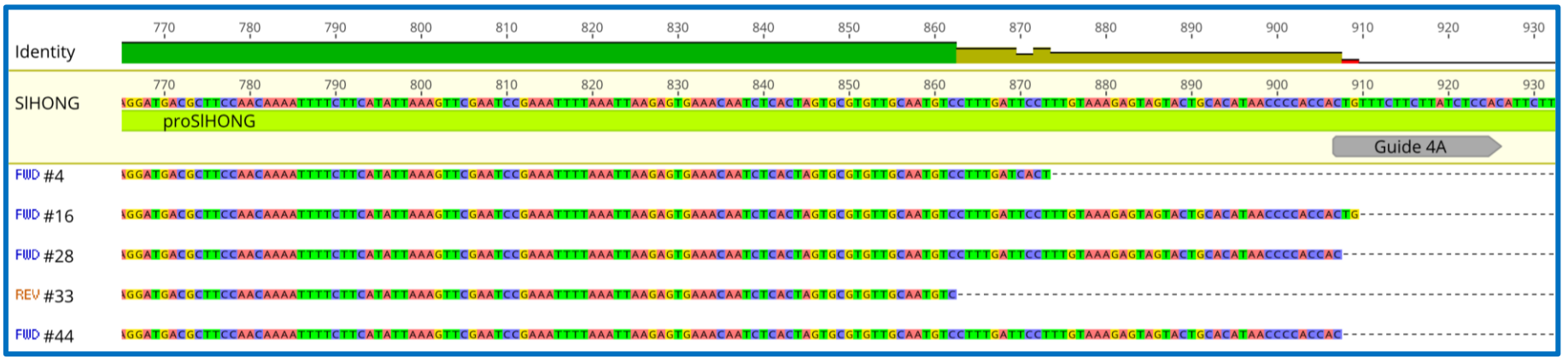


Figure 4.11 Five lines with *ΔSlHONG* deletions (#4, #16, #28, #33 and #44) were identified by PCR using forward and reverse primers shown in Figure 4.10 in the T0 generation. *S. lycopersicum* Money Maker plants were transformed using *A. tumefaciens* carrying the construct shown in Figure 4.2, in which two sgRNAs were designed to target *SlHONG* to create a 1.4kb deletion. Primer F and Primer R in Figure 4.10 were used for genotyping. The expected PCR products of wild-type allele and deletion allele were 2.4 kb and 1kb separately (Figure 4.10). Deleted bands are indicated by red arrows. WT, Wild Type.


A
B
C
Consensus
Identity
SlHONG
proSlHONG
5'U...
Exon 1
Exon 2
Exon 3
Exo...
3'UTR
FWD #4
FWD #16
FWD #28
REV #33
FWD #44


B
Identity
SlHONG
proSlHONG
Guide 4A
FWD #4
FWD #16
FWD #28
REV #33
FWD #44

C

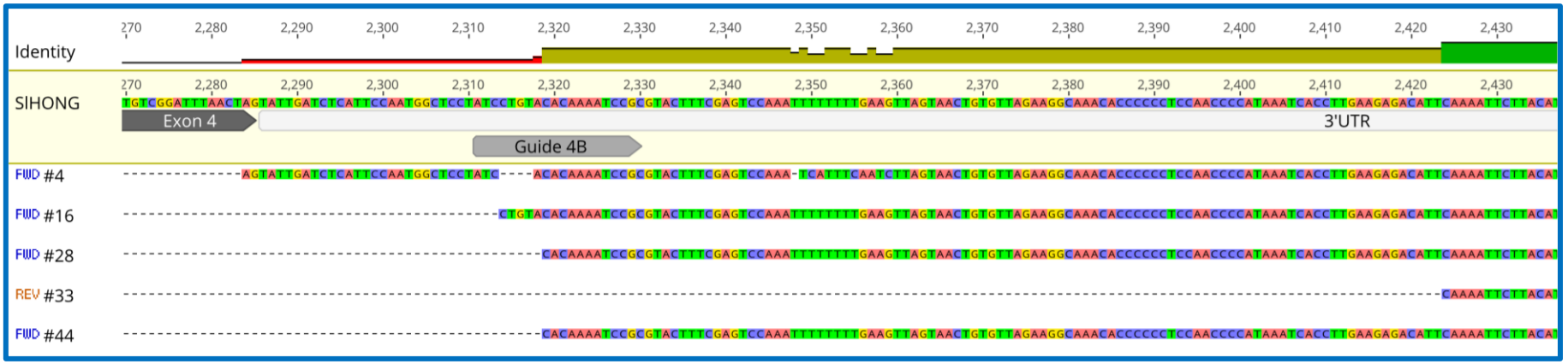


Figure 4.12 Sequencing results of the deleted bands in five lines with *ΔSlHONG* deletions (#4, #16, #28, #33 and #44), indicated in Figure 4.8. A. Schematic view of the deletions in *ΔSlHONG* positive lines. B, Sequencing results of the mutations caused by sgRNA Guide 4A. C, Sequencing results of the mutations caused by sgRNA Guide 4B.

have an effect and complicate the interpretation of the function of *SlHONG*. Therefore, it was essential to analyse the zygosity of the mutants before progressing to functional characterisation. Three regions were selected to check the presence of *SlHOHG* in the genome of transgenic lines, of which one covered the whole coding sequence (1.2 kb) and two covered the junctions of exon 1 and exon 2 (414 bp), and of exon 3 and exon 4 (304 bp) respectively (Figure 4.13 A). These regions could be amplified only with the appropriate primers (Figure 4.13 A) with at least one copy of *SlHONG* in the genome. It turned out that all five *ΔSlHONG* lines identified were heterozygous in the T0 generation, each retaining one wild type allele (Figure 4.12 B, C and D).

Many studies have shown that CRISPR/Cas9 mediated mutagenesis can generate various heritable mutations in plants. However, because Cas9 could still be functional as a nuclease in progeny plants, the maintenance of the CRISPR/Cas9 insert introduces the risk of off-target events. Therefore, it was necessary to segregate out CRISPR/Cas9 construct to produce stably heritable mutants. 16 lines of the T1 generation of *ΔSlHONG* #28 were genotyped to discover a plant which contained the successful deletion allele without the presence of Cas9 (Figure 4.14). However, this plant (~28-2) was heterozygous for the wild type *SlHONG* allele. This line was used for all subsequent analysis.

To further confirm the zygosity of the selected 16 T1 plants of *ΔSlHONG* #28, primers HONG-F and HONG-R (Figure 4.13 A) were used for the amplification of the entire genomic region of *SlHONG* coding sequence. According to the previous genotyping experiments, the competition between the amplification of WT allele and deleted allele could interfere with the results if conducted with genomic DNA, so the cDNA of these plants was used for the analysis. With the primers used, a 918 bp amplicon representing the whole coding sequence of *SlHONG* should result with WT cDNA, while no bands should be amplified by the same pair of primers with homozygous knock-out lines. All the 16 plants of *ΔSlHONG* #28 T1 generation were characterized with 918 bp bands, indicating that those plants contained as least one copy of *SlHONG* WT allele (Figure 4.14 C). No homozygous KO plants were identified in the initial screen of the T1 generation.

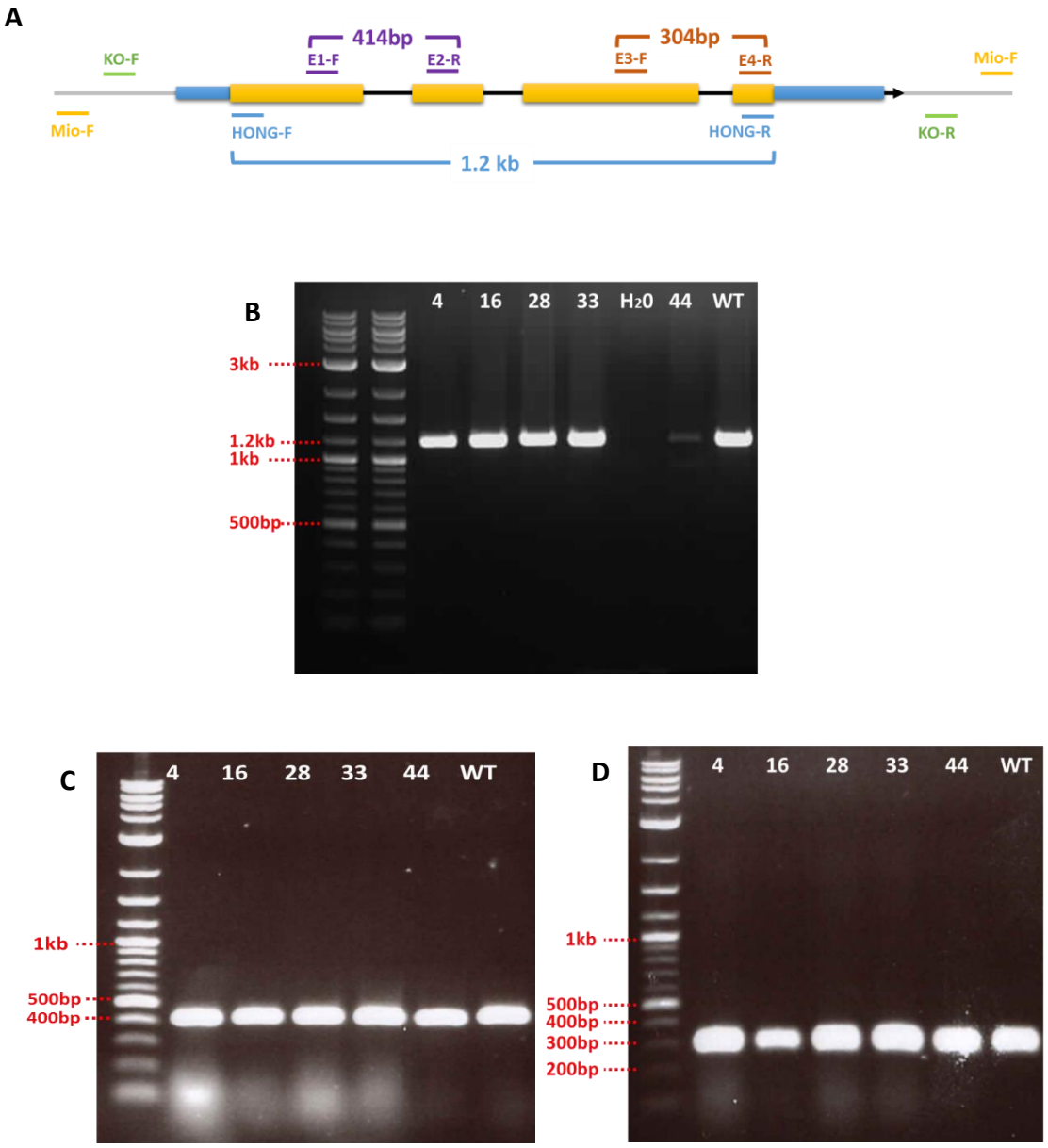


Figure 4.13 Five lines with *ΔSlHONG* deletion alleles (#4, #16, #28, #33 and #44) were identified by PCR as heterozygous in the T0 generation. A, Schematic view of the primers used for identifying the zygosity of the five lines with *ΔSlHONG* deletion alleles. Primer pairs HONG-F/HONG-R, E1-F/E2-R, and E3-F/E4-R were used to amplify the whole genomic region of *SlHONG* (1.2 kb, panel B), exon 1/exon2 junction (414 bp, panel C), and exon 3/exon 4 junction (304 bp, panel D). These regions could only be amplified with the presence of at least one copy of *SlHONG*. All the five lines were detected with the wild-type alleles of *SlHONG* by the indicated primer pairs, suggesting that these lines were heterozygous. H_2O, negative control. WT, Wild Type.

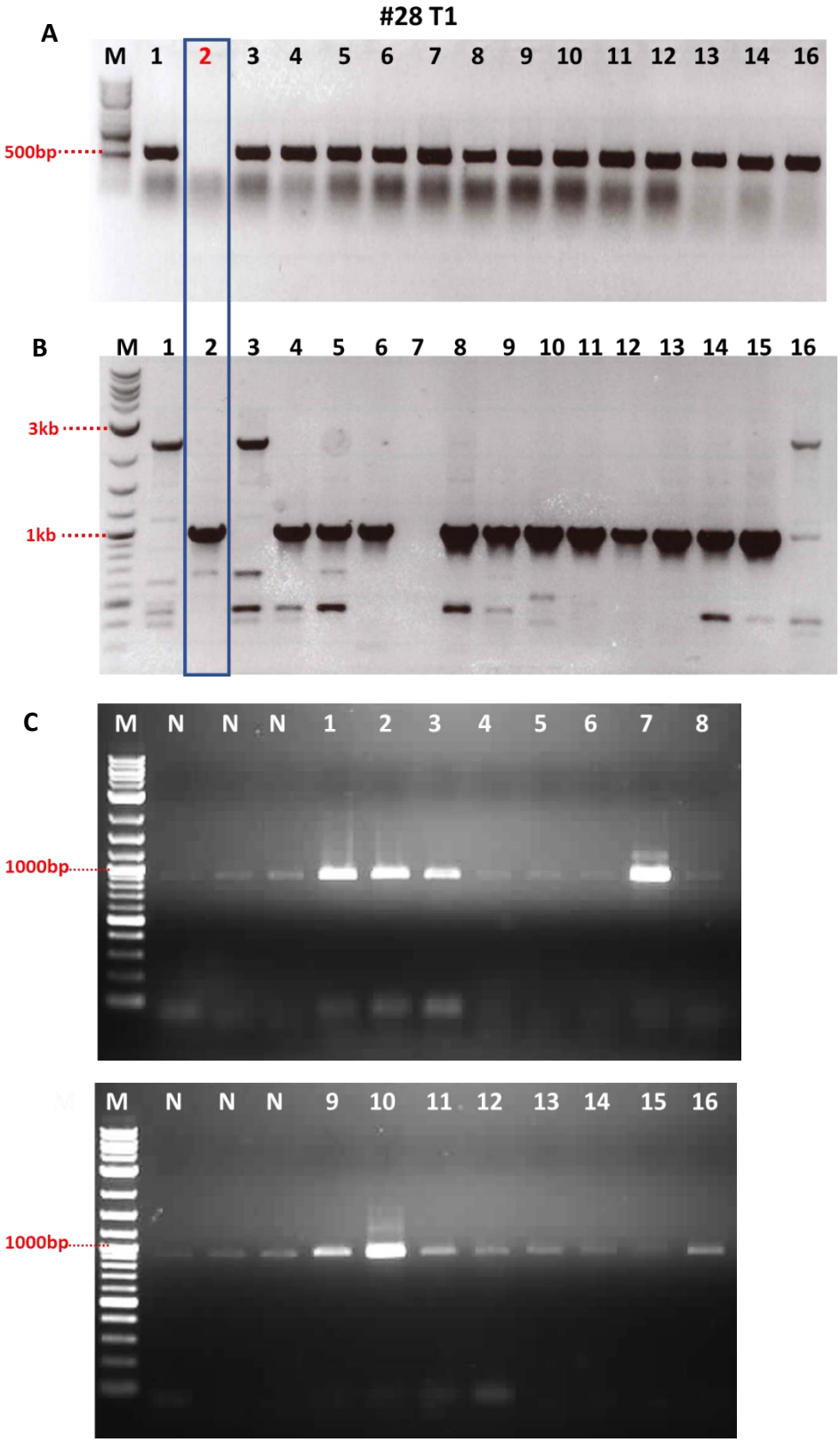
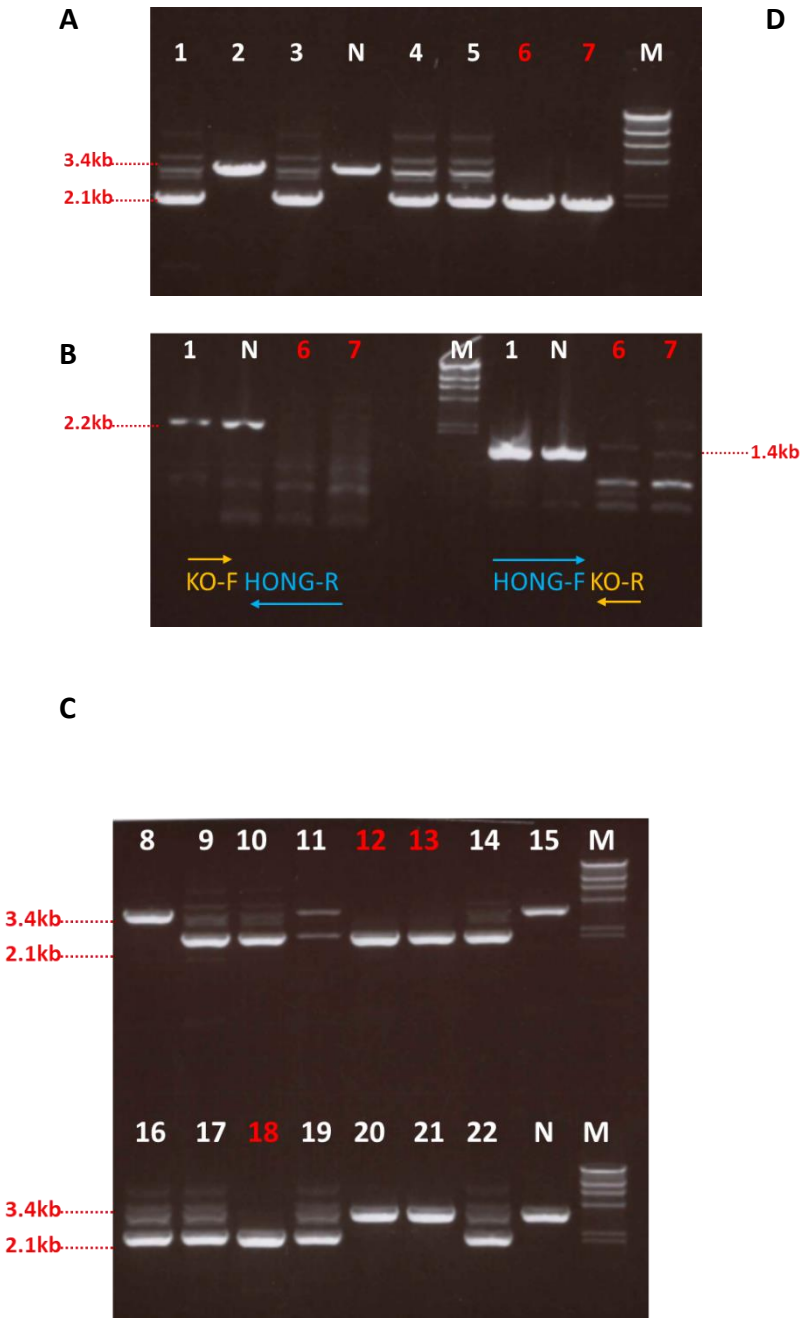


Figure 4.14 CRISPR-*HONG* #28 T1 generations were analysed by PCR. A, *Cas9*-specific primers were used to detect the presence of *Cas9* in the genome. In 16 CRISPR-*HONG* #28 T1 plants, #28-2 were identified as the absence of *Cas9*. B. Among the tested plants, 14 out of 16 plants contained *ΔSlHONG* deletion alleles, including #28-2, by PCR using the primer pair shown in Figure 4.10. C, Zygosity of the 16 plants were analysed by *SlHONG*-specific primers, HONG-F/HONG-R (Figure 4.13), using cDNA as templates. Wild-type alleles were 918 bp. All the 16 plants were identified as heterozygous. These data indicated that *Cas9* had been segregated out in #28-2. M, 2log DNA ladder. N, Non-silenced plant.

4.3.8 Homozygous *SlHONG*-KO (*ΔSlHONG*) plants were identified by PCR in the T1 generation of *ΔSlHONG* #4 and #28 T0 plants

With the help of Dr Eugenio Butelli, a new pair of primers was designed (named Mio-F and Mio-R) to characterise the *SlHONG* alleles in the 'mutants', following the same design principle as KO-F and KO-R, but giving relatively longer PCR fragments, ~3.4 kp and ~2.1 kb for WT allele and deletion alleles respectively, to minimise the competition between the amplification of *SlHONG* WT and deletion alleles during PCR (Figure 4.13 A). Seedlings were sown on agar without selection and leaf material was harvested from the first true leaves of progeny of T0 plants. Three homozygous *SlHONG*-KO (*ΔSlHONG*) plants were identified in progeny from line *ΔSlHONG* #4 and two homozygous *ΔSlHONG* plants were identified in the progeny of *ΔSlHONG* #28 (Figure 4.15 A). Another two pairs of primers were used to further confirm the homozygosity of these plants (*ΔSlHONG* #4 T1-8 and *ΔSlHONG* #4 T1-10), KO-F and HONG-R, and HONG-F and KO-R. Because for each primer pair, one of the primers, HONG-R and HONG-F was designed that lay within the deleted region, no predicted bands of the size in WT (about 2.2 kb and 1.4 kb respectively), could be amplified in homozygous *ΔSlHONG* mutants. Thus, *ΔSlHONG* #4 T1-8 and *ΔSlHONG* #4 T1-10 were confirmed as homozygous *SlHONG*-KO (*ΔSlHONG*) plants (Figure 4.15 B). The same genotyping was conducted in other *ΔSlHONG* #4 and #28 T1 plants. For T0 line *ΔSlHONG* #4, 7 more T1 plants were characterised, and 3 *ΔSlHONG*/ *ΔSlHONG*, 6 *SlHONG*/ *ΔSlHONG* and 2 *SlHONG/SlHONG* were identified; Among the 11 plants tested amongst the T1 progeny of the T0 line *ΔSlHONG* #28, the numbers of plants harbouring *ΔSlHONG*/ *ΔSlHONG*, *SlHONG*/ *ΔSlHONG* and *SlHONG/SlHONG* alleles were 2, 6 and 3, respectively (Figure 4.15 A and C), which is summarised in Figure 4.15 D. The segregation ratios of both lines were very close to a Mendelian segregation ratio of 1:2:1. Therefore, five homozygous *ΔSlHONG* plants were identified in total, from two independent deletion events.



D

No.	Samples	Genotype
1	ΔSlHONG #28 T1-1	*HONG/hong*^{del}
2	ΔSlHONG #28 T1-3	*HONG/HONG*
3	ΔSlHONG #28 T1-4	*HONG/hong*^{del}
4	ΔSlHONG #4 T1-2	*HONG/hong*^{del}
5	ΔSlHONG #4 T1-6	*HONG/hong*^{del}
6	ΔSlHONG #4 T1-8	*hong*^{del}*/hong*^{del}
7	ΔSlHONG #4 T1-10	*hong*^{del}*/hong*^{del}
8	ΔSlHONG #28 T1-6	*HONG/HONG*
9	ΔSlHONG #28 T1-7	*HONG/hong*^{del}
10	ΔSlHONG #28 T1-9	*HONG/hong*^{del}
11	ΔSlHONG #28 T1-13	*HONG/hong*^{del}
12	ΔSlHONG #28 T1-14	*hong*^{del}*/hong*^{del}
13	ΔSlHONG #28 T1-15	*hong*^{del}*/hong*^{del}
14	ΔSlHONG #28 T1-17	*HONG/hong*^{del}
15	ΔSlHONG #28 T1-18	*HONG/HONG*
16	ΔSlHONG #4 T1-11	*HONG/hong*^{del}
17	ΔSlHONG #4 T1-12	*HONG/hong*^{del}
18	ΔSlHONG #4 T1-13	*hong*^{del}*/hong*^{del}
19	ΔSlHONG #4 T1-14	*HONG/hong*^{del}
20	ΔSlHONG #4 T1-16	*HONG/HONG*
21	ΔSlHONG #4 T1-19	*HONG/HONG*
22	ΔSlHONG #4 T1-20	*HONG/hong*^{del}

Figure 4.15 Five *SlHONG*-KO lines were identified. A and C, Primers Mio-F/Mio-R (Figure 4.13) were used to distinguish *SlHONG* WT alleles (3.4 kb) and *SlHONG* deletion alleles (2.1 kb); B, Primers KO-F/HONG-R, and HONG/KO-R were used to amplify parts of the genomic region of *SlHONG.* KO-F and KO-R were designed within the deletion region (Figure 4.13). D, Information and genotyping result of each sample was summarised in D. M, DNA marker. N, Negative control.

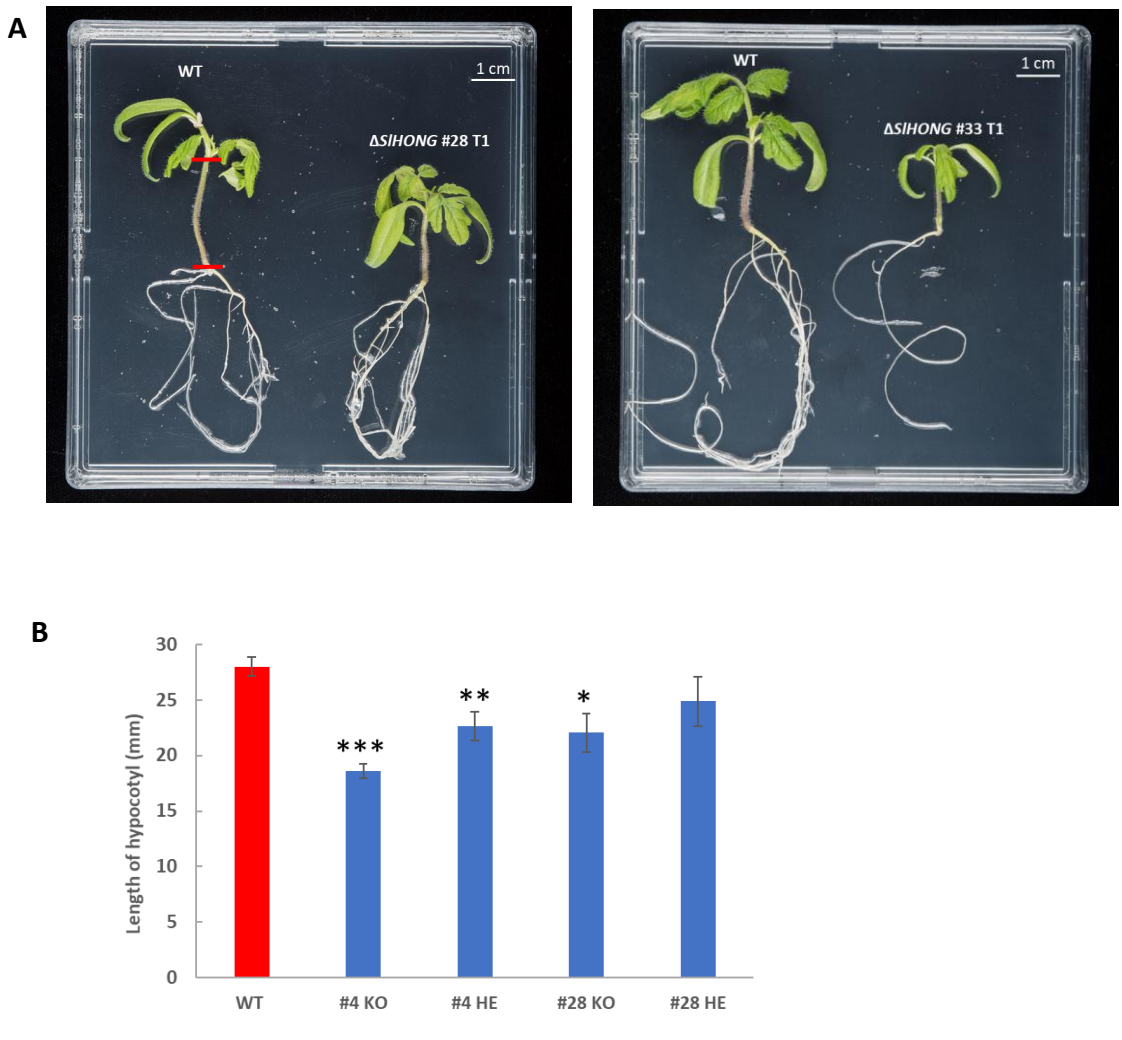


Figure 4.16 Hypocotyl length of *ΔSlHONG* lines was significantly smaller compared to that of WT. A, Representative hypocotyl lengths of wild type and heterozygous *ΔSlHONG* lines (#28 and #33). The hypocotyl was indicated as the part between the red lines; B, Qualification of hypocotyl lengths of the wild type (WT), homozygous *ΔSlHONG* lines (#4 and #28). KO, $hong^{del}/hong^{del}$; HE, $HONG/hong^{del}$. Statistical analysis was performed using a two-tailed t test. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. Data is the average ± s.e.m. of at least 10 biological replicates.

4.3.9 Homozygous *ΔSlHONG* exhibited significantly reduced hypocotyl elongation compared to WT seedlings

It was observed that T1 progeny of *ΔSlHONG* #28 and *ΔSlHONG* #33 exhibited shorter hypocotyls compared to the WT seedlings of the same age (Figure 4.16 A). More precise measurements were performed on the same seedlings with confirmed genotypes from the progeny of *ΔSlHONG* #4 and #28, which were used in the identification of homozygous *ΔSlHONG* lines. The hypocotyl length of the homozygous *ΔSlHONG* lines was significantly shorter than WT Money Maker seedlings. (Figure 4.16 B). Reduced hypocotyl elongation was also observed in heterozygous lines compared to WT (Figure 4.16 B). The alteration in hypocotyl length caused by the deletion of *SlHONG* demonstrated that SlHONG plays a positive role in hypocotyl elongation.

Due to the difficulties I encountered in identifying homozygous *ΔSlHONG* lines, the phenotype of the fruit of the heterozygous *ΔSlHONG* /*SlHONG* T0 plants (#28) were examined for effects on regulation of lycopene biosynthesis in tomato fruit. In the meanwhile, genotyping of more T1 progeny of *ΔSlHONG* #28 and of progeny of other *ΔSlHONG* lines was undertaken to identify homozygous *ΔSlHONG* tomato plants, and functional characterisation of homozygous *ΔSlHONG*/*ΔSlHONG* lines will follow. To capture initially the functions of SlHONG in the transcriptional regulation of lycopene biosynthesis in tomato fruit, heterozygous T0 line *ΔSlHONG* #28-2 was used for the subsequent analysis. Functional analysis of the homozygous deletion alleles will be completed in the future.

4.3.10 Heterozygous Δ*SlHONG* lines showed a similar phenotype to IL 2-1 compared to WT during fruit ripening.

It has been shown that during fruit ripening, introgression line IL2-1 turns red more slowly than M82 (Figure 3.8), a feature that was closely linked to the rate of accumulation of lycopene (Chapter 3). A time-course of fruit ripening compared the colour phenotypes between the heterozygous Δ*SlHONG* line #28-2, the isogenic wildtype control (Money

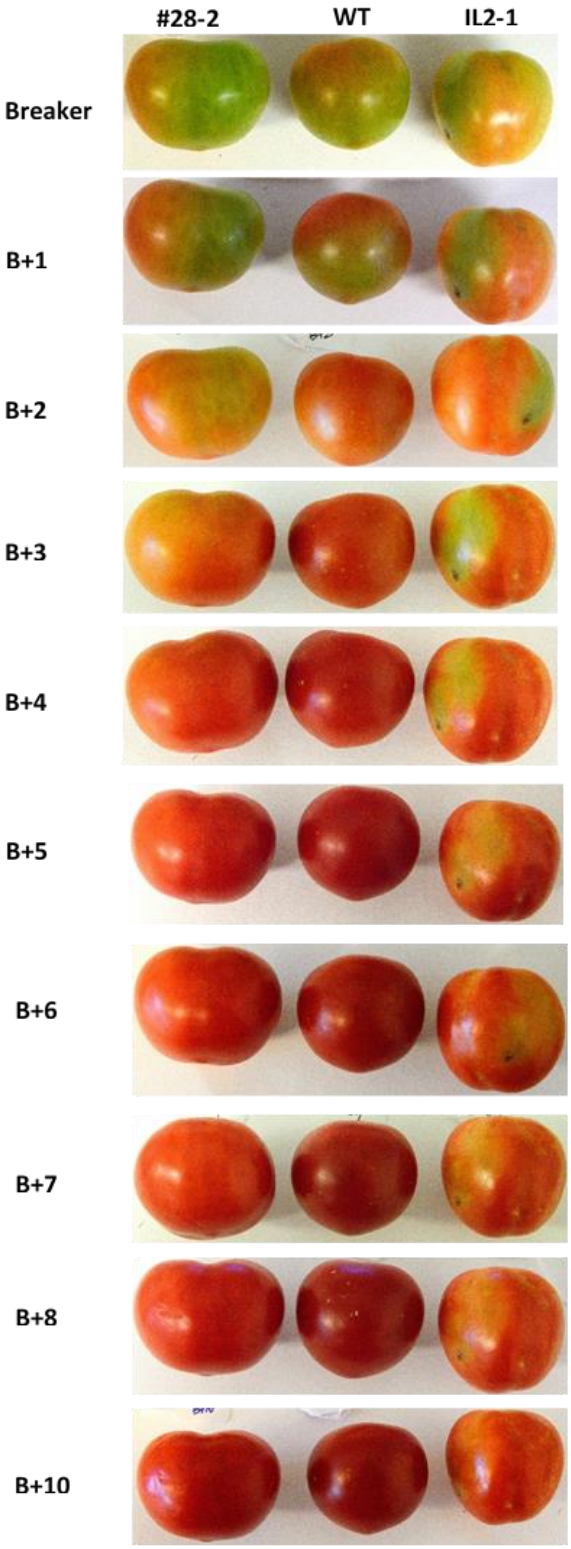


Figure 4.17 Positive Δ*SlHONG* line #28-2 showed delayed lycopene accumulation during fruit development, which was similar to the phenotype of IL2-1. This phenotype, which is possibly associated with the slower accumulation of carotenoids, especially lycopene, needs further validation in lines homozygous for the Δ*SlHONG* allele. B+N, N day (s) after breaker.

Maker) and IL 2-1 (Figure 4.17). Compared to Money Maker wildtype fruit, #28-2 showed delayed red colouration from the beginning of the breaker stage, which was similar to the phenotype observed in IL 2-1. The colour differences increased during fruit ripening and reached a maximum about seven days after breaker (Figure 4.17), consistent with the observation that peak levels of lycopene accumulation occurred between five days and ten days after breaker (Figure 4.7 A). By the end of the observation period, which was 20 days after breaker, the differences in red colouration remained. Although the rate of colouration was distinct between #28-2 and wildtype, lycopene appeared to accumulate homogeneously in fruit of both lines, which was different from the patchy pattern of pigmentation observed in IL 2-1 (Figure 4.17). Moreover, the colouration in fruit of IL 2-1 was slower than in #28-2 (Figure 4.17). These differences in phenotype were likely due to the heterozygosity of *ΔSlHONG* #28-2, or perhaps due to the activity of other genes replaced in IL 2-1 by *S.pennellii* homologs that contributed to the variation between #28-2 and IL 2-1. Further validation is needed to establish the accumulation of lycopene in fruit of homozygous Δ*SlHONG* mutants.

4.3.11 *SlPSY1* is up-regulated in fruit-specific *SlHONG* overexpression lines

Two types of transgenic tomatoes with stably overexpressed *SlHONG* were generated following a standard tomato stable transformation protocol. *SlHONG* was constitutively overexpressed in the whole plant driven by the double cauliflower mosaic virus (CaMV) 35S promoter. Forty 35S::*SlHONG* positive lines were identified in the T0 generation (Figure 4.18). The expression levels of *SlHONG* in young leaves were assessed by RT-qPCR. Ten lines were identified with more than ten-fold higher expression of *SlHONG* compared to wildtype, especially in #9 and #12, where the mRNA abundance was nearly 60 times higher than in Money Maker (Figure 4.19). These lines were selected for further analysis of the function of *SlHONG* in different tissues (particularly in flowers and vegetative tissues).

Although the CaMV 35S promoter can boost the expression of genes of interest in plants, this promoter does not confer any developmental stage specificity or tissue specificity to

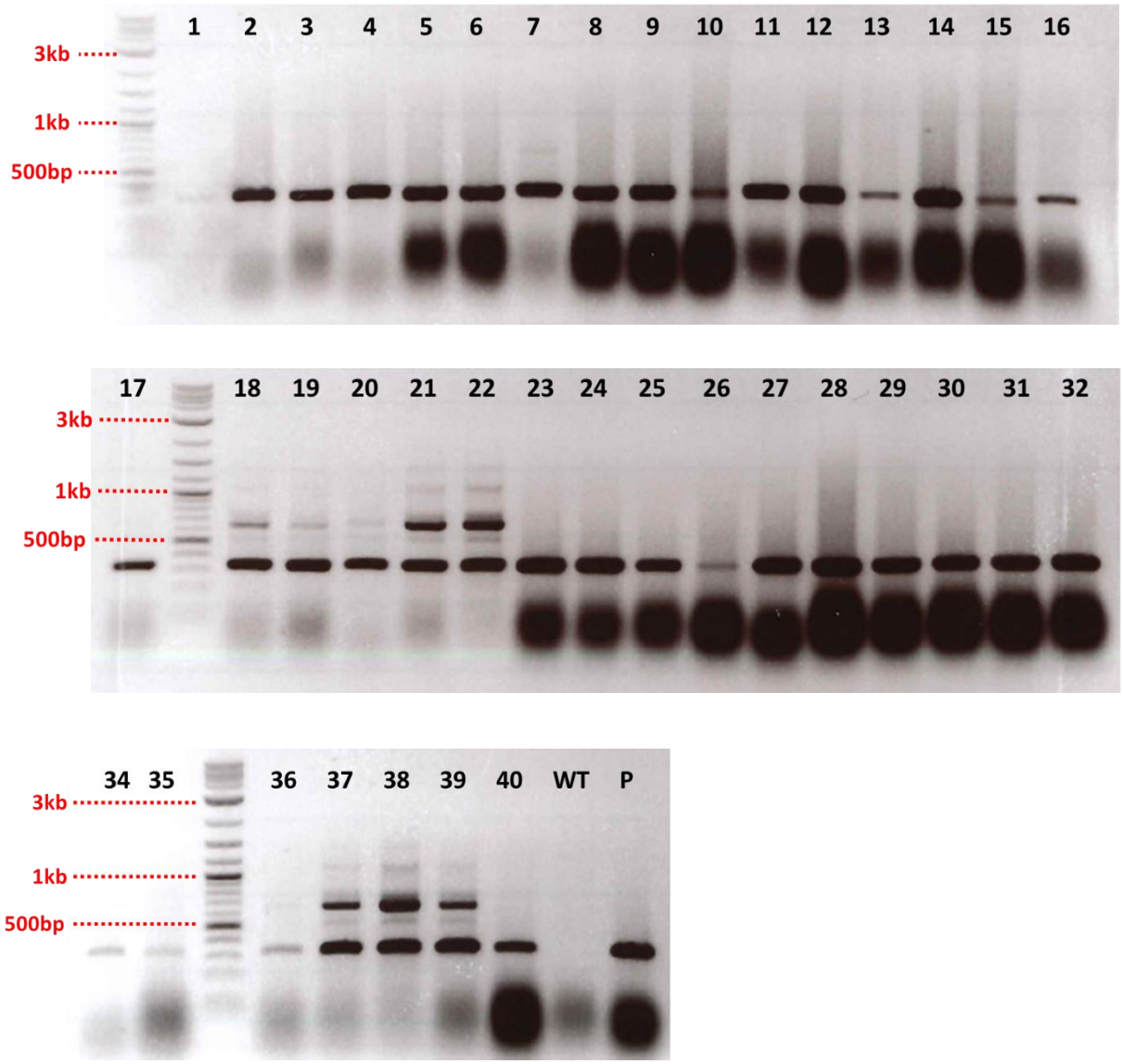


Figure 4.18 40 positive lines were identified by PCR in the T0 generation of 35S::*HONG* transgenic lines. The forward and reverse primers used were designed based on the sequences of 35S promoter and *SlHONG* separately, and the PCR products were 344 bp or/and 683 bp. WT, Wildtype (negative control); P, plasmid pBin19-2x35S::*HONG* (positive control, Figure 4.1).

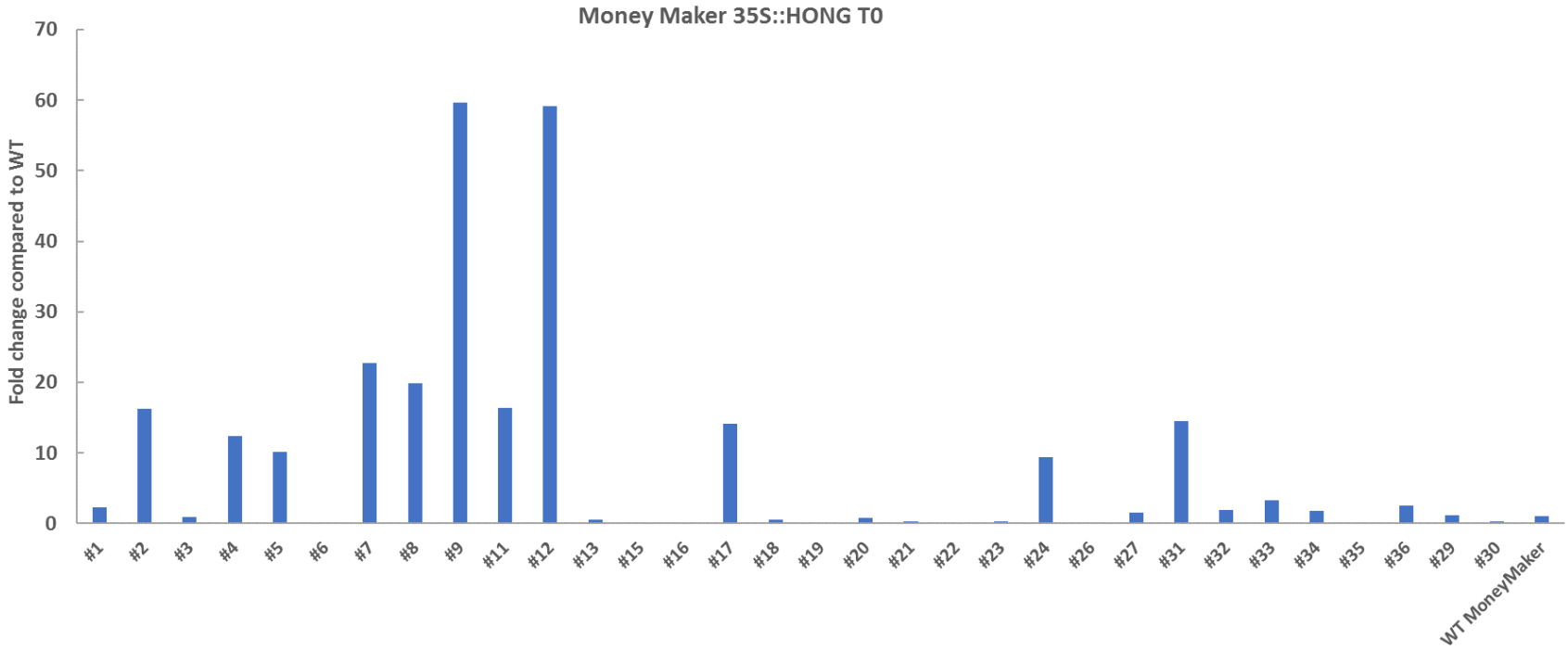


Figure 4.19 Fold changes of expression levels of *SlHONG* in leaves of 35S::*SlHONG* T0 generation, compared to WT Money Maker. The expression of *SlHONG* in leaves was analysed using quantitative PCR. Expression levels shown are relative to that for *ACTIN* (*Solyc03g078400*).

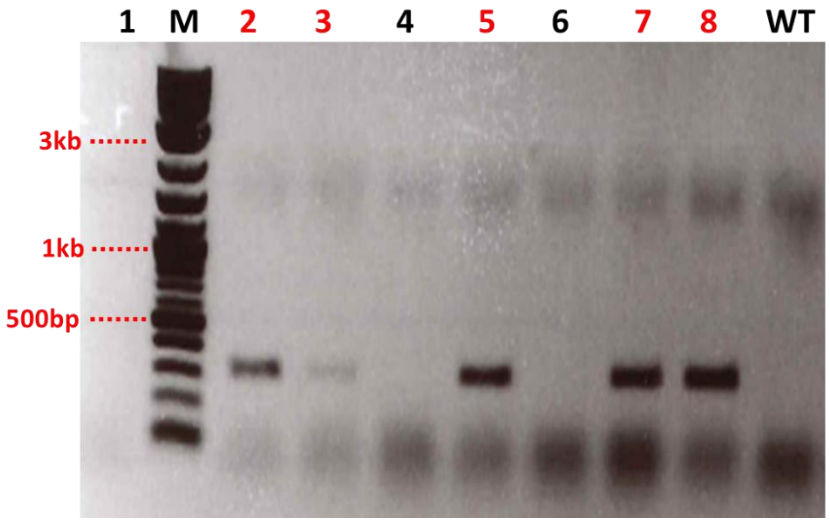


Figure 4.20 Five positive lines (#2, #3, #5, #7 and #8) were identified by PCR in the T0 generation of E8::*HONG* transgenic lines. The forward and reverse primers used were designed based on the sequences of 35S promoter and *SlHONG* separately, and the PCR products were 309 bp. WT, Wildtype (negative control); M, marker.

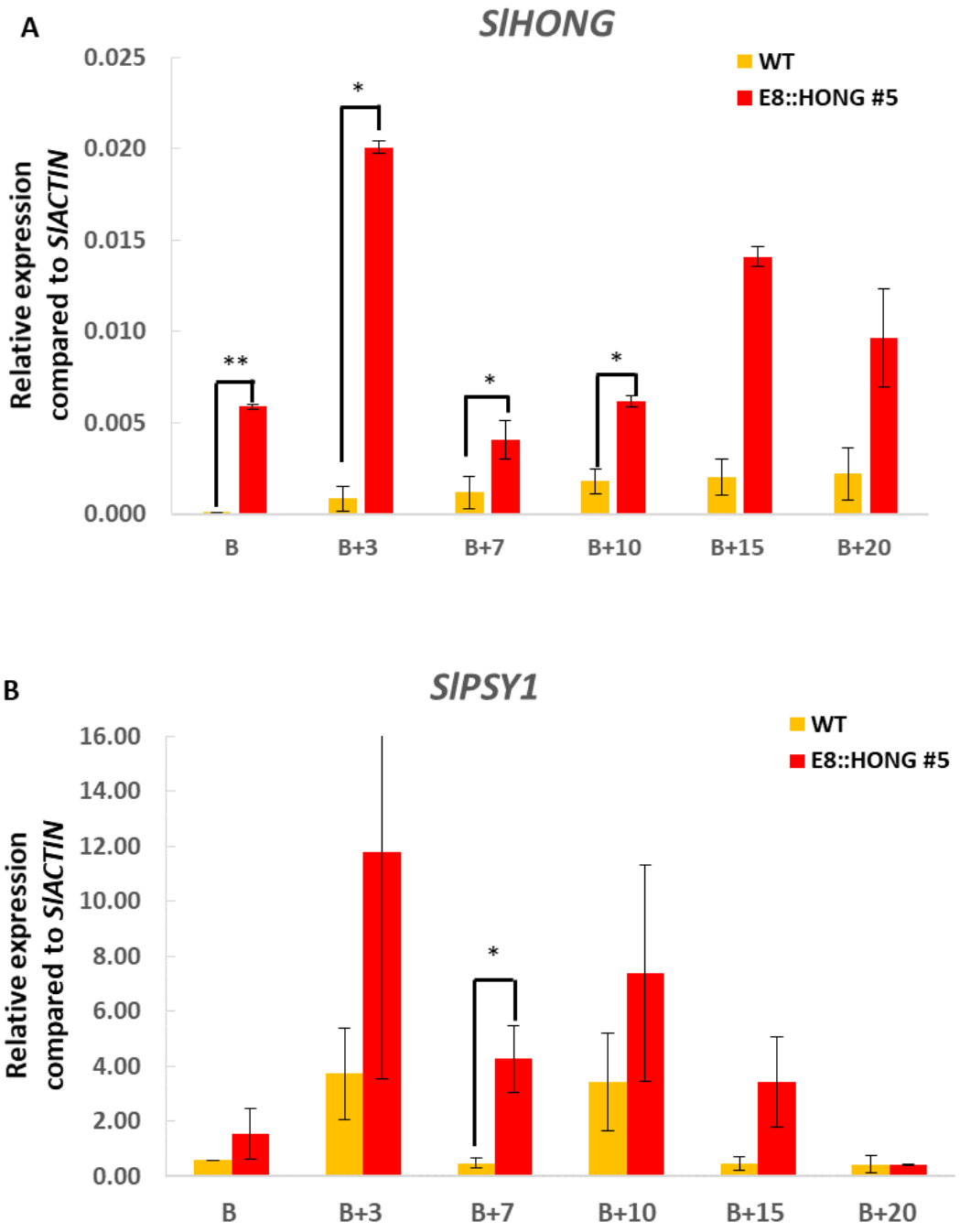


Figure 4.21 The expression level of *SlPSY1* was significantly increased in E8::*SlHONG* overexpression transgenic line #5 (T1). The expression of was analysed using quantitative PCR. Expression levels shown are relative to that for *ACTIN* (Solyc03g078400). Statistical analysis was performed using a two-tailed *t* test. * $p < 0.05$, ** $p < 0.01$. Error bars, s.e.m.; three biological replicates, each with three technical replicates. B, Breaker. B+N, N days after breaker.

the expression of the target gene. The expression of SlHONG suggested that it might play multiple roles in different plant tissues outside of fruit, which might complicate analysis of its role in controlling lycopene accumulation in fruit. For example, possible growth defects at the early stage of development caused by the overexpression of the target gene might interfere with the characterisation of gene function in later stages of fruit ripening. Moreover, previous studies in the lab had suggested that genes driven by the 35S promoter were not highly overexpressed in the pericarp of the fruit, where most of the lycopene accumulates in tomato. For these reasons, the fruit-specific, E8 promoter was also used to generate transgenic tomatoes with SlHONG overexpressed specifically in tomato fruit from breaker stage, when lycopene production is initiated.

Five E8::*SlHONG* transgenic lines, #2, #3, #5, #7 and #8, were identified by PCR using forward and reverse primers binding the E8 promoter and *SlHONG* coding sequences respectively (Figure 4.20). RT-qPCR was performed to check the mRNA levels of *SlHONG* in the successful, stable transformants. Compared to wildtype, the expression level of *SlHONG* stayed at very high levels in E8::HONG #5 after the breaker stage and remained higher than that in wildtype throughout fruit ripening (Figure 4.21 A). The expression level of *SlPSY1* was also significantly increased in this line (Figure 4.21 B), which further supported the role of *SlHONG* as a positive regulator of *PSY1*.

4.4 Discussion

Different carotenoids differentially accumulate in different tissues in tomato plants; lycopene mostly accumulates in fruit plastids, and is usually undetectable in leaves where it serves as an intermediate in the biosynthesis of xanthophylls. Tomato has evolved diverse carotenoid pathways in different tissues by recruiting different genes encoding isoforms of biosynthetic enzymes. Tomato has evolved three *PSY* genes, while Arabidopsis only has one in its genome. *PSY1* is very highly expressed in ripe-stage fruit and its natural mutant *yellow flesh* resulted in no accumulation of lycopene in the fruit. The silencing of its paralogous genes, *PSY2* and *PSY3* have much less effect on lycopene accumulation (Li et al., 2008b;

Fantini et al., 2013). Distinct from the carotenoid structural genes, *SlHONG* is expressed in both vegetative and reproductive tissues, indicating that in addition to it being a candidate positive regulator in lycopene biosynthesis during fruit ripening, it might serve different roles in other physiological processes in other tissues (pleiotropy).

As mentioned in the introduction to this chapter, no transcription factor has been found which regulates carotenoid biosynthesis in tomato fruit specifically, and all transcriptional regulators that have been identified have broad effects mainly on fruit ripening (*nor, rin* etc). In this chapter, the expression of *SlHONG* was detected in both vegetative and reproductive tissues, not just in fruit, suggesting that SlHONG likely contributes to the regulation of other physiological processes besides lycopene accumulation in fruit. This may explain why mutants regulating lycopene production have not been found in tomato, perhaps due to their pleiotropic effects and their consequent pleiotropic phenotypes.

From the data obtained, SlHONG not only serves a positive role in the transcriptional regulation in lycopene biosynthesis in tomato fruit during ripening, but likely also contributes to other physiological processes in other tissues, such as hypocotyl elongation, which is consistent its high expression level in green tissues (Figure 4.4) and the shorter hypocotyl phenotypes of *ΔSlHONG* heterozygotes and homozygous lines. Based on the phylogenetic analysis of tomato and Arabidopsis bHLH transcription factors, SlHONG is closely related, structurally to two Arabidopsis bHLH proteins, *BR-ENHANCED EXPRESSION 2* (BEE2, AT4G36540.1) and HBI1 (HBI1, AT2G18300.3), which are involved in brassinosteroid signalling (Malinovsky et al., 2014; Fan et al., 2014; Singh et al., 2017). BRs play important roles in a wide range of physiological processes, such as seed development and germination, cell division and elongation. It has been shown that defects in BR signalling result in reduction in seed size in Arabidopsis (Jiang et al., 2013a). Defects in BR metabolism often lead to shorter hypocotyl length due to interference in cell elongation. The gene encoding HBI1, one of the two proteins most similar to *SlHONG* in Arabidopsis, is preferentially expressed in hypocotyl and cotyledons, and has been shown to act as a positive regulator of cell elongation and photomorphogenesis (Bai et al., 2012). Significantly reduced hypocotyl length of *ΔSlHONG* lines was observed. This demonstrated further that

SlHONG has a positive effect on hypocotyl elongation (Figure 4.16). The abnormal phenotypes associated with cell elongation, were not only detected in the seedling stage, but were also observed as the plants reached reproductive phase. The adult plants of homozygous *ΔSlHONG* #28 T1-14 and *ΔSlHONG* #28 T1-15 were shorter compared to the plants harbouring two wild-type *SlHONG* alleles, such as *ΔSlHONG* #28 T1-19 (*SlHONG*/*SlHONG*) (Figure 4.22 B). *ΔSlHONG* #4 T1-8 and *ΔSlHONG* #4 T1-10 *ΔSlHONG* #4 T1-8 and *ΔSlHONG* #4 T1-10 were the first two lines confirmed to be homozygous for *ΔSlHONG* alleles (Figure 4.15). These two lines were smaller compared to WT and heterozygous plants under the same growth conditions (Figure 4.22 A). Particularly, *ΔSlHONG* #4 T1-8 had severe growth defects and could not achieve full height for the mature stage (Figure 4.22 A). This observation may partially explain that why no homozygous *ΔSlHONG* lines were identified in the initial screening of the T1 plants from *ΔSlHONG* #28 (Figure 4.14), because the leaf samples used for genotyping were harvested from fruiting plants, when it was likely that the homozygous plants like *ΔSlHONG* #4 T1-8 were dead already. The phenotypes caused by overexpression or knock-out of *SlHONG* were very similar to those resulting from defects in BR signalling in Arabidopsis (Bai et al., 2012). This suggested that the SlHONG may participate in responses in BR signalling in tomato, which will be discussed more extensively in the next chapter.

SlHONG is likely involved in responses to brassinosteroid signalling in tomato because lines heterozygous or homozygous for *ΔSlHONG* alleles showed similar shorter hypocotyl phenotypes to those seen in Arabidopsis mutants of BEE2. From the phenotyping of *ΔSlHONG* mutants in the T1 generation the contribution of the *HONG* gene to hypocotyl elongation may be similar to that of *BEE2* in Arabidopsis.

Figure 4.22 Adult plants of *ΔSlHONG* #4 and #28 T1 generation. The plants shown were on the same bench of those in Figure 4.15, which were about three-month old. The ruler shows centimetres.

Chapter 5

Investigation of the regulatory mechanism of SlHONG in transcription regulation of carotenoid biosynthesis in fruits

5.1 Introduction

5.1.1 Transcriptional regulation of carotenoid biosynthesis in tomato

Carotenoids serve a wide range of functions in diverse biological processes and, in tomato, regulatory networks incorporating different kinds of transcription factors maintain the production and consumption of specific carotenoids in a dynamic balance in response to internal and external signals. For example, lutein and zeaxanthin are two carotenoids operating in non-photochemical quenching and protect chlorophyll from photo-oxidation (Muller, 2001). These compounds are synthesised in substantial amounts in green tissues such as leaves and are present in green tomato fruit. Due to the functions of carotenoids as pigments, and as the precursors of lots of volatile compounds, they serve crucial roles in the bright colour, aroma and flavours of tomato fruit and flowers, attracting seed dispersers and pollinators (Ronen et al., 2000a; Vogel et al., 2010). Biosynthesis and catabolism of carotenoids, especially the lycopene biosynthetic pathway, have been studied extensively, with the identification of isogenes responsible for specific reactions, as described in Chapter 3. The composition of carotenoids in tomato fruit changes very substantially during fruit development, especially as mature green fruit ripen and turn red. The regulatory mechanisms controlling carotenoid production in green fruit are likely similar to those in leaves, but the accumulation of lycopene during fruit maturation involves a new set of transcriptional changes as described in Chapter 3.

Several studies have described the regulation of carotenoid biosynthesis at the molecular level in plants (Cunningham and Gantt, 1998; Hirschberg, 2001; Liu et al., 2004). Although the catalytic steps of the carotenoid biosynthetic pathway have been well-characterized, the regulatory mechanisms that control carotenoid accumulation remain poorly understood. Different types of regulatory mechanisms working at transcriptional and posttranscriptional levels, have been suggested to be involved in the accumulation of specific carotenoids (Lee et al., 2012; Sauret-Gueto et al., 2006). In addition to ethylene, light, the availability of substrates produced through the MEP pathway, some transcription factors have been found to contribute to the control of carotenoid levels (Chung et al., 2010;

Lee et al., 2012; Manning et al., 2006; Toledo-Ortiz et al., 2010; Vrebalov et al., 2009; Vrebalov et al., 2002). However, all of these transcription factors have broad effects on ripening and none specifically regulate a single fruit metabolic pathway. Upto now, no transcription factor has been identified directly regulating the structural genes of carotenoid biosynthesis.

5.1.2 Basic Helix-Loop-Helix (bHLH) transcription factor family in plants

Basic helix-loop-helix transcription factors are a group of proteins containing a bHLH domain, responsible for DNA binding and dimerization capabilities, which has been revealed as one of biggest transcription factor super families in plants (Pires and Dolan, 2010). bHLH proteins have been well characterised in non-plant systems, especially in mammals, as very important regulatory components in transcriptional regulation in a wide range of biological processes (Beck et al., 2014; Massari and Murre, 2000). The TFs in this super family harbour a typical bHLH domain, composed of about 60 amino acids, including two functionally important regions: 1) a basic domain, which includes about 18 hydrophilic and basic amino acids, responsible for DNA binding , and 2) a HLH structure, which is formed by two amphipathic α-helices separated by an intervening loop of variable length which is required for dimerization (Murre et al., 1994; Ferré-D'Amaré et al., 1993; Nair and Burley, 2000). bHLH TFs typically function as dimers, forming homodimers or heterodimers with closely related members in the family, while others have broader dimerization activities (Littlewood and Evan, 1998).

The first 10-18 amino acids corresponding to the basic domain, are functional in interactions with the target DNA sequences (Ferré-D'Amaré et al., 1993). Based on the specificities of the sequences of DNA binding motif and their functions, bHLH transcription factors have been subgrouped into six subfamilies (A to F) in animals (Jones, 2004). Most animal bHLH proteins recognise and bind to a hexanucleotide sequence (5'-CANNTG-3'), known as the E-box, and bound by glutamic acid at the position 9 of the bHLH domain, responsible for the interaction with the CA nucleotides within the E-box (Buck and Atchley, 2003; Ferré-

D'Amaré et al., 1993). Similarly, many plant bHLH domains have a configuration of His-Glu-Arg (H-E-R) at positions 5, 9 and 13 indicating the binding ability to E-boxes, including the motif 5'-CACGTG-3, one of the variations of the E-box, classically known as the G-box (Buck and Atchley, 2003; Heim et al., 2003). Recognition of this core binding motif is mainly determined by conserved residues in the basic region of bHLH proteins, while other amino acids in the domain play a role in binding to a specific type of E-box (Robinson et al., 2000). In some cases, binding specificity may be affected by the nucleotide sequences flanking the core recognition site (Massari and Murre, 2000; Littlewood and Evan, 1998). The α-helix loops are responsible for the formation of hetero- or homo-dimerization between bHLH transcription factors. It has been shown that the structure of a dimer is stabilized by the hydrophobic amino acids isoleucine (I), leucine (L), and valine (V) in the bHLH domain. The positions of these residues are highly conserved in animals and plants (Ferré-D'Amaré et al., 1993; Atchley et al., 1999; Pires and Dolan, 2010). The combination of E-box recognition and dimerization of bHLH proteins generates a wide range of functions in the transcriptional regulation of various physiological processes (Fairman et al., 1993). Apart from working as an activator of a transcriptional programme, non-DNA binding HLH proteins may be functional as negative regulators by forming heterodimers with other bHLH proteins (Littlewood and Evan, 1998).

High sequence similarity of bHLH proteins between plants and animals, especially of the DNA-binding basic domain and in the hydrophobic amino acids of the helical domains, suggests the conservation of the functions of members of this super transcription factor family between plants and animals. Several phylogenetic analyses have been undertaken in different plant species, revealing that most major groups of land plants have large numbers of bHLH transcription factors, particularly in *Arabidopsis thaliana* and *Oryza sativa*, with 166 and 173 bHLH transcription factors respectively (Bailey et al., 2003; Heim et al., 2003; Toledo-Ortiz, 2003; Li et al., 2006; Pires and Dolan, 2010). The bHLH transcription factors identified in Arabidopsis, rice and Chinese cabbage genomes have been classified further in to 21, 22 and 24 subfamilies respectively (Heim et al., 2003; Li et al., 2006; Song et al., 2014), only a few of which have been characterised functionally.

5.2 Materials and methods

5.2.1 Plant material and plasmids

The wild species were collected from TGRC (Tomato Genetics Resource Centre), including *S. corneliomulleri* (LA0107), *S. chmielewskii* (LA1028), *S. peruvianum* (LA1278), *S. pimpinellifolium* (LA1578), *S. harbrochaites* (L1778), *S. chilense* (LA1969), *S. neorickii* (LA2133) and *S. arcanum* (LA2172).

Golden Braid cloning and the plasmids used in the transactivation assay were kindly provided by Dr Diego Orzaez (Sarrion-Perdigones et al., 2011, 2013).

5.2.2 Transactivation assays

In eukaryotic cells, transcriptional regulation of gene expression often achieved by the binding of regulatory proteins to the specific DNA motif (*cis*-acting element) usually in the promoter of the target gene. The promoter of a gene is defined as the *cis*-acting DNA region, containing the TATA-box motif required for the recruitment of the basal transcriptional machinery, and lies upstream of the coding sequences of the target gene, and defines the transcription start site (TSS). Normally, the entire promoter region is divided into three sub-regions: the core promoter (~80-100 bp around the TSS), the proximal promoter (~250-1000 bp upstream of the core promoter) and the distal promoter which is further upstream. However, the *cis*-acting element is not always found in the defined promoter region, and can be located far more downstream or even within the gene. To test the interaction between the HONG transcription factor and the promoters of its potential target genes, I undertook transactivation assays using the Dual-Luciferase® Reporter (DLR™) Assay System Kit from Promega. In this dural-reporter system, the actificties of two luciferases are measured sequentially from the same sample. In the DLR™ assay, the firefly (*Photinus pyralis*) luciferase was used as the reporter reflecting the binding activity between transcription factor and the promoter of target gene, and the luminescent signal generated

by Renilla (*Renilla reniformis* or sea pansy) luciferase was used as a standard after the activity of firefly luciferase had been quenched.

Plasmid construction The defined promoters of genes to be tested, coding sequences of transcription factor candidates (SlHONG, SpHONG and SlRIN), as well as negative control (GFP) were amplified by PCR and cloned into the pDONR207 vector to generate entry vectors. Gateway™ cassettes were domesticated into the domestication vector (pUPD) following Golden Braid cloning protocol (Sarrion-Perdigones et al., 2011). The promoters, transcription factor candidates and GFP were introduced into the Golden Braid system by LR reactions. Transcription factor candidates as well as the CaMV 35S promoter were assembled in the destination vector pDGB2_α1 to from a complete transcription unit. Similarly, the promoter of interest with the reporter luciferase, firefly luciferase, and the TNos terminator were assembled in pDGB_α2. The resulting units were then combined in pDGBΩ1, as a 35S:TF:TNos-promoter:firefly_luciferase:TNos construct. More information about the vector sequences can be found in Golden Braid Database (https://gbcloning.upv.es/).

N. benthamiana **agroinfiltration** Plasmids constructed, as well as the pre-assembled unit 35S:Renilla:TNos-35S:p19:TNos, were transformed into *A. tumefaciens* strain GV3101:pMP90 by electroporation. Overnight-grown bacterial cultures were pelleted and resuspended in agroinfiltration medium (10 mm MES, pH 5.6, 10 mm $MgCl_2$, and 200 µm acetosyringone) to an optical density at 600 nm of 0.5. Infiltrations were carried out using a needle-free syringe in leaves of about 3-week-old *N. benthamiana* plants (growing conditions: 24°C day/20°C night in a 16-h-light/8-h-dark cycle). Leaves were harvested 3 days post infiltration.

Transactivation assays To measure the activity of firefly/ Renilla luciferase reporters (Grentzmann et al., 1998), three or four *N. benthamiana* leaves were agroinfiltrated with the *AGL1* stains containing the test constructs following the agroinfiltration of *N. benthamiana* leaves as described above. Leaves were harvested three days after infiltration.

Firefly luciferase and Renilla luciferase were assayed in the leaf extracts following the DLR™ Assay System (Promega) standard protocol and were quantified by a GloMax 96 Microplate Luminometer (Promega).

5.2.3 Construction of the phylogenetic trees for bHLH proteins

The phylogenetic analysis of bHLH proteins was conducted with considerable help from Dr Paul Bailey. HMMSEARCH from the HMMER3 software suite (version 3.1b2; Eddy, 2011) was used to search for all bHLH transcription factors from the proteomes of eight species: *A. thaliana* (Cheng et al., 2017), *O. sativa* (Kawahara et al., 2013), *S. lycopersicum* (Tomato Genome Consortium, 2012), *S. pennellii* (Bolger et al., 2014), *S. pimpinellifolium* (Tomato Genome Consortium, 2012), *S. tuberosum* (Sharma et al., 2013), *C. annuum* CM-334 (Kim et al., 2014) and *M. polymorpha* (Bowman et al., 2017). A hidden Markov model (HMM) of the bHLH domain was built from a wide range of known Arabidopsis and rice bHLH sequences and used for the search. The sequences identified were aligned to the HMM using HMMER3 HMMALIGN. The resulting alignment of the bHLH domain was converted to FASTA format and gap columns not part of the HMM were removed. Sequences with < 70% coverage across the alignment were removed from the dataset and the longest sequence for each gene out of the available set of splice versions was used for phylogenetic analysis. Phylogenetic analysis was carried out using the MPI version of RAxML (v8.2.9; Stamatakis, 2014) with the following method parameters set: -f a, -x 12345, -p 12345, -# 100, -m PROTCATJTT. The tree was mid-point rooted and visualized using the Interactive Tree of Life (iToL) tool (Letunic and Bork, 2016).

5.2.4 Yeast transformation

The yeast transformation protocol was modified from Gietz and Woods, 2002. A fresh yeast colony from the plate was incubated in 5 mL of YPD media overnight at 30 °C shaking at 220rpm. A sample of overnight culture, with OD_{600} of 0.5, was inoculated into 25 mL YPD medium and grown at 200 rpm, 30 °C for about 4 hours until the OD_{600} reached 1. The liquid

culture was then transferred to a sterile 50 mL falcon tubes, yeast cells were harvested by centrifuging at 1750 g (high speed in clinical centrifuge) for 2 min. The supernatant was carefully removed, and the pellet was resuspended in 1 mL of sterile nano-pure H_2O. The wash step was repeated carefully. The pellet was resuspended in 1 mL of 100 Mm lithium acetate (LiAc) and incubated at 30 °C for 5 min. The cells were gently resuspended and 100 µL was transferred into a 1.5 mL sterile Eppendorf tube for the transformation reaction. The following solutions were added sequentially: 240 µL of PEG (50% w/v), 36 µL of 1.0M LiAc, 5 µL of prepared salmon sperm DNA (10 mg/ml), 100 ng of each plasmid DNA and 16 µL of sterile nano-pure dH_2O. The mixture was vortexed for at least 1 min followed by incubation at 42°C for 15 min. The cells were centrifuged at top speed for 10 seconds and the supernatant was removed using a pipet. The pellet was gently resuspended in 200 uL of sterile nano-pure dH_2O by pipetting. 100 uL of the cell suspension was plated onto SD/-Leu-Trp plate to select the colonies with successful transformation of plasmids and onto SD/-Leu-Trp-His-Ade plate to check the protein-protein interactions. The plates were incubated at 30 °C.

5.2.5 Yeast two-hybrid assays

The MATCHMAKER GAL4 Two-Hybrid System (Clontech) was applied. Yeast strain AH109 was co-transformed with pGADT7-ECD2 and pGBKT7 fused to the transcription factor candidate to be tested using the yeast transformation protocol described above. The transformants were spotted on SD medium lacking Trp/Leu or SD medium lacking Trp/Leu/His/Ade and examined for growth. Each tested transcription factor candidate was co-transformed with an empty vector (pGADT7-ECD2 or pGBKT7) to check the autoactivation. Interaction activities were scored visually based on the growth of the colonies.

5.3 Results

5.3.1 SlHONG regulates the expression of carotenoid biosynthetic genes through binding directly to their promoters

Nearly all the genes encoding enzymes in the lycopene biosynthetic pathway were significantly down-regulated when *SlHONG* was silenced using VIGS (Figure 4.5), especially *SlPSY1*, encoding the reportedly rate-limiting step of the pathway, which was 17 times less highly expressed compared to control, unsilenced tomato fruit tissue (Figure 4.8). Stable overexpression of *SlHONG* under the control of the fruit-specific, E8 promoter resulted in increased transcript levels of *SlPSY1*, the gene encoding the rate-limiting step in lycopene biosynthesis during fruit ripening (Figure 4.21). The association between the expression of *SlHONG* and the structural genes indicated that *SlHONG* might act as a positive regulator of lycopene biosynthesis. However, up to this point I had no evidence showing whether SlHONG activates the expression of the structural genes through direct binding or whether it contributes to the transcriptional regulation of the lycopene biosynthetic pathway indirectly, through other regulators.

SlHONG is a basic helix-loop-helix (bHLH) transcription factor belonging to a superfamily of bHLH transcription factors that bind to specific DNA motifs. Extensive phylogenetic analyses of bHLH transcription factors has been undertaken in Arabidopsis and rice, based on plant bHLH proteins which can be classified into 15-25 subgroups (Heim et al., 2003; Li et al., 2006; Toledo-Ortiz, 2003; Buck and Atchley, 2003), that participate in a wide range of biological processes. In a previous genome-wide study in tomato, 152 putative bHLH transcription factors were identified through a Hidden Markov Model, of which 20 were grouped into subfamily XII based on the sequence of N-terminal region of the bHLH domain (Wang et al., 2015). An alignment based on the protein sequences of these 20 type XII bHLH transcription factors, revealed that there was very low level of similarity outside the bHLH domain (Wang et al., 2015). SlHONG groups as one member of bHLH subfamily XII, having a 48-amino acid bHLH domain (Wang et al., 2015), in which the basic domain is believed to be responsible

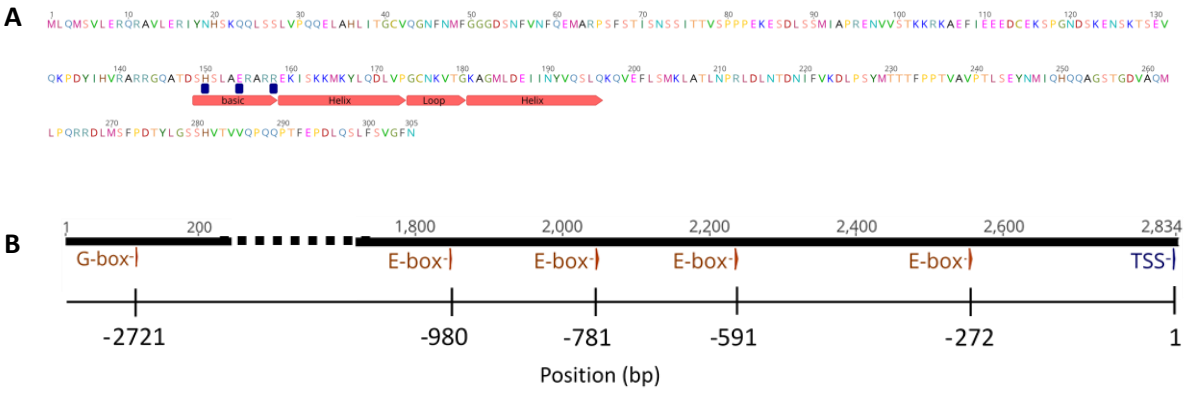


Figure 5.1 Representative information for SlHONG and the *SlPSY1* promoter. A, protein sequence of SlHONG, in which the bHLH domain is indicated by a red arrow. B, distribution of E-boxes (including the G-box) within the promoter region of *SlPSY1*. TSS, transcription start site; E-box, CANNTG; G-box, CAGCTG.

for binding DNA. This domain includes the key His-Glu-Arg (H-E-R) residues, which have been proposed to play a crucial role in binding to the E-box (Heim et al., 2003) and were present in the basic domain of SlHONG (Figure 5.1 A), consistent with other members in subfamily XII, except SlbHLH 047 and SlbHLH 055, which were identified as non-DNA binding, HLH proteins.

Five E-box motifs (including one G-box), the cognate binding motif of bHLH transcription factors, were identified in the genomic region upstream of the *SlPSY1* coding sequences, located between -3000 bp upstream and the transcription start site (TSS). One G-box (CAGCTG), was present at the distal end (-2721) of the analysed region (Figure 5.1 B). The presence of these motifs suggested that SlHONG might function as a transcriptional activator of *SlPSY1* through direct interaction with the *SlPSY1* promoter.

To characterise further the underlying mechanism linking SlHONG to transcriptional activation of PSY1, a Dual-Luciferase Reporter (DLR) Assay (Promega) was undertaken to test whether SlHONG could activate transcription from the *SlPSY1* promoter. The dual-reporter assay utilises firefly luciferase in combination with renilla luciferase within a transient leaf expression system to make ratiometric measurements. Renilla luciferase is used as an internal control to which measurements of firefly luciferase are normalised. In measurements of PSY1 promoter activity, this system was used in transient transfections of tobacco leaves, where the vector containing the experimental reporter gene, encoding firefly luciferase, was coupled to the promoter of *SlPSY1* together with *SlHONG* expressed under the control of the CaMV 35S promoter (Figure 5.2 A). This vector was co-transfected with a second vector containing the renilla luciferase gene that served as an internal control for reporter activity. The renilla luciferase gene was driven by a constitutive promoter, CaMV 35S promoter. Use of this dual luciferase assay helped minimise the variation in reporter gene activity under different experimental conditions.

Cloning and characterisation of the *SlPSY1* promoter region from S.*lycopersicum*, including its 5' UTR, confirmed that several potential E-boxes were present, which might contribute

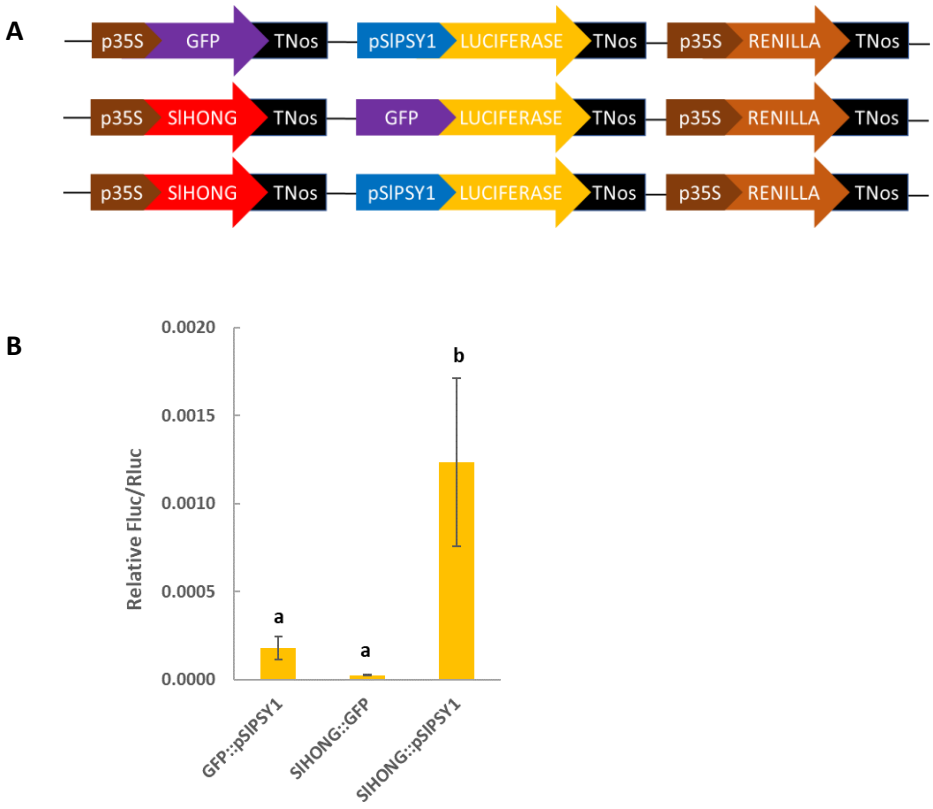


Figure 5.2 Transactivation assays to test the transcriptional activation between SlHONG and the *SlPSY1* promoter in by agroinfiltration in *N. benthamiana* leaves. A, Scheme of multigene constructs used. In each construct, full length coding sequence of tested transcription factor (*SlHONG*), or control, was driven by CaMV 35S promoter. The control or the promoter of the tested gene (*SlPSY1*) was fused with *firefly luciferase* gene coding sequence. The reference part was composed of *renilla luciferase* gene under the control of the CaMV 35S promoter. GFP was used as the control. B, Transactivation assay in *N. benthamiana* leaves. *x* axis, the tested combinations of transcription factor (SlHONG) and the promoter region of the *SlPSY1*. *y* axis, relative firefly luciferase (Fluc) activity normalised by renilla luciferase activity as reference. Error bars, s.e.m.; ten biological replicates. Statistical analysis was performed using a two-tailed *t* test and different letter indicated the significant difference, $p < 0.05$.

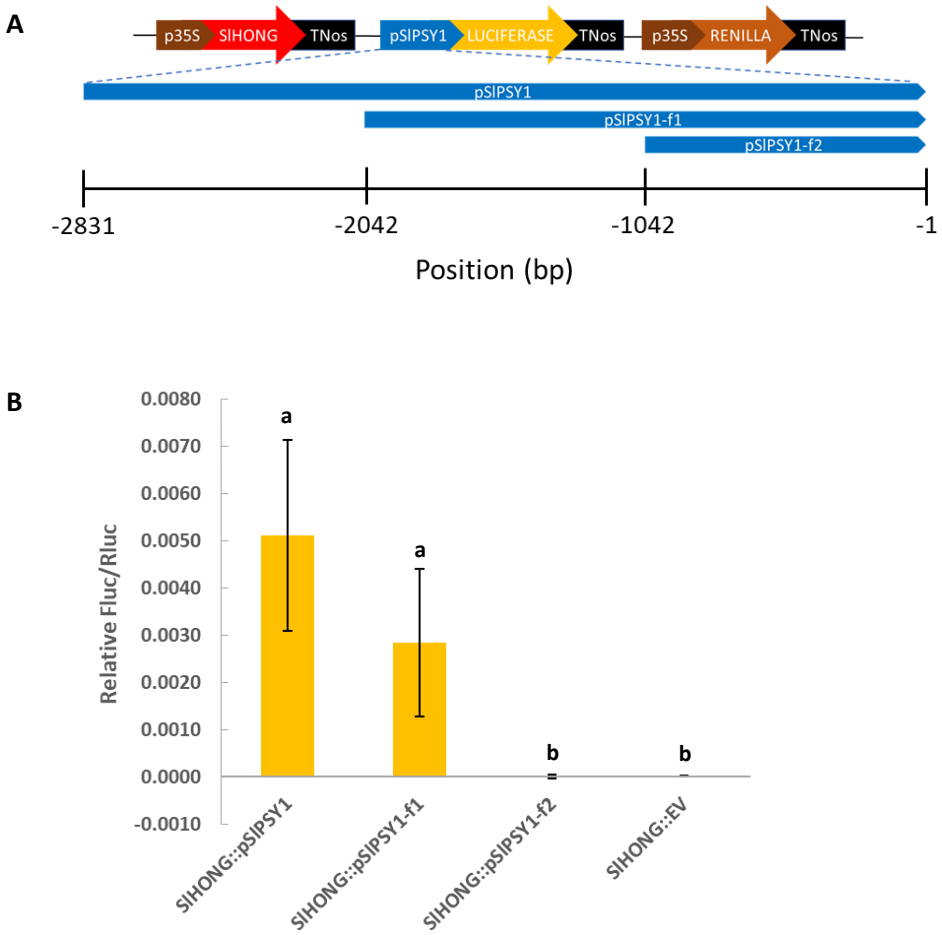


Figure 5.3 Transactivation assays to test transcriptional activitivation between SlHONG and truncated *SlPSY1* promoters by agroinfiltration in *N. benthamiana* leaves. A, Scheme of multigene constructs used. Full length coding sequence of *SlHONG* was driven by the CaMV 35S promoter. Full length or truncated *SlPSY1* promoters were fused to *firefly luciferase* gene. The tested *SlPSY1* promoter regions were indicated by the position according to TSS (transcription start site, +1). B, Transactivation assay in *N. benthamiana* leaves. Full length *SlPSY1* promoter (pSlPSY1) and the fragment, pSlPSY1-f1 (-1 to -2042), could be activated by SlHONG. *x* axis, the tested combinations of transcription factor (SlHONG) and the promoter regions of the *SlPSY1*. *y* axis, relative firefly luciferase (Fluc) activity normalised by renilla luciferase activity as reference. Error bars, s.e.m.; five biological replicates. Statistical analysis was performed using a two-tailed *t* test, and different letter indicated the significant difference, $p < 0.05$.

to the activation of this gene by SlHONG. In the dual luciferase assay, overexpression of SlHONG, significantly increased the signal from firefly luciferase when it was driven by the full-length *SlPSY1* promoter, compared to controls, in which either the *SlPSY1* promoter or *SlHONG* CDS were replaced by GFP (Figure 5.2). This suggested that SlHONG activated the expression of *SlPSY1* through direct binding to its promoter. To narrow down the region contributing to the direct interaction with SlHONG further, the *SlPSY1* promoter was dissected into shorter fragments as shown in Figure 5.3 A, which were then cloned separately into the DLR constructs driving firefly luciferase to test the corresponding response to the expression of SlHONG. As previously observed, the full-length *SlPSY1* promoter was activated by SlHONG. A similar signal was detected when firefly luciferase was under the control of first fragment of *SlPSY1* promoter (pSlPSY1-f1, -1 to -2042), whereas no activation was observed for fragments that were truncated further (pSlPSY1-f2, -1 to -1042) (Figure 5.3 B), indicating that the region of pSlPSY1-f1 that did not overlap with pSlPSY1-f2 must contain the motif(s) responsible for interaction with SlHONG. Because no E-boxes were found in the non-overlapping region between pSlPSY1-f1 and pSlPSY1-f2 (-1042 to -2042), other experimental methods are needed to define the exact binding motif of *SlHONG*, such as yeast one-hybrid assays. Candidate binding motifs, identified by these functional assays could then be confirmed by chromatin immunoprecipitation (ChIP) provided an antibody to SlHONG is available or that epitope-tagged versions of SlHONG are developed in tomato.

The Dual-Luciferase Reporter Assay was performed to test whether SlHONG could activate the promoter regions of any other carotenoid biosynthetic genes, including *SlDXS1*, *SlPDS*, *SlZISO*, *SlCRTISO* and *SlCYCB*, while GFP was used as the negative control (Figure 5.4 A). In addition to the *SlPSY1* promoter, which could be activated by SlHONG in the DLR assay, the promoter region of *SlDXS1* was also able to initiate the transcription of firefly luciferase in response to the expression of SlHONG (Figure 5.4 B). A weaker signal was detected for the combination of SlHONG and the *SlPDS* promoter (Figure 5.4 B). SlHONG could not self-activate its own expression (data not shown). The induction of firefly luciferase activity was observed for other promoters of lycopene biosynthetic genes. However, the induction of firefly luciferase in these cases was much smaller and none of them showed significantly higher induction compared to controls. More repeats of these experiments are needed to

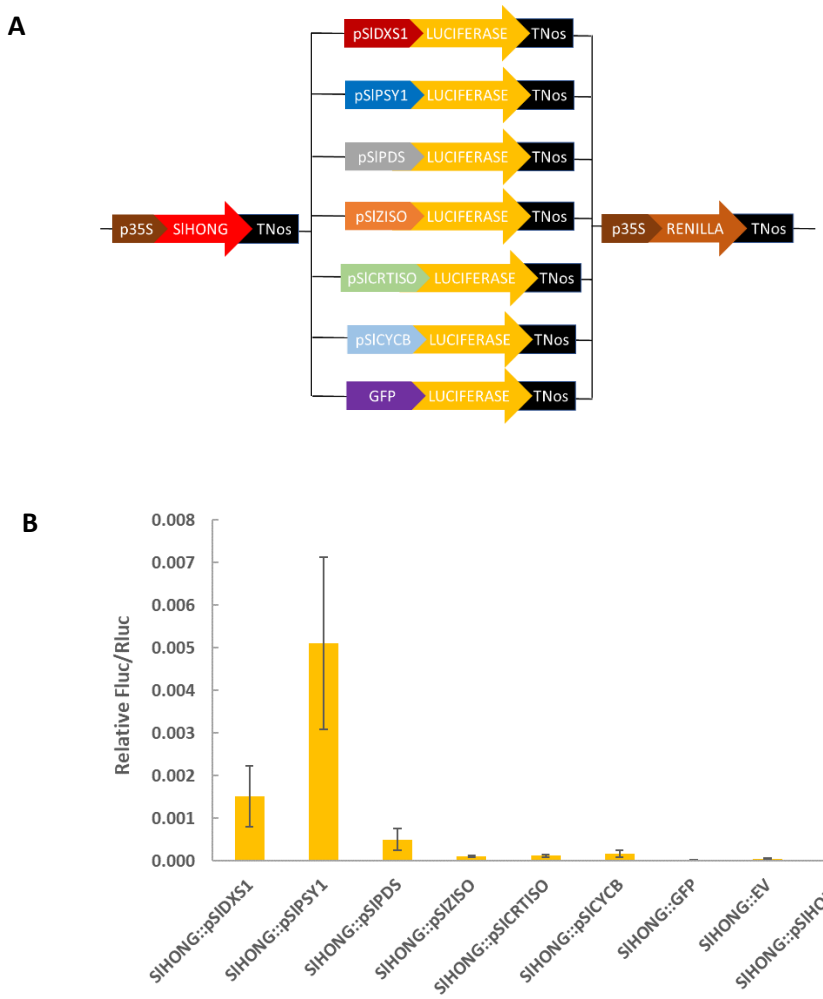


Figure 5.4 Transactivation assays to test binding activity between SlHONG and promoters of lycopene biosynthetic genes by agroinfiltration in *N. benthamiana* leaves. A, Scheme of multigene constructs used. In each construct, full length coding sequence of tested transcription factor (SlHONG), or control, was driven by CaMV 35S promoter. The control or the promoters of the structural genes were fused with *firefly luciferase* gene coding sequence. The reference part was composed of *renilla luciferase* gene under the control of the CaMV 35S promoter. GFP was used as the control. B, Transactivation assay in *N. benthamiana* leaves. *x* axis, the tested combinations of transcription factor (SlHONG) and the promoter region of the carotenoid biosynthetic genes. *y* axis, relative firefly luciferase (Fluc) activity normalised by renilla luciferase activity as reference.

complete these analyses. *SlZDS*, one of the structural genes in the lycopene biosynthetic pathway, was not included in this assay, due to a lack of success in cloning of its promoter region into the DLR vector. The promoter of *SlZDS* needs to be tested for activation by HONG in the future.

IL2-1 contains a *cis*-eQTL for *SlHONG* (Chapter 3, Figure 3.11). In this IL, *SlHONG* has been replaced by its ortholog from *S.pennellii*, *SpHONG*. The transcript abundance of *SpHONG* is less than 5% of that in fruit of other ILs in the M82 background or in M82 itself (Chapter 3, Figure 3.11). Correspondingly, most of the lycopene biosynthetic genes are down-regulated in ripe fruit of IL2-1 compared to fruit of M82. This was the basis on which IL2-1 was identified as carrying a *trans*-eQTL for lycopene biosynthesis. The differences in transcript levels for lycopene biosynthetic genes could be because *S. pennellii* HONG (SpHONG) is non-functional as a transcription factor or because new factors regulate the expression of *SlHONG*, but not *SpHONG* in ripe tomato fruit. When SpHONG was tested for its ability to drive expression of firefly luciferase from the *PSY1* promoter in DLR assays, similar activation of firefly luciferase was detected with the expression of SpHONG as with SlHONG. The luciferase signal was significantly higher than controls after being normalised to renilla (Figure 5.5). The predicted protein sequences of SlHONG and SpHONG were aligned and revealed that the key His-Glu-Arg (H-E-R) residues were present in the bHLH domain of SpHONG and were 100% identical to the residues in SlHONG (Figure 5.6), suggesting that SpHONG has the same transcriptional activiation as SlHONG, so explaining the activation of the *SlPSY1* promoter by SpHONG observed in Figure 5.5.

SpHONG could also activate the expression of firefly luciferase when it was driven by other promoter regions from the genes encoding enzymes of lycopene biosynthesis (Figure 5.7 A). Direct interaction between SpHONG and *SlDXS1* and *SlPDS* promoters was observed, consistent with the data for SlHONG. In addition, firefly luciferase induction was also detected from the *SlCRTISO* and *SlCYCB* promoters separately in response to SpHONG (Figure 5.7 B). Because the DLR assay is based on transient expression by Agroinfiltration, the variation between different experiments needs to be taken into consideration, meaning that it is very difficult to compare the data obtained from different experiments, which

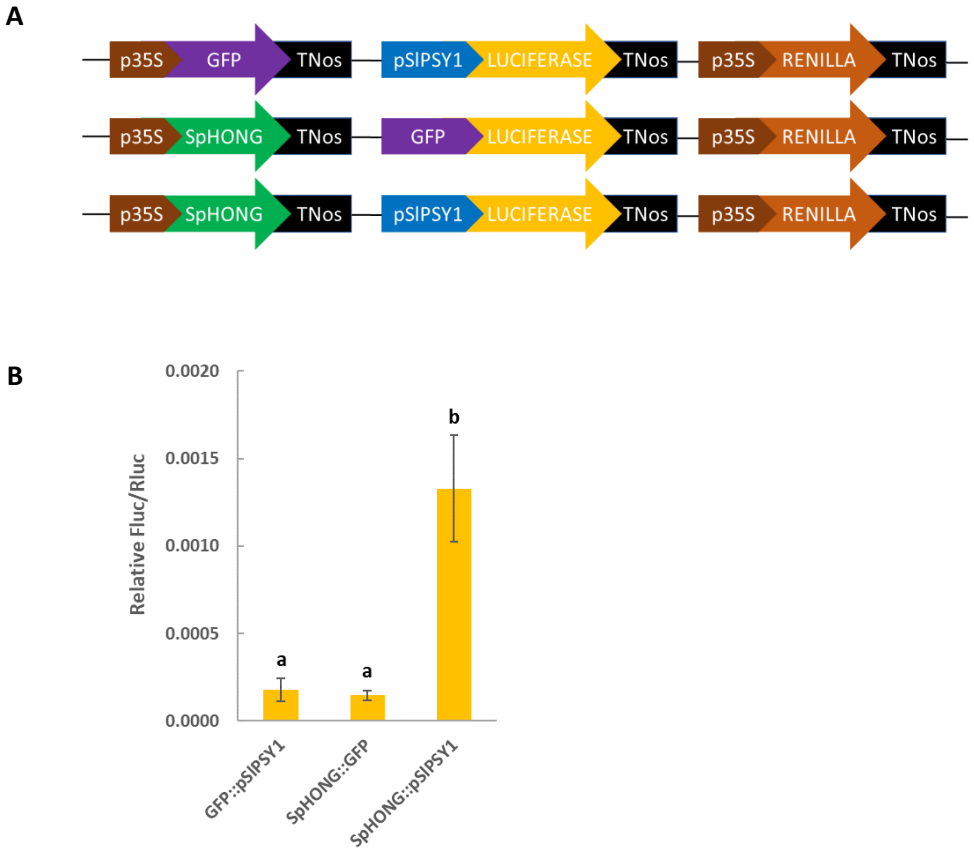


Figure 5.5 Transactivation assays to test binding activity between SpHONG and *SlPSY1* promoter in by agroinfiltration in *N. benthamiana* leaves. A, Scheme of multigene constructs used. In each construct, full length coding sequence of tested transcription factor (*SpHONG*), or control, was driven by CaMV 35S promoter. The control or the promoter of the tested gene (*SlPSY1*) was fused with *firefly luciferase* gene coding sequence. The reference part was composed of *renilla luciferase* gene under the control of the CaMV 35S promoter. GFP was used as the control. B, Transactivation assays in *N. benthamiana* leaves. *x* axis, the tested combinations of transcription factor (*SpHONG*) and the promoter region of the *SlPSY1*. *y* axis, relative firefly luciferase (Fluc) activity normalised by renilla luciferase activity as reference. Error bars, s.e.m.; ten biological replicates. Statistical analysis was performed using a two-tailed *t* test and different letter indicated the significant difference, $p < 0.05$.

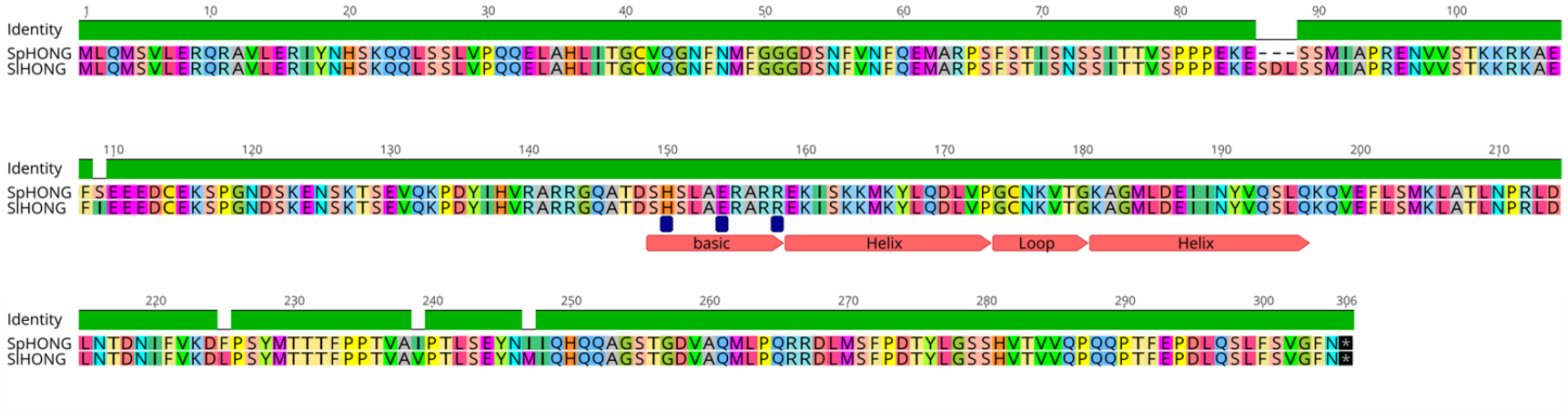


Figure 5.6 Alignment of protein sequences between SlHONG and SpHONG. The basic helix-loop-helix domain is indicated by red arrows.

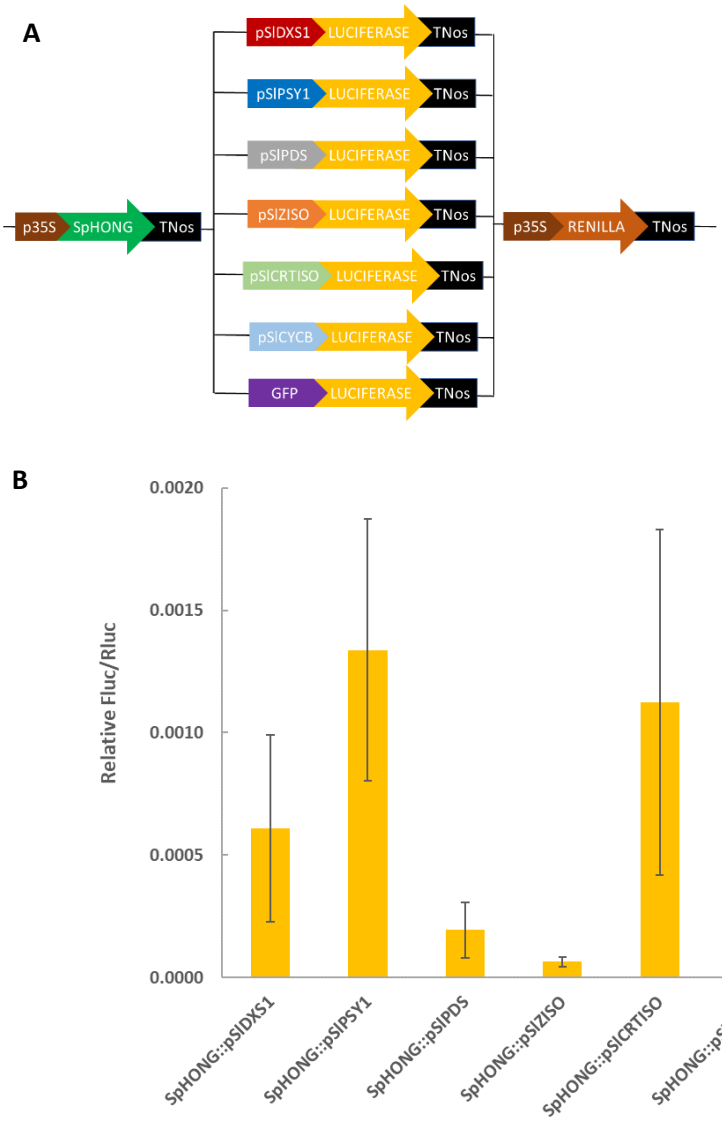


Figure 5.7 Transactivation assays to test the binding activity between SpHONG and the promoters of lycopene biosynthetic genes by agroinfiltration in *N. benthamiana* leaves. A, Scheme of multigene constructs used. In each construct, full length coding sequence of tested transcription factor (SpHONG), or control, was driven by CaMV 35S promoter. The control or the promoters of the structural genes were fused with *firefly luciferase* gene coding sequence. The reference part was composed of *renilla luciferase* gene under the control of the CaMV 35S promoter. GFP was used as the control. B, Transactivation assay in *N. benthamiana* leaves. *x* axis, the tested combinations of transcription factor (SpHONG) and the promoter region of the carotenoid biosynthetic genes. *y* axis, relative firefly luciferase (Fluc) activity normalised by renilla luciferase activity as reference. Error bars, s.e.m.; five biological replicates.

explains the large error bars in the figures. To further confirm the direct interactions suggested, these DLR assays need to be repeated with more replicates in each experiment. Given that SpHONG appears to be functionally identical to SlHONG, the differences in activity of HONG in fruit of M82, IL2-1 and *S. pennellii* may be due to differences in the level of expression of the *HONG* gene in ripe fruit. This was confirmed by the fact that IL2-1 carries a strong *cis*-eQTL for *HONG*, with much lower expression of the *S. pennellii* allele than the *S. lycopersicum* allele in ripe fruit (Chapter 3, Figure 3.11). Consequently, it is likely that other regulatory factors contribute to the regulation of carotenoid accumulation in tomato fruit, either acting in parallel with HONG or controlling the expression of HONG during fruit ripening.

5.3.2 *SlHONG* is regulated by the MADS-domain transcription factor, RIN, through a CArG motif in the promoter region of *SlHONG*

RIN, a MADS-box transcription factor, directly regulates fruit ripening genes, and demethylation occurs near RIN binding sites during fruit ripening (Zhong et al., 2013). It has been reported that the RIN protein and mRNA are absent during the initial pre-ripening phase of pericarp development (0–35 DPA), but accumulate early during ripening. RIN protein is detected slightly before the breaker stage, and its expression is maintained throughout ripening (up to 10 days after breaker), which correlates with the expression profile of carotenoid biosynthetic genes (Figure 5.8). On the basis of ChIP-seq with an antibody to RIN (Zhong et al., 2013), *SlPSY1* was identified as a potential direct target of RIN, which was confirmed by the DLR assay (Figure 5.9). However, *SlPSY1* is the only gene in the carotenoid biosynthetic pathway which can be direly activated by RIN. For example, *SlPDS*, encodes the protein catalysing the second step of the pathway, converting phytoene into ζ-carotene. The *SlPDS* promoter showed interaction with HONG, but could not be activated by RIN in the DLR assay (data not shown). Therefore, the RNA expression profiles of RIN cannot explain fully the changes in expression of structural genes contributing to carotenoid accumulation during tomato fruit ripening.

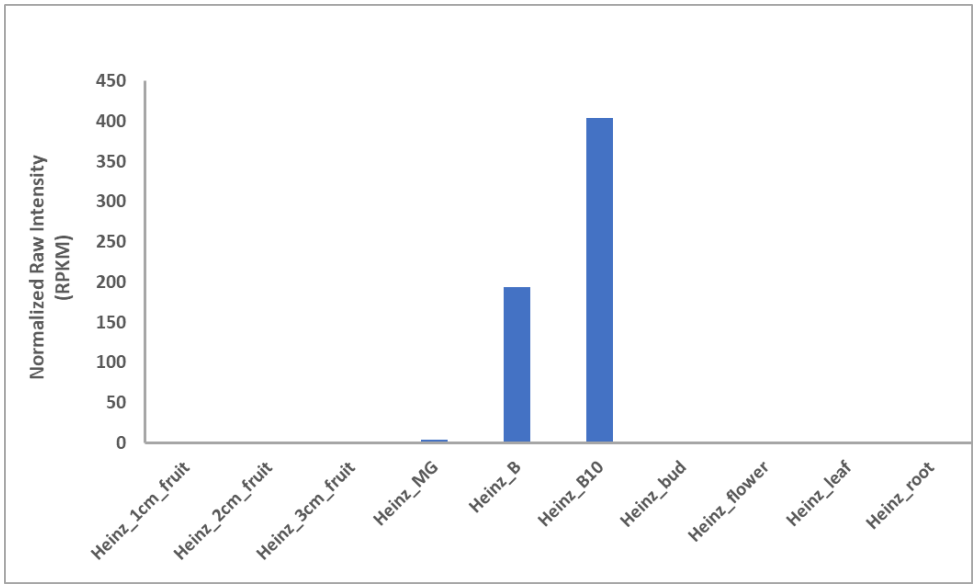


Figure 5.8 Normalized expression levels of *SlRIN* across the tomato plant. The data was from Tomato Genome Consortium, 2012. MG, Mature Green. B, Breaker stage. B10, 10 days after breaker.

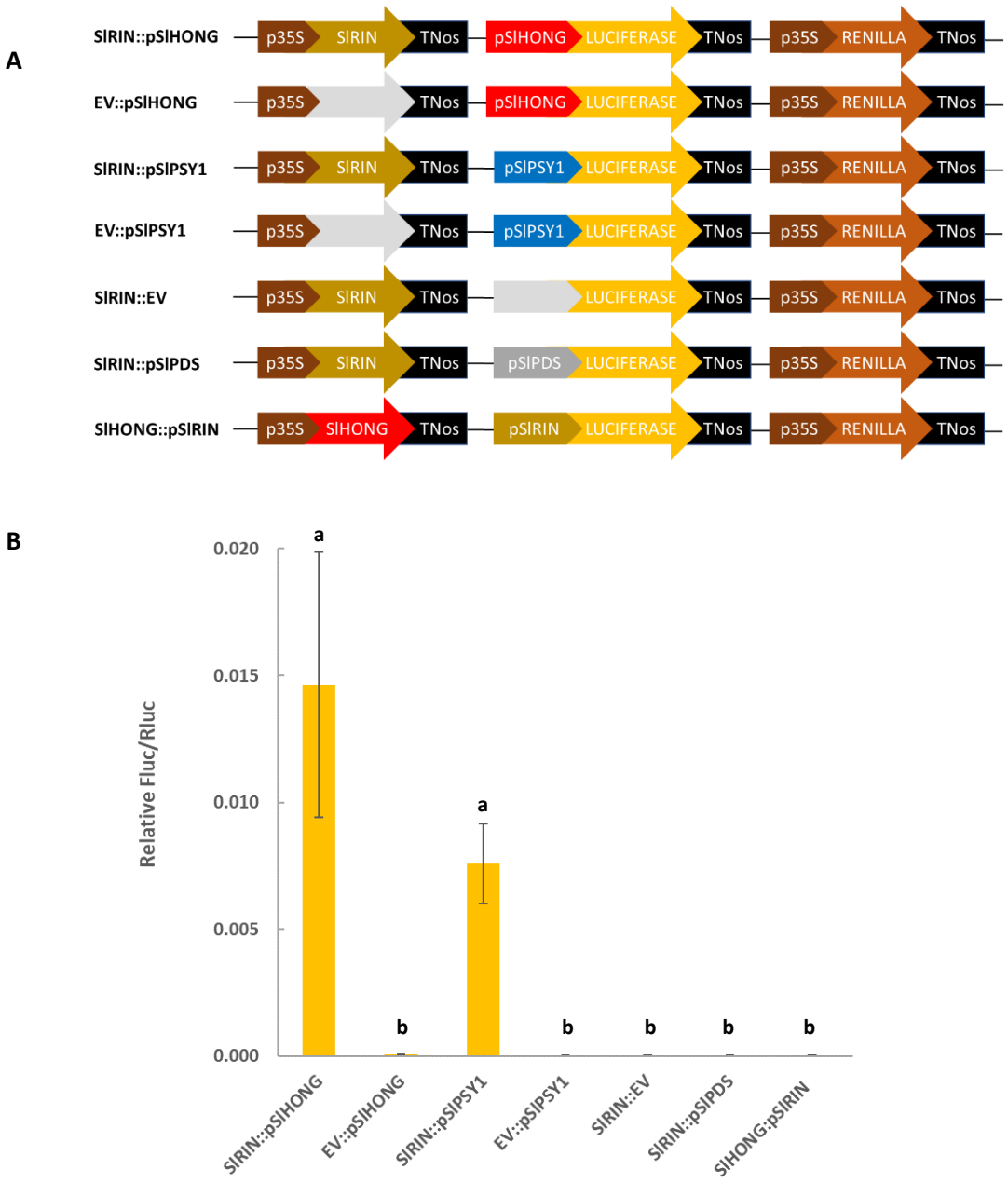


Figure 5.9 Transactivation assays to test binding activity between SlRIN and the promoters of selected genes by agroinfiltration in *N. benthamiana* leaves. A, Scheme of multigene constructs used. In each construct, full length coding sequence of tested transcription factor (SlRIN or SlHONG), or control, was driven by CaMV 35S promoter. The control or the promoters of the tested genes were fused with *firefly luciferase* gene coding sequence. The reference part was composed of *renilla luciferase* gene under the control of the CaMV 35S promoter. Empty vector (EV) was used as the control. B, Transactivation assay in *N. benthamiana* leaves. *x* axis, the tested combinations of transcription factor and the promoter region of the tested genes. *y* axis, relative firefly luciferase (Fluc) activity normalised by renilla luciferase activity as reference. Error bars, s.e.m.; five biological replicates. Statistical analysis was performed using a two-tailed *t* test and different letter indicated the significant difference, $p < 0.05$.

SlHONG was identified as a potential target of SlRIN by examination of existing ChIP-seq data generated with antibodies to RIN, which was confirmed with help from the lab of Dr. Silin Zhong. Enrichment of sequences in the *HONG* promoter upstream of the *SlHONG* TSS was observed in chromatin immunoprecipitated with the RIN antibody (Figure 5.10). Previous studies have reported the ability of RIN to bind to the CArG motif (the consensus MADS-box motif) in vitro and that a number of promoters bound by RIN are enriched in CArG motifs (Ito et al., 2008; Fujisawa et al., 2011). Two putative CArG motifs, defined as the consensus sequence C(C/T) (A/T)6(A/G) G, were identified in the promoter region of *SlHONG*, proximal to the transcription start site (around -443 nucleotides, upstream of the translation start site) (Figure 5.12). Interestingly, these two 10-nucleotide motifs, CCTATAATAG and CCTATTATAG, on the forward and reverse strands respectively, overlap with each other over nine nucleotides, making this short region a very strong candidate as a SlRIN binding site (Figure 5.12). The ability of SlRIN to activate transcription from the *SlHONG* promoter was confirmed by a DLR assay, which showed significantly induced firefly luciferase activity resulting from the activity of RIN on the *SlHONG* promoter. This response by the *SlHONG* promoter was even stronger than the activation of the *SlPSY1* promoter by RIN (Figure 5.9).

To confirm that RIN binding was associated with the CArG motif, the *SlHONG* promoter was truncated to a smaller region upstream of the transcription start site (pSlHONG-f1, -1 to -1060), and used to drive the firefly luciferase gene, with 35S:RIN in DLR assays (Figure 5.11 A). Even though the truncated promoter (pSlHONG-f1) was only about half the length of the full-length promoter (pSlHONG, -1 to -1960), the association of SlRIN with SlHONG was maintained and RIN activated firefly luciferase activity from the truncated *SlHONG* promoter, most likely because of the presence of the two CArG motifs in the short *SlHONG* promoter (Figure 5.11 B).

It had been observed that expression of *SlHONG* was positively associated with the transcript levels of lycopene biosynthetic genes, as well as the direct binding between SlHONG and the promoters of some structrue genes, especially *SlPSY1*, encoding the proposed rate-limiting enzyme. Very similar binding activity by SlHONG and SpHONG was

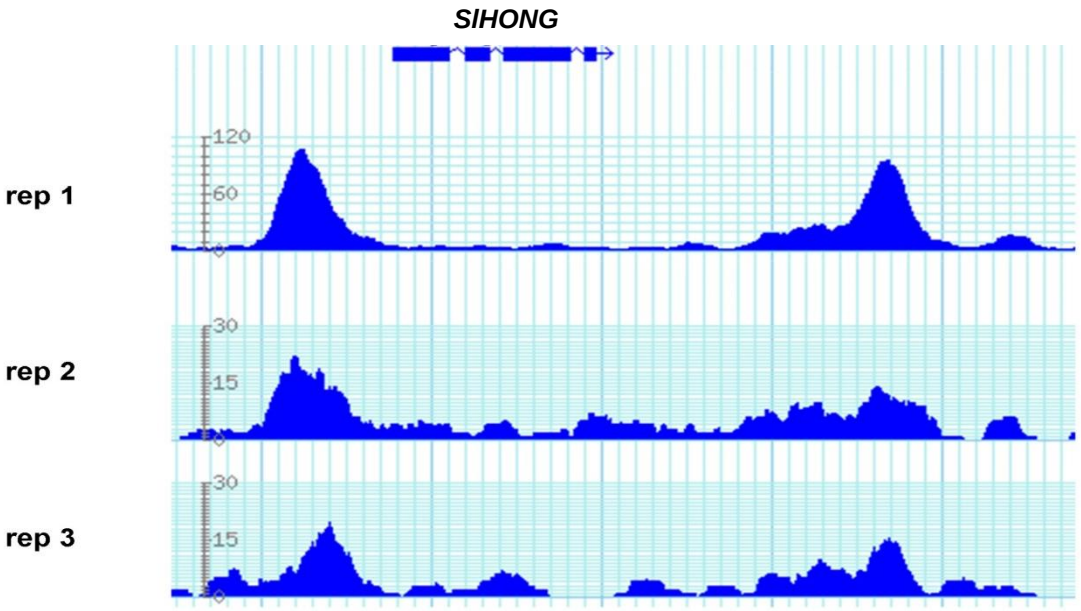


Figure 5.10 Enrichment of sequences in the *SlHONG* promoter upstream of *SlHONG* TSS was observed in chromatin immunoprecipitated with SlRIN antibody, Silin Zhong, *et al*., 2013, *Nature Biotechnology*. The number of reads is shown on the *y* axis. For the region of the *SlHONG* promoter precipitated by the SlRIN antibody in ChIP-seq. The position of the *SlHONG* coding sequence is shown to scale. Three biological replicates are shown (rep 1, rep 2 and rep 3). The full data can be accessed from the tomato epigenome database (http://ted.bti.cornell.edu/epigenome/).

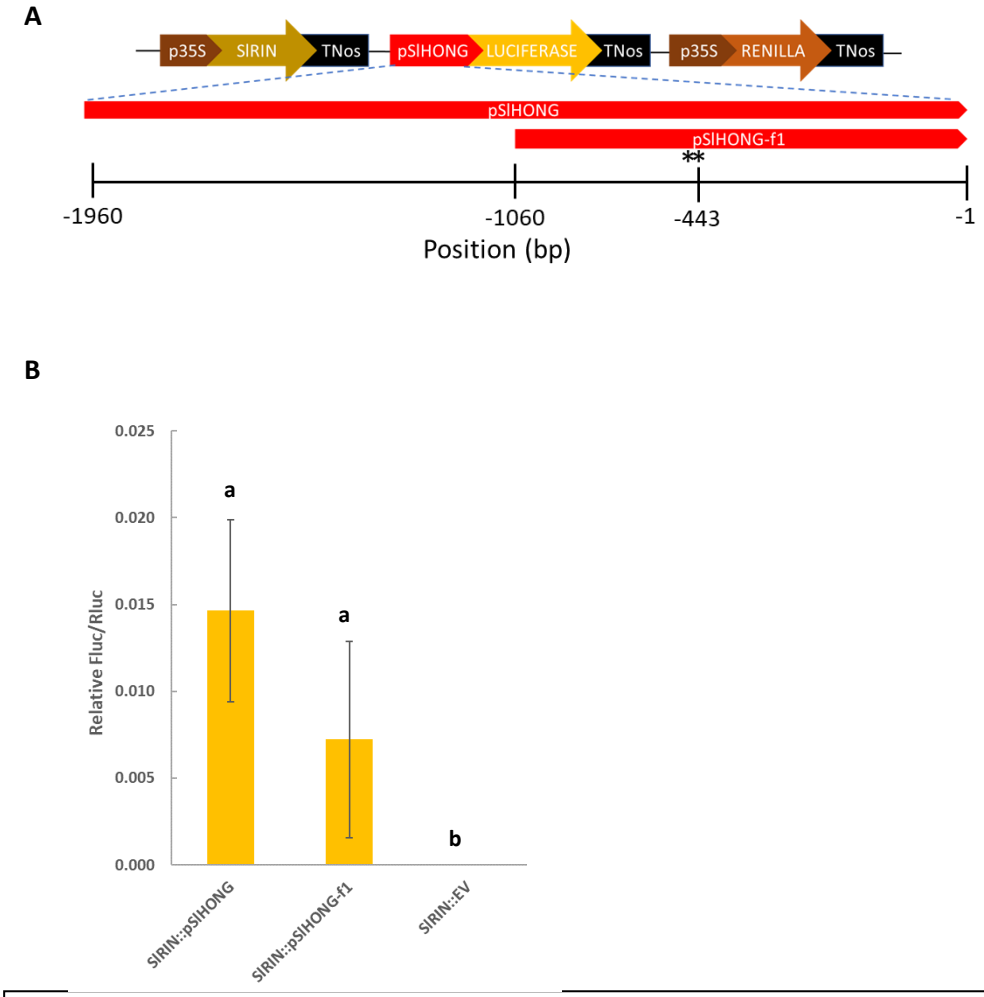


Figure 5.11 Transactivation assays to test the transcriptional activation by SlRIN on truncated *SlHONG* promoters by agroinfiltration in *N. benthamiana* leaves. A, Scheme of multigene constructs used. Full length coding sequence of *SlRIN* was driven by the CaMV 35S promoter. Full length or truncated *SlHONG* promoters were fused to *firefly luciferase* gene. The tested *SlHONG* promoter regions were indicated by the position according to TSS (transcription start site, +1). Asterisks show position of CArG boxes. B, Transactivation assay in *N. benthamiana* leaves. Full length *SlHONG* promoter (pSlHONG) and the fragment, pSlPSY1-f1 (-1 to -1060), could be activated by SlRIN. *x* axis, the tested combinations of SlRIN and the promoter regions of the *SlHONG*. *y* axis, relative firefly luciferase (Fluc) activity normalised by renilla luciferase activity as reference. Error bars, s.e.m.; five biological replicates. Statistical analysis was performed using a two-tailed *t* test, and different letter indicated the significant difference, $p < 0.05$.

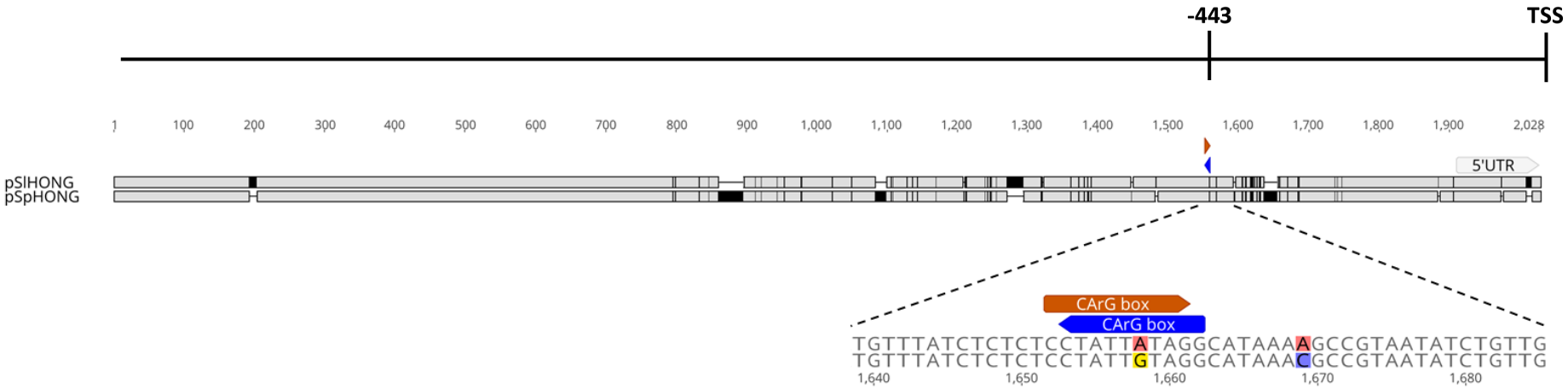


Figure 5.12 Alignment of promoter regions of *SlHONG* and *SpHONG*. One A to G SNP was detected in the *SpHONG* promoter, which abolished two CArG motifs. Black bars indicate SNPs. TSS, transcription start site.

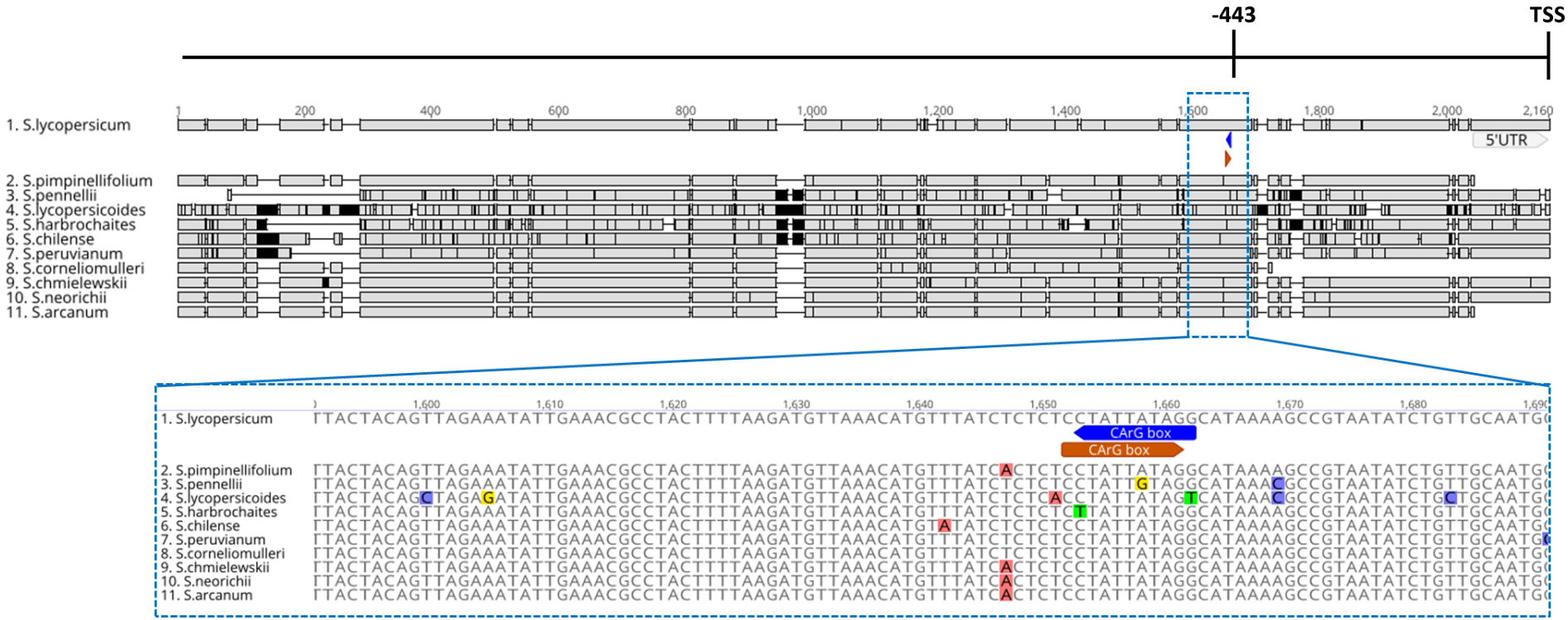


Figure 5.13 Alignment of promoter regions of HONG of *S. lycopercisum* and its wild relatives. Black bars indicate SNPs. TSS, transcription start site.

detected, because of the identical bHLH domains contained in the two proteins. The mRNA level of *HONG* in IL2-1 was more than 90% reduced, when *SlHONG* was replaced by *SpHONG* in the introgressed region of IL2-1, and there was corresponding down regulation of the transcripts of most of the lycopene biosynthetic genes. Based on the differential expression of *SlHONG* and *SpHONG* in fruit ripening and that *SlHONG* was identified as a direct target of SlRIN (Figure 5.9), it was necessary to check the presence of the CArG motifs in *SpHONG*, which likely contributed substantially to the activation of the transcription of *SlHONG* by SlRIN. However, both overlapping CArG motifs, are abolished by an A to G SNP in the promoter of *SpHONG*, together with the absence of other C(C/T) (A/T)6(A/G) G sites lying within 2kb upstream of the transcription start site of *SpHONG* (Figure 5.12). This suggests that *SpHONG* might not be activated by RIN, which could explain the very low transcript level of *SpHONG* in IL2-1. DLR assays need to be undertaken to confirm this conclusion. *SlHONG* is therefore a potential target of SlRIN, while this is not the case for *SpHONG*.

Solanaceae contains more than 3000 species in around 90 genera, among which the largest genus is Solanum L. with about 1500 species (Knapp et al., 2004; Bohs, 2007). Cultivated tomato and its wild relatives have been classified in one of the sections within this genus, *Lycopersicon*, in which there are 13 species, including *S. pennellii* (Peralta et al., 2008). *S. pennellii* was selected as the species to establish functionality in regulating carotenoid biosynthesis is because it does not make lycopene in its fruit, resulting in green fruit when ripe (Figure 5.14). Within the *Lycopersicum* clade, apart from *S. lycopersicum* and its close relative *S. pimpinellifolium*, bearing small red fruit when ripe, all the others are non-red fruit species (Figure 5.14). As the genome sequences of *S. lycopersicum*, *S. pennellii* and *S. lycopersicoides* are available from the SOL Genomics Network (SGN) (https://solgenomics.net/), the promoter regions of HONG of another eight wild species were cloned and sequenced to estimate the likelihood of being direct targets of RIN according to the presence of CArG motifs. From the alignment of the *HONG* promoters among 11 species, both CArG motifs were missing in *S. harbrochites*, which is a green-fruited species, caused by a C to T SNP (Figure 5.13). In addition, in another green-fruited species, *S. lycopersicoidies*, one of the adjacent CArG boxes was abolished while the other one was intact, which shed light on the possible differential importance of the two adjacent RIN

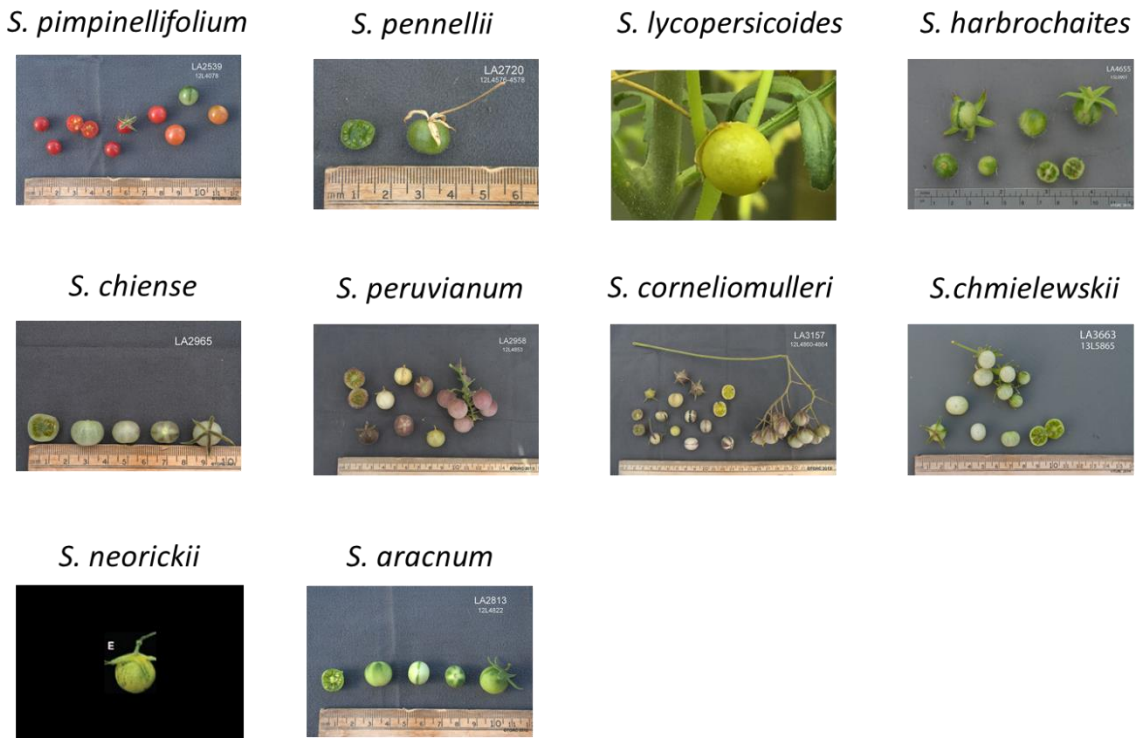


Figure 5.14 Representative pictures of fruit of selected tomato wild relatives. Pictures were from Tomato Genetics Resource Centre Website (http://tgrc.ucdavis.edu/)

binding motifs (Figure 5.13). *S. pimpinellifolium* had the same CArG-boxes as *S. lycopersicum*, consistent with its red-fruit phenotype, reflecting the successful accumulation of lycopene (Figure 5.13, 5.14). There were no differences observed between the *HONG* promoters of the other wild species analysed and cultivated tomato, which means that the direct binding between *RIN* and HONG promoters in these species could still be activated, indicating that the difference of accumulation of lycopene was more likely to be the result of other differences, perhaps involving the absence of CArG boxes further upstream in the *HONG* promoter, in these species (Figure 5.14). This conclusion needs to be validated further experimentally, such as with DLR assays.

These results were consistent with the hypothesis that SlRIN influences lycopene accumulation both through direct regulation of the expression of *SlPSY1*, the reported rate-limiting step in the carotenoid biosynthetic pathway, and/or indirectly by controlling the expression of genes encoding specific transcription factors, such as *SlHONG*, which regulate the expression of structural genes by direct binding.

5.3.3 Genome-wide phylogenetic study of basic helix-loop-helix transcription factors reveals SlHONG might be involved in the brassinosteroid signalling pathway.

Basic helix-loop-helix proteins serve various roles in the regulation of a wide range of physiological processes in plants. Most genome-wide phylogenetic studies of the transcription factors within this superfamily were conducted in *Arabidopsis thaliana* (Heim et al., 2003; Toledo-Ortiz, 2003). However, as a climacteric plant species, biological processes involved in plant growth and development in tomato were largely divergent from the model plant, and much less attention has been paid to the systematic study of bHLH superfamily in this very important crop, which has impeded the identification and functional characterisation of bHLH proteins in tomato. Genome-wide phylogenetic analysis was performed in Arabidopsis and tomato, with the identification of 159 bHLH proteins in tomato, which is very similar to the number of bHLH genes in Arabidopsis. To understand the evolutionary relationship of the tomato bHLH proteins, a Neighbour-joining tree of the

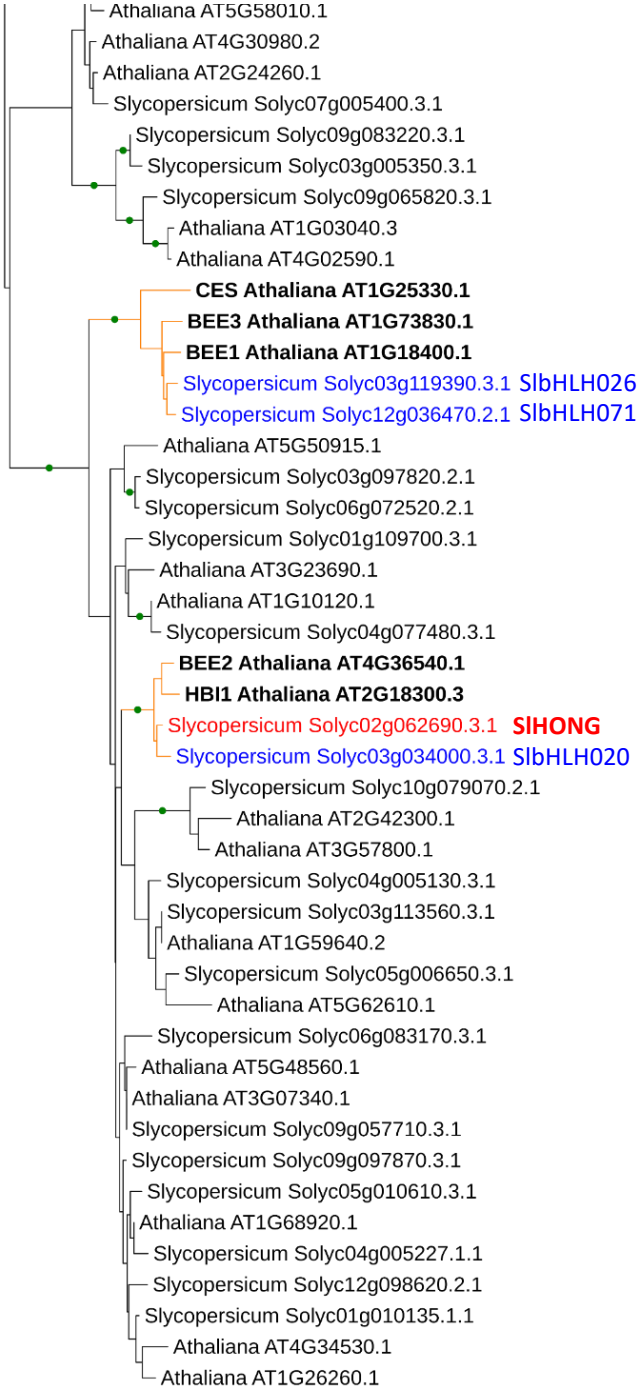


Figure 5.15 Neighbour-joining tree of tomato and Arabidopsis bHLH transcription factors in subgroup XII based on alignment of the bHLH domain. Red shows SlHONG. Blue shows SlbHLH071, SlbHLH026 and SlbHLH020.

identified transcription factors was drawn based on the multiple sequence alignment of the typical bHLH domains in these two species (Figure 5.15). In Arabidopsis, three members of the bHLH subgroup XII, BEE1 (AT1GT18400.1), BEE2 (AT4G36540.1) and BEE3 (AT1G73830.1), have been shown to take part in a wide range of pathways controlling multiple aspects of plant growth and development (Friedrichsen et al., 2002). These bHLH transcription factors, as well as HBI1 (AT2G18300.3), a homolog of BEE2, are closely related to each other and function redundantly in the early response pathway to BR signalling as positive regulators (Bai et al., 2012; Friedrichsen et al., 2002; Malinovsky et al., 2014). CESTA, another bHLH protein (AT1G25330.1), interacts with BEE1, and binds to the promoter region of one of the BR biosynthetic genes, *CPD* (*Constitutive Photomorphogenesis and Dwarfism*) to activate its expression (Poppenberger et al., 2011). Based on the phylogenetic analysis of tomato and Arabidopsis bHLH transcription factors, SlHONG was identified as most similar to two Arabidopsis bHLH proteins, BEE2 (AT4G36540.1) and HBI1 (AT2G18300.3). BEE2 is reported to form a complex with BEE1 (AT1G18400.1) and BEE3 (AT1G73830.1) which functions as a positive regulator of brassinosteroid signalling. HBI1 is a homolog of BEE2, overexpression of which results in increased hypocotyl and petiole elongation in Arabidopsis, contributing to related BR responses. It is possible that SlHONG might work with other bHLH proteins in tomato to control the BR signalling pathway as in Arabidopsis.

5.3.4 SlHONG plays a crucial role in forming BEE complexes with other bHLH proteins in tomato.

Phylogenetic analysis of other subgroup XII bHLH proteins in tomato, and comparison to Arabidopsis (Figure 5.15), showed three other tomato bHLH proteins that were highly similar in terms of their protein amino acid sequences to the members of BEE complex in Arabidopsis; SlbHLH071 (Solyc12g036470), SlbHLH026 (Solyc03g119390), SlbHLH020 (Solyc03g034000). It has been shown that BEE1, BEE2, BEE3 and CESTA (AT1G25330.1), a bHLH transcription factor in subgroup XII interacting with BEE1, can form a complex through direct interaction, which serves a positive role in regulating responses to BR. To capture the expression profiles of these three BEE protein candidates during fruit ripening, the

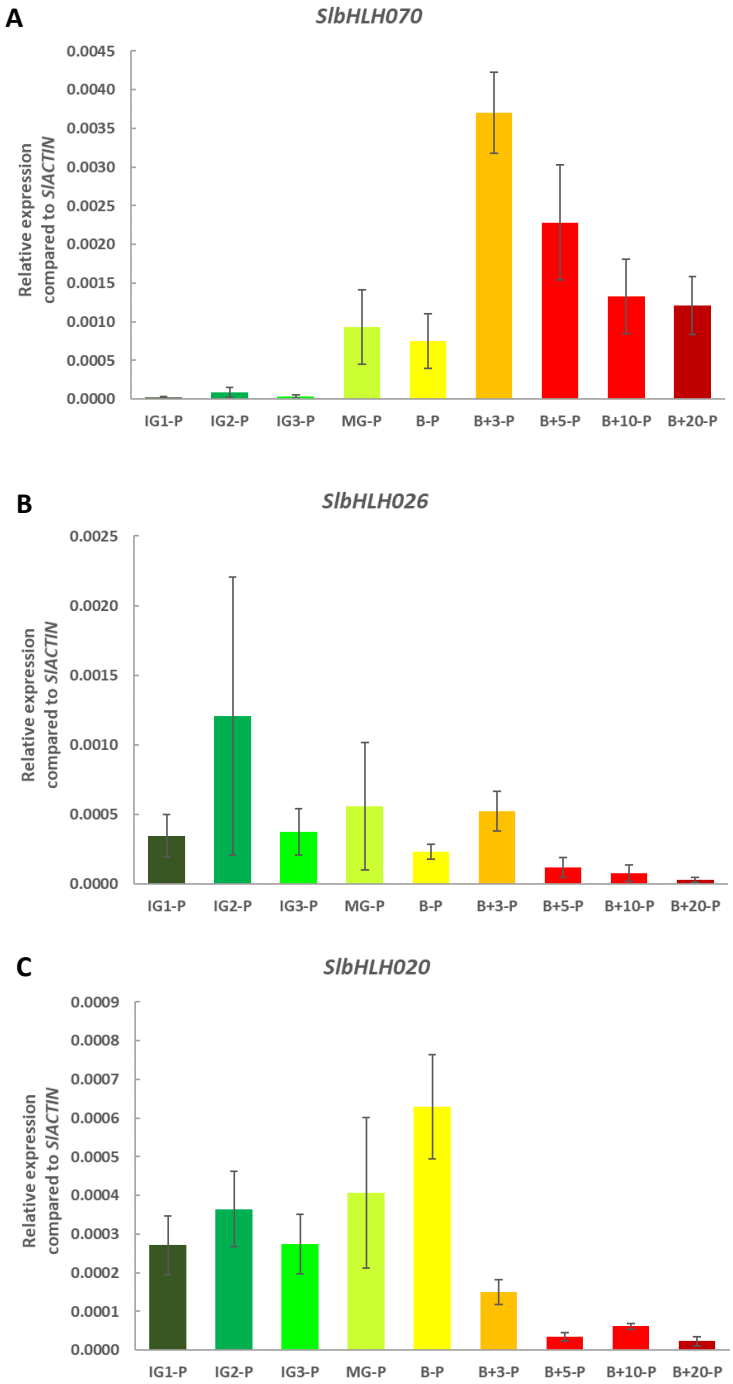


Figure 5.16 Expression of *SlbHLH070*, *SlbHLH026* and *SlbHLH020* in tomato fruit during ripening. The expression was analysed using quantitative PCR. Expression levels shown are relative to that for *ACTIN* (*Solyc03g078400*). Error bars, s.e.m.; three biological replicates, each with three technical replicates. IG1, 7 days after anthesis; IG2, 19 days after anthesis; IG3, 29 days after anthesis; MG, mature green (39 days after anthesis); B, breaker (42 days after anthesis); B+3, 3 days after breaker; B+5, 5 days after breaker; B+10, 10 days after breaker; B+20, 20 days after breaker; P, pericarp.

transcript levels were measured by RT-qPCR using gene-specific primers. Measurements were made using the same fruit samples as those used for measuring *SlHONG* expression levels. *SlbHLH020*, encoding the bHLH protein most closely related to SlHONG, was expressed at a very high level from the beginning of fruit development, which was very similar to *SlHONG* expression, as its expression reached a peak at the breaker stage. Expression of *SlbHLH20* decreased during fruit ripening (Figure 5.16 C). The expression pattern of *SlbHLH070* was very similar to that of lycopene biosynthetic genes, in which the transcripts stayed at very low levels during the green stages, and started to increase dramatically during the mature green (MG) stage, and reached a maximum approximately three days after breaker (B+3), and then decreased during fruit maturation (Figure 5.16 A). *SlbHLH026* also showed dynamic changes in its expression, although it was relatively more highly expressed in green stages than that in red stages (Figure 5.16 B).

A yeast two-hybrid experiment was performed to determine whether any direct interactions could be detected between these bHLH proteins of tomato. Firstly, each bHLH transcription factor, including SpHONG, was tested, for whether it could bind itself and form a homodimer. SlbHLH026 showed strong auto activation, when fused to the binding domain, so interactions with this protein must be interpreted with caution. All 5 bHLH proteins tested could homodimerize (Figure 5.17 B), although the positive result for SlbHLH026 could be due to autoactivation. SlHONG and SpHONG showed identical binding properties in the yeast two-hybrid experiment, which was expected because exactly the same bHLH domain was present in both proteins. Both SlHONG and SpHONG could interact with the other three putative bHLH proteins of the BEE complex when fused to the GLAL4 DNA-binding domain, and the other BEE proteins were fused to the activation domain of GAL4, while a similar broad binding activity was not detected for the other two transcription factors (SlbHLH071 and SlbHLH020) when they were fused to the Gal4 DBD. Assessment of the ability of SlbHLH026 to bind to the other TFs was confounded by the autoactivation but when fused to the GAL4 AD it interacted with SlbHLH071 fused to the GAL4 DBD. This suggested that SlHONG could play a central role in forming the putative BEE complex in tomato, through direct interaction with other bHLH proteins within the complex. Other interactions among BEEs were observed in Arabidopsis, such as between BEE1 and CES. *SlbHLH071*, *SlbHLH026*

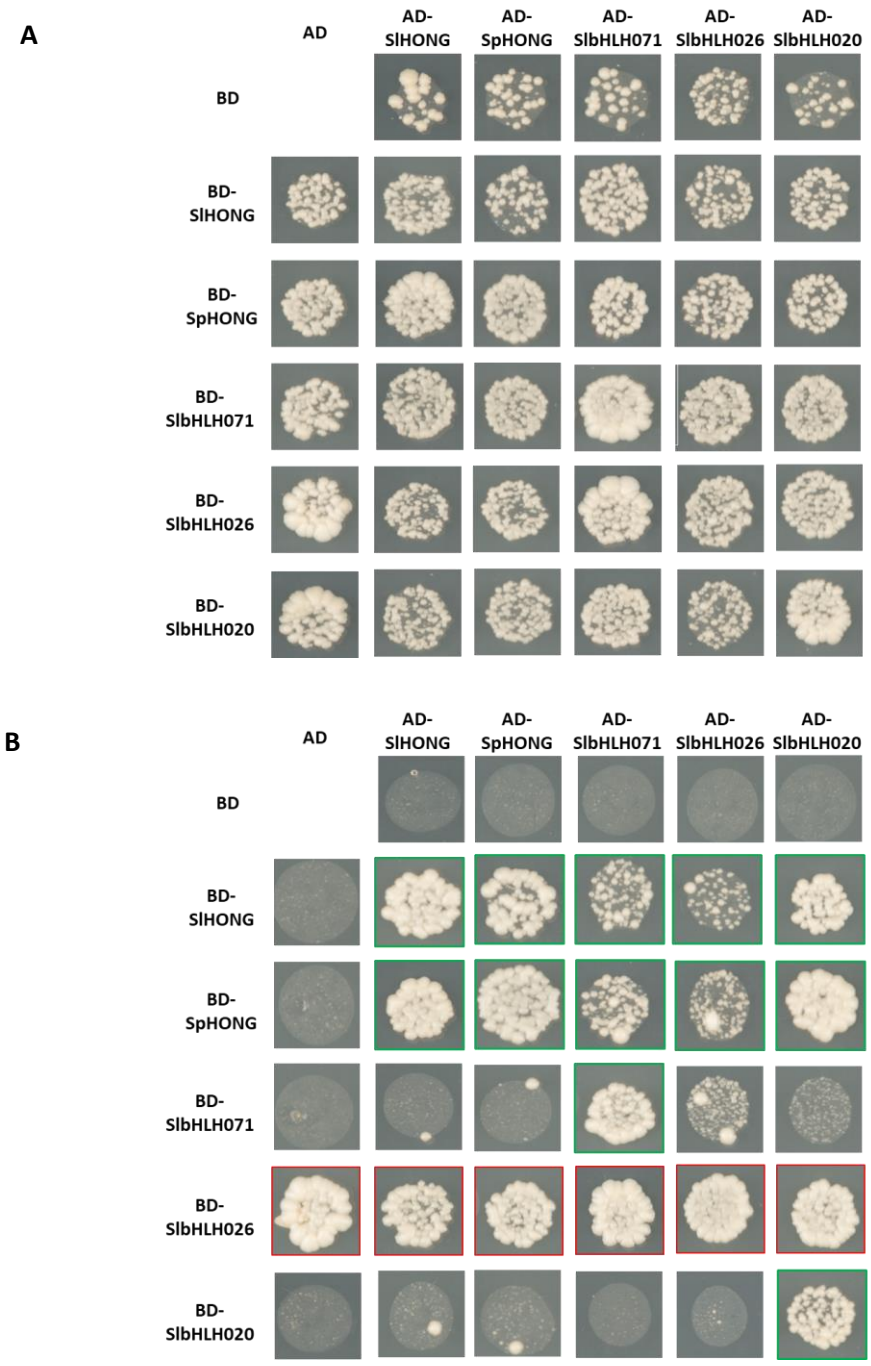


Figure 5.17 Growth of yeast cells co-transformed with the listed constructs. Transformants were spotted on SD/-Leu-Trp (A) or SD/-Leu-Trp-His-Ade (B) medium. A, the growth of yeast colonies indicates the successful transformation of the constructs. B, the growth of yeast colonies indicates the interactions between proteins encoding by the constructs contained in AD and BD. Autoactivation is indicated by red outline.

and *SlbHLH020* displayed distinct expression patterns during fruit ripening (Figure 5.16), and *SlHONG* maintained a relatively high transcript level throughout the different stages of ripening (Chapter 4, Figure 4.3 A). As all the three bHLH proteins can interact with SlHONG, the BEE complex in tomato may be regulated by the dynamic activities of *SlbHLH071*, *SlbHLH026* and *SlbHLH020*, as well as by regulators of SlHONG, such as SlRIN.

5.3.5 SlHONG can interact with other bHLH transcription factors, suggesting that it functions in regulating a wide range of biological processes.

Apart from the direct interactions detected among the members of the putative BEE complex in tomato, SlHONG could interact with other bHLH proteins. To identify other bHLH transcription factors potentially functioning with SlHONG by forming heterodimers regulating lycopene biosynthesis, the introgressed regions in the *trans*-eQTL hotspots characterised in Chapter 3 were re-examined, which were IL 2-1, IL 9-3 and IL12-1-1, each showing reduced expression of the genes of lycopene biosynthesis (Chapter 3, Figure 3.7). However, there were no bHLH transcription factors annotated within these candidate ILs. According to the co-expression analysis (Figure 3.7), lycopene biosynthetic genes are also down-regulated in IL8-2/IL8-2-1/ IL8-3, indicating that the region replaced by the corresponding genome of *S. pennellii* in these three ILs might harbour candidate regulators (Figure 3.7). Taking the gene expression profiles into consideration, four bHLH transcription factors were selected for further tests, including Solyc08g076820 (SlbHLH146), Solyc08g076930 (SlbHLH 147), Solyc08g08114 (SlbHLH090) and Solyc08g089240. Solyc03g120530 in IL 3-5 was also selected as a candidate regulator. Another three bHLH proteins, Solyc01g096050, Solyc09g083220 (SlbHLH060) and Solyc06g051260 (SlbHLH043) were included in the Y2H test based on literature mining (Ye et al., 2015). Similar yeast two hybrid assays were undertaken to detect interactions between the eight bHLH candidate transcription factors and HONG from *S. lycopersicum* and *S. pennellii* separately. In general, SlHONG and SpHONG showed the same binding activities with the tested TF candidates, in line with the presence of exactly the same bHLH domains in both proteins (Figure 5.6).

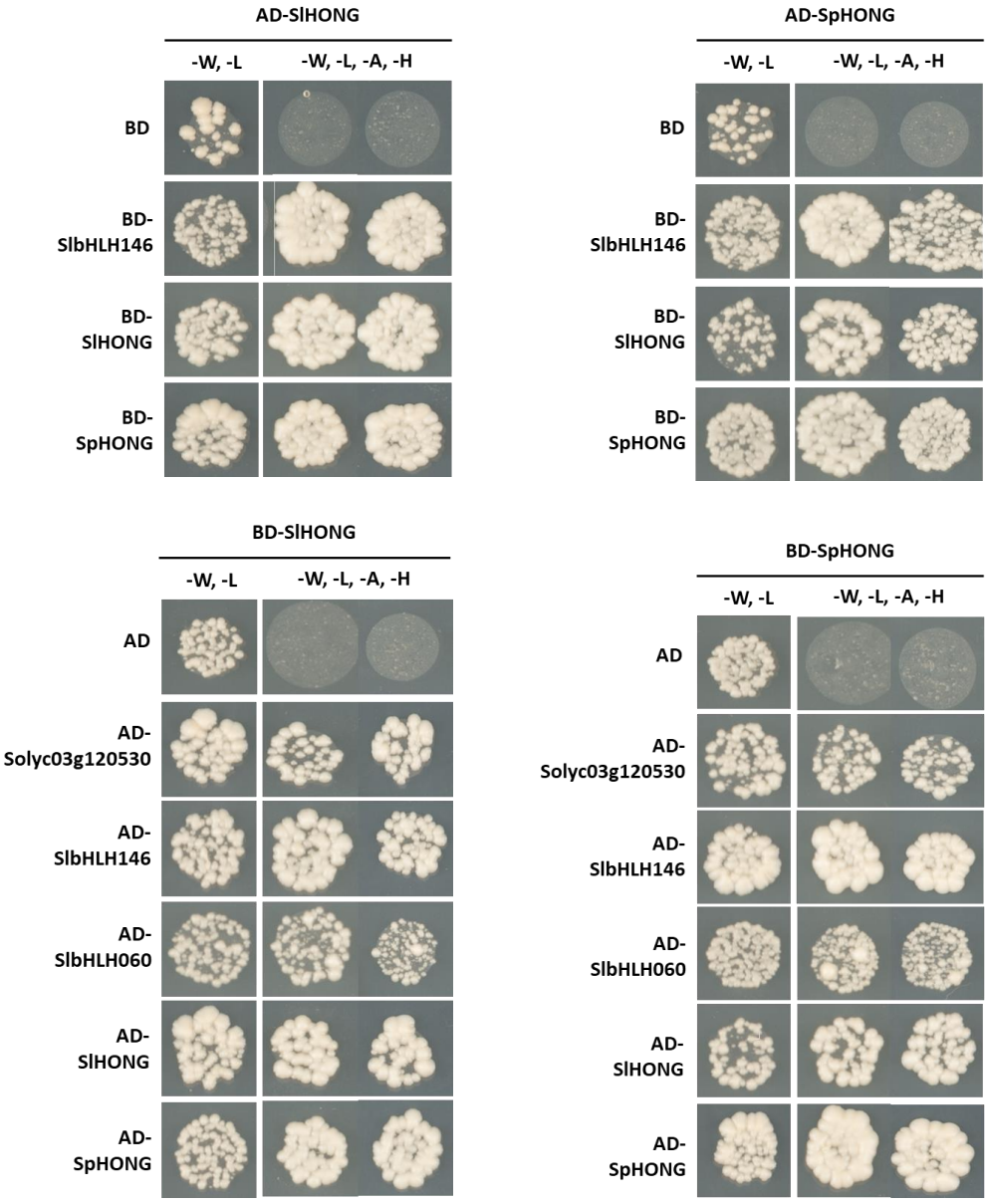


Figure 5.18 Growth of yeast cells co-transformed with the listed constructs, revealing the bHLH or HLH proteins interacting with SlHONG or SpHONG. Transformants were spotted on SD/-Leu-Trp (-W, -L) or SD/-Leu-Trp-Ade-His (-W, -L, -A, -H) medium. The growth of yeast colonies on (-W, -L) SD medium indicates the successful transformation of the constructs and the growth of yeast colonies on (-W, -L, -A, -H) SD medium indicates the interactions between proteins encoding by the constructs contained in AD and BD. The strength of the interaction was indicated by the growth of yeast colonies.

Strong interactions were detected between HONG and Solyc03g120530, SlbHLH146 and SlbHLH060, which were indicated by the growth of the corresponding yeast colonies (Figure 5.18). Based on the homology of bHLH domains, *AT1G03040*.3 and *AT4G02590*.1 encode the most similar bHLH proteins to SlbHLH060 in Arabidopsis (Figure 5.19 A). No functional characterisation has been undertaken for SlbHLH060, or for its close relatives, and further experiments are needed to analyse the roles of the heterodimer of SlHONG and SlbHLH060 in tomato plants. AT3G24140.1 was identified as the protein most closely related to SlbHLH146 in tomato (Figure 5.19 B). AT3G24140.1 has been shown to be a FAMA protein involved in the regulation of cell division and cell fate in stomatal development (Hachez et al., 2011). According to the Tomato eFP Browser, SlbHLH146 is also most highly expressed in leaves, consistent with the high transcript level of SlHONG detected in vegetative tissues, indicating that SlHONG might play a role in the regulation of stomatal development working together with SlbHLH146. Tomato fruit does not have stomata however.

In bHLH proteins, the Helix-loop-helix (HLH) region is functionally distinct from the basic domain, composed of hydrophobic amino acids contributing to the formation of homodimers or heterodimers with other bHLH proteins (Toledo-Ortiz et al. 2003). The basic region is involved in binding DNA (Robinson et al., 2000). bHLH dimers can serve as transcriptional activators or suppressors of their target genes depending on the properties of the corresponding bHLH proteins. It has been suggested that non-DNA-binding bHLH proteins can negatively regulate the transcription of target genes either by forming an inactive heterodimer with another non-DNA-binding bHLH or by sequestering a DNA-binding bHLH by direct binding, a process also referred to as 'squelching' (Littlewood and Evan 1995, Ikeda et al. 2012). According to the results of the yeast two-hybrid experiment, Solyc03g120530 can interact with SlHONG. The key residues, His-Glu-Arg (H-E-R), which are responsible for binding to the G-box (E-box), are absent in Solyc03g120530, which suggests that Solyc03g120530 might be a non-DNA-binding bHLH protein. According to the Tomato eFP browser, expression of *Solyc03g120530* is highest in roots, followed by flowers, and lowest in ripening fruit, partially distinct from the expression of *SlHONG*, which was highest in flowers as well as green tissues, and lower in roots. It could be that Solyc03g120530 acts as a repressor of carotenoid biosynthetic pathway in a tissue-specific way, by forming a

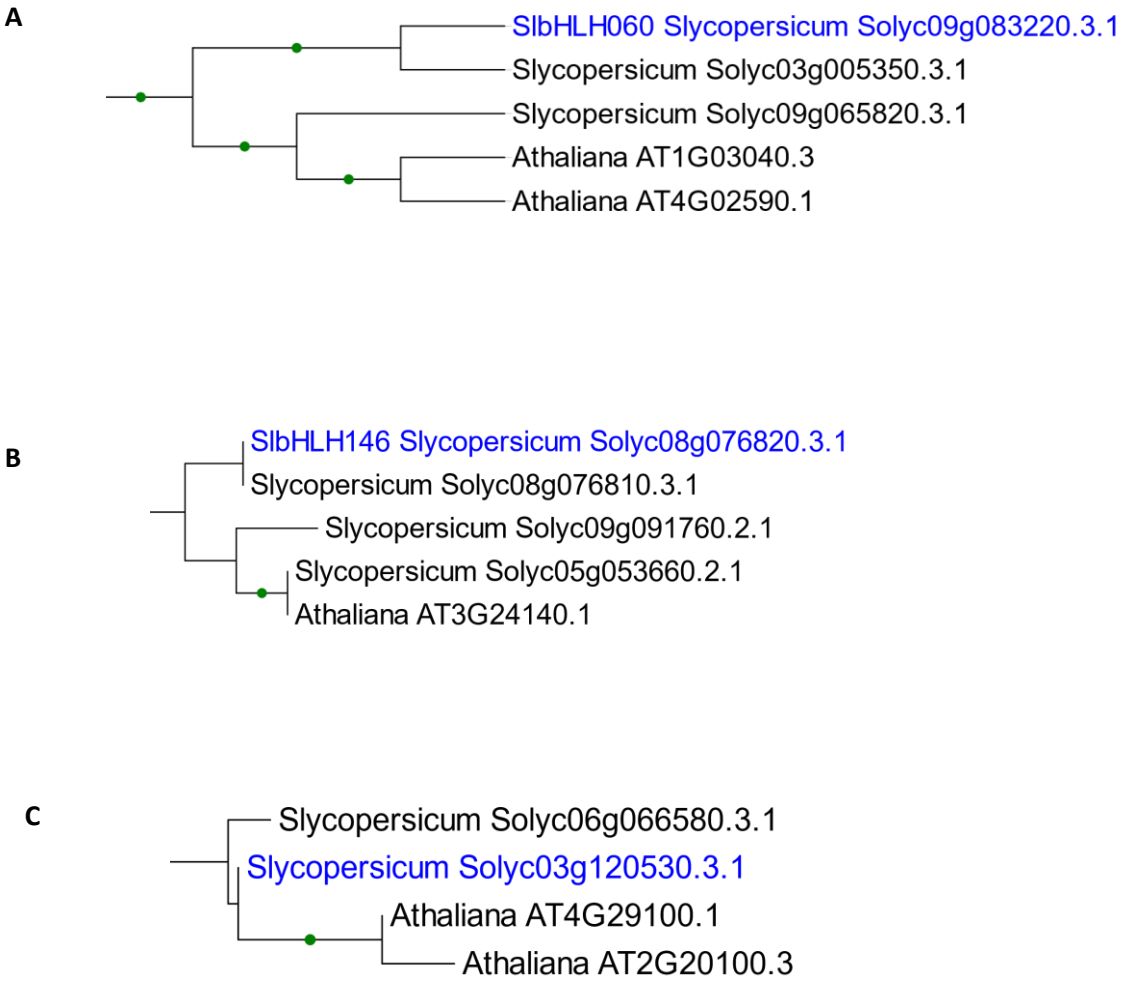


Figure 5.19 Parts of phylogenetic trees from Neighbour-joining tree of tomato and Arabidopsis bHLH transcription factors based on the amino acid sequence of bHLH domain**.** SlbHLH060, SlbHLH146 and Solyc03g120530 could interact with SlHONG according to the yeast two hybrid experiments (Figure 5.18).

heterodimer with SlHONG, or, of course, that it is functional with or without SlHONG, in other physiological processes.

Two Arabidopsis bHLH transcription factors At4G29100 (AtbHLH68) and AT2G20100 were identified as the proteins most similar to of Solyc03g120530 by analysis of the homology between the proteins based on the sequences of both of the entire protein and of bHLH domains (Figure 5.19 C). It has been reported that AtbHLH68, which is expressed in the vascular tissues of Arabidopsis, takes part in root development, and tolerance to drought stress, possibly in an ABA-dependent way (Le Hir et al., 2017). Overexpression of AtbHLH68 in Arabidopsis results in defects in lateral root development, increased resistance to drought stress and down-regulation of the genes involved in ABA signalling (Le Hir et al., 2017). It has been shown that AtbHLH68 can down-regulate the expression of two DNA-binding bHLH proteins AtMYC2 and AtbHLH122, which have been characterised as positive regulators of drought tolerance (Abe et al. 2003, Liu et al. 2014). However, no direct interaction has been reported between AtbHLH68 and these two bHLH proteins. This leaves how AtbHLH68, as a non-DNA-binding protein, functions as a repressor of the transcription of genes involved in ABA signalling and drought responses as an open question. The high expression of Solyc03g120530 in tomato root indicated that it is possible that it has a function similar to its Arabidopsis homolog. The direct interaction between Solyc03g120530 and SlHONG may shed light on the mechanism underlying the function of AtbHLH68 as a negative regulator. ABA is derived through an oxidative cleavage reaction in plastids from C_{40} epoxycarotenoid precursors, produced by the carotenoid biosynthetic pathway (Liming Xiong, 2003). I have shown that SlHONG may play a positive role in carotenoid biosynthesis. It would be interesting to test root development and drought tolerance in the SlHONG mutants, in both knock-out lines and overexpression lines, to characterise the potential function of SlHONG in the regulation of ABA homeostasis.

5.4 Discussion

5.4.1 SlHONG may activate the expression of lycopene biosynthetic genes in tomato fruit through direct binding under the control of the master regulator of fruit ripening, SlRIN.

Several transcription factors have been identified that participate in the transcriptional regulation of carotenoid accumulation in tomato fruit, most of which influence carotenoid biosynthesis through broad effects on fruit ripening, such as SlEF6, *SlAP2a and* TAGL1 (Chung et al., 2010; Vrebalov et al., 2009; Lee et al., 2012). RIN (ripening inhibitor), a MADS-box transcription factor, a master regulator of fruit ripening, is expressed specifically in fruit tissues from the breaker stage. RIN has been demonstrated to regulate carotenoid accumulation by interacting with the *SlPSY1* promoter (Martel et al., 2011). Ethylene serves a crucial role in fruit ripening which, is triggered by a dramatic increase in ethylene production. It has been shown that expression levels of *SlPSY1* and *SlPDS* are correlated with ethylene levels, contributing to the corresponding accumulation of lycopene and β-carotene (Marty et al., 2005). RIN interacts directly with the promoters of genes involved in ethylene biosynthesis, such as *ACC SYNTHASE2* (*ACS2*) and *ACS4*, as well as, *ETHYLENE RECEPTOR 3* (*ETR3/NR*) (Martel et al., 2011; Fujisawa et al., 2013, 2011). Therefore, carotenoid accumulation in tomato fruit is regulated by SlRIN either through direct binding to the *SlPSY1* promoter, or by regulating ethylene production. From the DLR assays, I found that both SlHONG and SpHONG could activate the transcription of some lycopene biosynthetic genes, especially *SlPSY1* through direct binding. Two CArG motifs, responsible for SlRIN-binding to DNA, were found in the promoter region of *SlHONG*, which may be responsible for the direct binding observed. Both CArG motifs were abolished in the promoter of *SpHONG* by a SNP, which may contribute to differential performance of HONG initially observed between IL2-1 and IL2-1-1, resulting in the differential transcript abundance of lycopene biosynthetic genes in ripening fruit. Thus, SlRIN might control lycopene accumulation not only through direct interaction with the *SlPSY1* promoter, but also through SlHONG, which can bind to the *SlPSY1* promoter and activate gene expression.

5.4.2 SlHONG may play a central role in a putative BEE complex in tomato controlling brassinosteroid signaling.

In Arabidopsis, four bHLH transcription factors, BEE1, BEE2/HBI1, BEE3 and CESTA, form a complex that functions as a positive regulator in the early response in brassinosteroid signaling. Genome-wide phylogenetic analysis of bHLH proteins in tomato and Arabidopsis revealed that SlHONG was most similar to BEE2 and HBI1. Another three bHLH transcription factors were identified that could comprise a BEE complex in tomato, which were SlbHLH020, SlbHLH026 and SlbHLH071. According to the yeast-two hybrid assays, SlHONG was the only protein within the complex which could interact with all the other three bHLH transcription factors, suggesting that it plays a crucial role in forming the BEE complex in toamto. The phenotypes described in Chapter 4 caused by overexpression or knock-out of SlHONG, such as significantly reduced hypocotyl elongation in homozygous SlHONG knock-out lines, were very similar to those resulting from defects in BR signalling in Arabidopsis (Bai et al., 2012). SlHONG was initially identified by eQTL analysis of *S. lycopercium* x S. *pennellii* population in the trans-eQTL hotspot, IL2-1 (Chapter 3). However, the other three bHLH proteins, SlbHLH020, SlbHLH026 and SlbHLH071, which may form the tomato BEE complex with SlHONG, were not identified in my analysis. This might have been because my *trans*-eQTL hotspot identification was primarily focused on the biosynthetic gene-free regions (Chapter 3, Figure 3.3). Apart from *SlHONG*, the other bHLH proteins were co-localized with at least one of the carotenoid biosynthetic genes. For example, the gene encoding SlbHLH020, the most similar bHLH protein to SlHONG, was co-localized with *SlPSY1* in IL3-2, encoding the rate-limiting enzyme of lycopene biosynthesis. *SlbHLH071* and *SlbHLH026* were co-localized with *SlZISO* in IL12-3 and *SlPDS* in IL3-5 respectively. Co-localization made any effects of these transcription factors on the transcript abundance of the biosynthetic genes were masked by their own *cis*-eQTLs. Carotenoid accumulation during fruit ripening is influenced by a lot of internal and external factors, among which phytohormones play an important role in the signaling networks (Lee et al., 2012). For example, regulatory mechanisms involving ethylene regulation of carotenoid biosynthesis in tomato fruit have been extensively studied. Auxin, abscisic acid and jasmonic acid also take part in regulation of fruit ripening, as well as carotenoid accumulation. (Jones et al., 2002; Galpaz et al., 2008; Liu et al., 2012).

2,4-Epibrassinolide (EBR) has been shown to regulate carotenoid accumulation (Vardhini and Rao, 2002). Furthermore, EBR-treated pericarp discs of *Nr* accumulated more carotenoids than those of the control, suggesting the existence of a BR-induced carotenoid accumulation pathway independent of NR-mediated ET signal transduction (Liu et al., 2014a). The transcription factor BRASSINAZOLE RESISTANT1 (BZR1) is a key component of BR signaling. Tomato fruit overexpressing the *Arabidopsis BZR1-1D* gene exhibited enhanced carotenoid accumulation and increased soluble solid, soluble sugar, and ascorbic acid contents during fruit ripening (Liu et al., 2014a). Because of the central role of SlHONG in forming putative BEE complex in tomato, as well as the direct interaction detected between SlHONG and the *SlPSY1* promoter, and between SlRIN and the *SlHONG* promoter, SlHONG might be involved in an ethylene-independent regulatory network involved in controlling carotenoid accumulation in tomato fruit.

Chapter 6

Genome-wide association analysis of carotenoids in fruits of a natural population

6.1 Introduction

6.1.1 Metabolite-based genome-wide association studies (mGWAS) in plants

Genome-wide association studies (GWAS) have been used extensively in plants to identify a wide range of loci controlling complex traits with rapidly increasing numbers of accessions in the mapping population (Nemri et al., 2010; Slavov et al., 2014). The exceptional diversity of metabolites in plants, which provide more than 200,000 structurally distinct compounds (Wurtzel and Kutchan, 2016), has led to a growing interest in the investigation of the natural variation in plant metabolism and its underlying genetic foundation by forward and reverse genetic approaches (Cardoso et al., 2014; Quadrana et al., 2014). With the rapid development of high-throughput profiling and genotyping technologies, such as mass-spectral profiling and next-generation sequencing, metabolite-based genome-wide association studies have emerged as a powerful tool for forward genetic approaches to elucidate the genetic and functional basis of metabolic diversity in plants. mGWAS, which is metabolite-based genome-wide association study, was firstly exploited in the model plant *Arabidopsis thaliana*, followed by successful applications in a number of different plant species, including rice, tomato and maize (Chan et al., 2011; Chen et al., 2014; Riedelsheimer et al., 2012; Rosenwasser et al., 2014; Zhu et al., 2018).

According to the specificity of targets in metabolite profiling, mGWAS are generally classified as targeted, non-targeted or widely-targeted studies. Targeted metabolic profiling is directed at a specific subgroup of preselected compounds. For instance, in Arabidopsis, mGWAS focusing on 43 glucosinolates (GSLs) detected in 96 accessions generated around 230, 000 single nucleotide polymorphisms (SNPs) used for association study, with the identification two major loci controlling GSL variation of this population, harbouring dozens of genes (Chan et al., 2011). mGWAS targeted at other metabolites have also been conducted in plants, such as branched-chain amino acids (BCAA), fatty acids and tocochromanols (Angelovici et al., 2013; Matsuda et al., 2015; Lipka et al., 2013). Unlike targeted metabolite profiling, which is normally achieved by HPLC, non-targeted metabolomics is dependent on either mass spectrometry (MS) or nuclear magnetic

resonance (NMR) to allow broader metabolic profiles to be detected in a high-throughput way. The differential application of targeted and non-targeted profiling in mGWAS, is determined manually by the scientific questions being addressed and the level of knowledge about the metabolites of interest. Recently, a widely-targeted metabolomics technology based on LC-MS has been developed, combining the advantages of accurate measurement performed by targeted profiling and the wide coverage of non-targeted analysis (Chen et al., 2014; Wen et al., 2014).

6.1.2 Integrated application of mGWAS and other genome-scale approaches for functional genomics

In studying the diversity of plant metabolism, it is important to elucidate how metabolites are synthesized and regulated. Extensive efforts have been made to understand the genetic foundation of plant metabolism, by using various mapping populations particularly those derived from two distinct parental lines (Gong et al., 2013) coupled to analysis of the metabolome within the mapping population. With the emergence of genome-scale approaches, mGWAS has also been applied to functional genomics. Candidate genes responsible for the accumulation of certain secondary metabolites can be identified through correlative analysis of various -omics datasets including genomics, transcriptomics, proteomics and metabolomics. This strategy is further facilitated by recent advances in next generation sequencing technologies. Mapping approaches in plants have typically been conducted by using with bi-parental populations through linkage mapping. However, dissecting the complex relationships between metabolite accumulation and genetic variation was largely constrained by the huge genome size integrating with environmental effects. Compared to traditional genetics approaches, next-generation sequencing with longer reads can improve the accuracy and efficiency of the analysis. Resequencing-assisted genetics strategies has been widely applied in unveiling the complex genetic basis of specific metabolites in various species (Pawelkowicz et al., 2015; Huang et al., 2009; Zhu et al., 2018). In addition, the analysis of natural variation using population genetics has been adopted widely. Plant metabolite diversity is usually investigated by linkage mapping, e.g. quantitative trait locus (QTL) mapping using bi-parental populations and/or by genome-

wide association studies (GWAS) using unrelated natural populations. Furthermore, with the increasing number of genome sequences available from closely or more distantly related species, it has become possible to combine comparative genomic analysis with metabolomics for gene identification and pathway elucidation. For example, in an mGWAS in rice, 36 candidate genes involved in the regulation of the levels of metabolites with important physiological and nutritional value were identified, five of which were characterised or annotated as genes encoding methyltransferases, a glucosyltransferase and three putative acyltransferases (Chen et al., 2014). Genome-wide transcriptomic analysis is also a genome-scale approach which can be combined with mGWAS, has been applied successfully to the characterisation of the biosynthesis of secondary metabolites, such as phenylpropanoids, flavonoids, steroidal glycoalkaloids (SGAs) (Alseekh et al., 2015; Cárdenas et al., 2016; Itkin et al., 2013).

6.2 Materials and Methods

6.2.1 Plant materials and growth conditions

A total of 154 tomato accessions were collected from USDA (United State Department of Agriculture), TGRC (Tomato Genetics Resource Centre), University of Florida, INRA (The National Institute for Agricultural Research) EU-SOL (The European Union-Solanaceae projce) and IVF-CAAS (The Institute of Vegetable and Flowers, Chinese Academy of Agricultural Science), including including two accessions of wild species (*S. cheesmaniae* and *S. peruvianum*), and 152 accessions from the red-fruited clade (*S. lycopersicum*, *S. lycopersicum* var *cerasiforme* and *S. pimpinellifolium*) (Appendix 4). Tomato plants were grown in the greenhouses of Agricultural Genomics Institute at Shenzhen, China. For the metabolic profiling, eight plants were grown for each accession. At least one fruit was selected from each plant at the red stage, and pooled together as one biological sample. Carotenoid profiles of each accession were generated from at least two independent biological samples.

6.2.2 SNP identification and annotation

The 154 accessions used were characterised by whole genome re-sequencing, which was performed in Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences (Zhu et al., 2018). DNA was extracted from young leaves for sequencing library construction with insert sized of about 500 bp following the manufactures's instructions (Illumina). The sequencing was conducted by Illumina HiSeq 2000 Platform with 100 bp and 125 bp paired end. The sequencing reads obtained were mapped to tomato reference genome by SOAP2 (Li et al., 2009b) according to the following parameters: -m 100, -x 888, -s 35, -l 32, -v 3. After being filtered to remove PCR duplicates, both pair-end and single-end reads were used for SNP calling by SOAPsnp based on following parameters: -L 100 -u -F 1 (Li et al., 2009a). According to SOAPsnp programme, SNPs with quality > = 40 and base quality > = 40 were selected for consensus calling and SNP detection.

6.2.3 Genome-wide association analysis

A total of 68,068 SNPs with minor allele frequency (MAF) >5% were used to perform the GWAS analysis, in which Factored Spectrally Transformed Linear module (FaST-LLM) was applied (Lippert et al., 2011). The genome-wide significance thresholds were determined by a modified Bonferroni correction, using the effective number of independent SNPs instead of the total number of SNPs. With a nominal level of 0.05, P_{LLM} = 1.44E-6 was used as the genome-wide significance threshold for all the carotenoids profiled.

6.2.4 Carotenoid extraction

300 mg of freeze dried fruit tissues were placed in a 15 mL Falcon tube. 5 mL of extraction solution, containing hexane: acetone: ethanol at 2: 1: 1 (v/v/v), with 0.01% (w/v) butylated hydroxytoluene(BHT), were added and vortexed vigorously. The mixture was sonicated for 30 min to remove bubbles and for homogenisation, followed by 10 min centrifugation at

3000 r/min at 4 °C. The supernatant was transferred into a new tube and the pellet was re-extracted twice by repeating the final step. The combined extract was vacuum concentrated and dissolved in 3.0 mL methyl tert-butyl ether (MTBE) with 0.01% BHT. The re-dissolved extract was then centrifuged at 13000 r/min at 4 °C for 10 min, and the supernatant was filtered through a microfilter (PTFE 0.22).

6.2.5 Carotenoid profiling by HPLC

The carotenoids were profiled by High Performance Liquid Chromatography (HPLC) using Waters® binary HPLC 1525 (PDA 2998). YMC C30 (Waters) was used, which is particularly designed for the separation of carotenoids, including isomers. The assay was performed under following conditions: column oven temperature, 28 °C; mobile phase, ACN: MeOH (3: 1, v/v); flow rate of mobile phase, 1 mL min^{-1}; injection volume, 20 µL; wavelength range, 200−750 nm. These analyses were performed by Dr Shouchuang Wang and Dr Wei Chen at Hazhong Agricultural University, China.

6.3 Results

6.3.1 Establishment of an HPCL method which can separate 11 carotenoids in a single run.

Tomato is one of the major sources of carotenoids in the human diet because it contains a broad range of different carotenoids present at different levels in fruit. A precondition of successful use of metabolomics to study the genetic basis of a specific pathway is to be able to distinguish fully all the targeted compounds by HPLC/MS methods. Two factors limit the accurate measurement of metabolites, one is the extraction method which is able to obtain most of the carotenoids from tomato fruit, and another is the establishment of a suitable HPLC programme to separate major carotenoids, clearly. After several attempts, my collaborators, Dr Shouchuang Wang and Dr Wei Chen from the group of Prof Jie Luo at

Hazhong Agricultural University, China, established a method that could detect 11 carotenoids at one time (Figure 6.1). At 450nm, around the point of maximum absorption for most of the carotenoids in which I was interested, including lycopene and β-carotene, nine carotenoids were clearly separated (the absorption peaks of phytofluene a/b were obtained at 320 nm, which is not shown in Figure 6.1). All measurements were made using this method.

To capture the abundance of different carotenoids in tomato, fruit samples from 24 accessions were randomly selected out of the total of 154 accessions and these were used for a trial mGWAS analysis. Substantial variance in the contents of each carotenoid analysed was observed, which offered the possibility to study the genetic basis of the regulation of carotenoid biosynthesis by mGWAS (Figure 6.2). Not surprisingly, among the total 11 carotenoids, the highest accumulation was obtained for lycopene and β-carotene, the most abundant carotenoids found in tomato fruit (Figure 6.3). The contents of lutein were also variable with a relatively high abundance from among all the carotenoids (Figure 6.3). This indicated that lycopene, β-carotene and lutein were very likely the most reliable compounds for identification of loci harbouring the genes underpinning the control of carotenoid biosynthesis in tomato fruit. The skewness and kurtosis analysis of the contents of these three major carotenoids showed that the distribution of lycopene and β-carotene was close to normal distribution, while lutein distributed in a more asymmetry manner. This could be explained that in cultivated tomatoes, the lutein contributes to a relatively small portion in total carotenoids compared to lycopene and β-carotene and the genetic basis of lutein biosynthesis is divergent from β-carotene biosynthesis.

6.3.2 Genetic basis of natural variation in major carotenoids revealed by mGWAS

We collected a total of 154 tomato accessions, including two accessions of wild species (*S.cheesmaniae* and *S. peruvianum*), and 152 accessions from the red-fruited clade (*S.lycopersicum*, *S. lycopersicum* var *cerasiforme* and *S. pimpinellifolium*). To investigate potential loci controling the natural diversity in carotenoid accumulation, mGWAS were

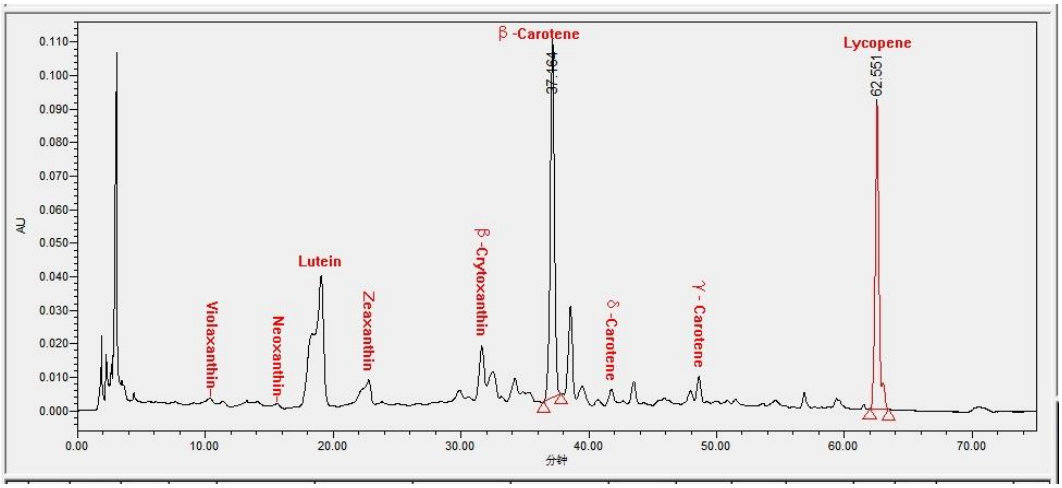
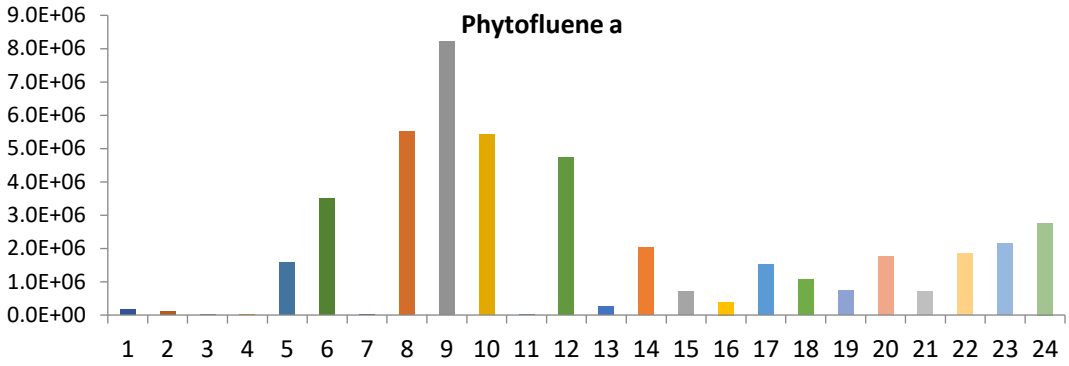
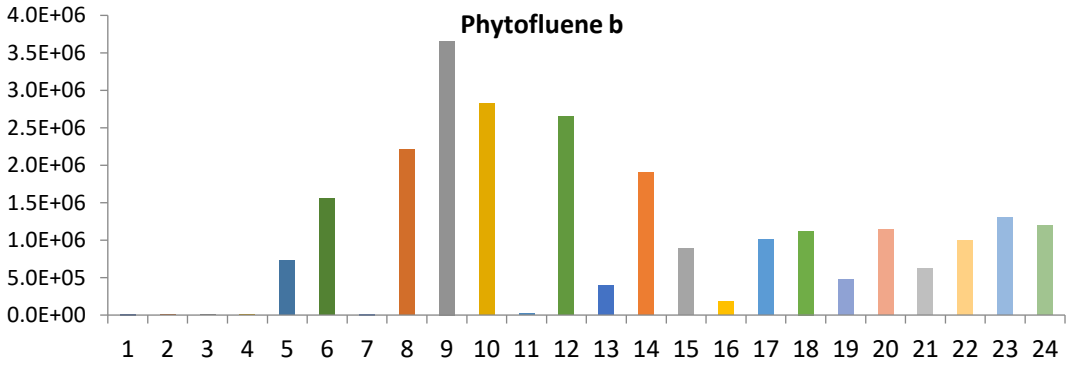
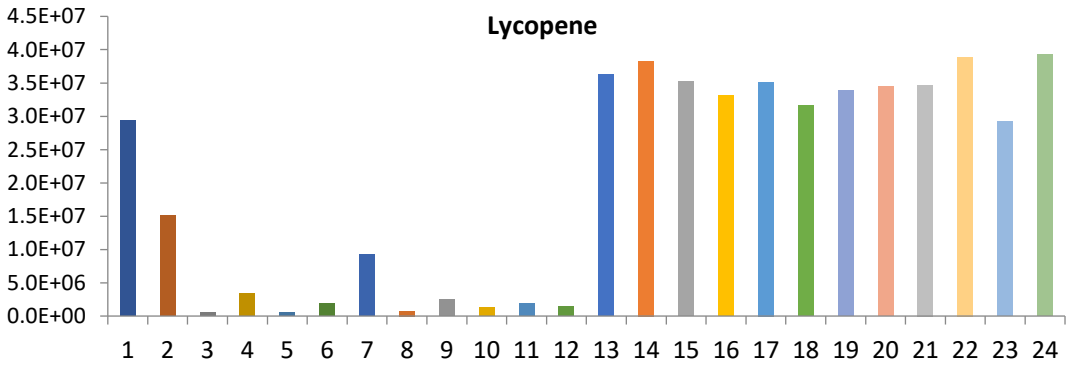
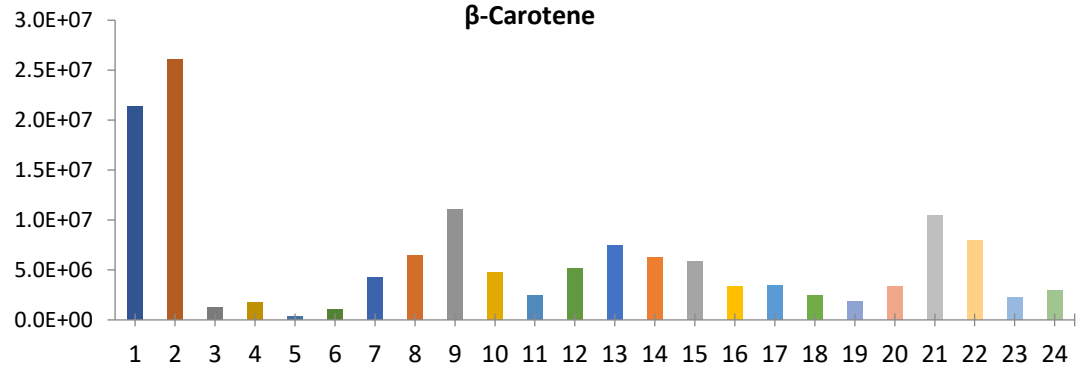


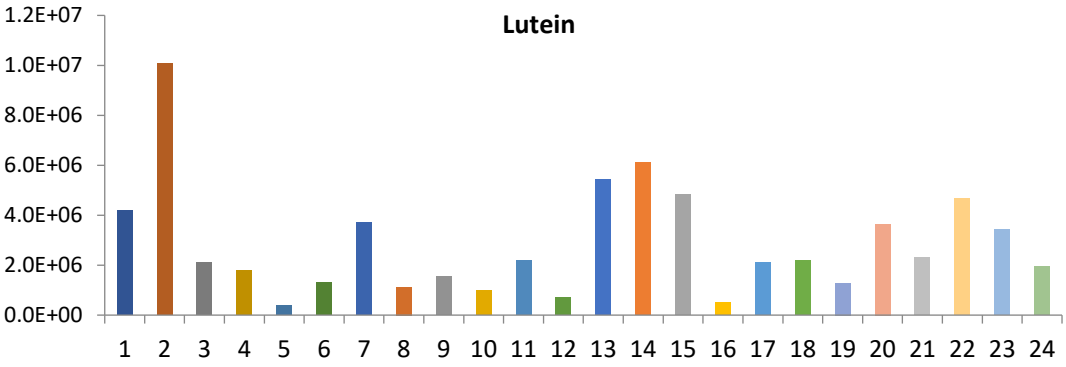
Figure 6.1 HPLC chromatogram of carotenoid extracts of tomato fruit detected at λ = 450 nm. All the carotenoids tested could be clearly separated. Phytofluene a/b were not included in the figure as their λ_{max} is 320 nm. *x* axis, retention time (minutes). *y* axis, the intensity of absorbance (in units of mAU). See method for chromatographic conditions.

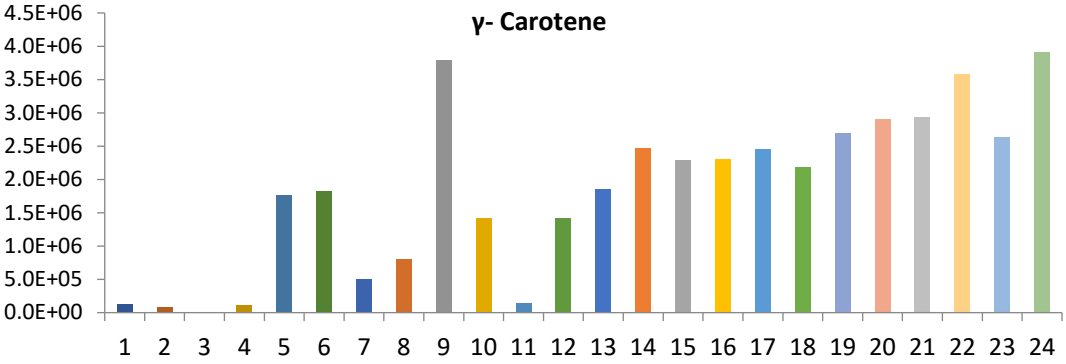

Phytofluene a
9.0E+06
8.0E+06
7.0E+06
6.0E+06
5.0E+06
4.0E+06
3.0E+06
2.0E+06
1.0E+06
0.0E+00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

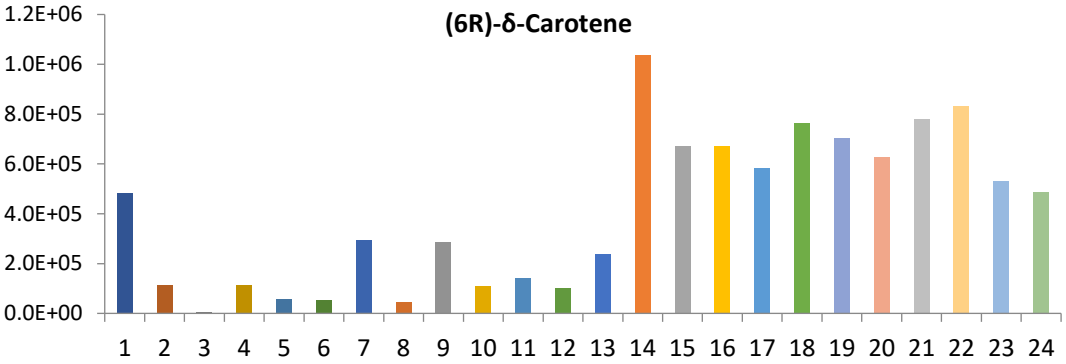

Phytofluene b
4.0E+06
3.5E+06
3.0E+06
2.5E+06
2.0E+06
1.5E+06
1.0E+06
5.0E+05
0.0E+00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

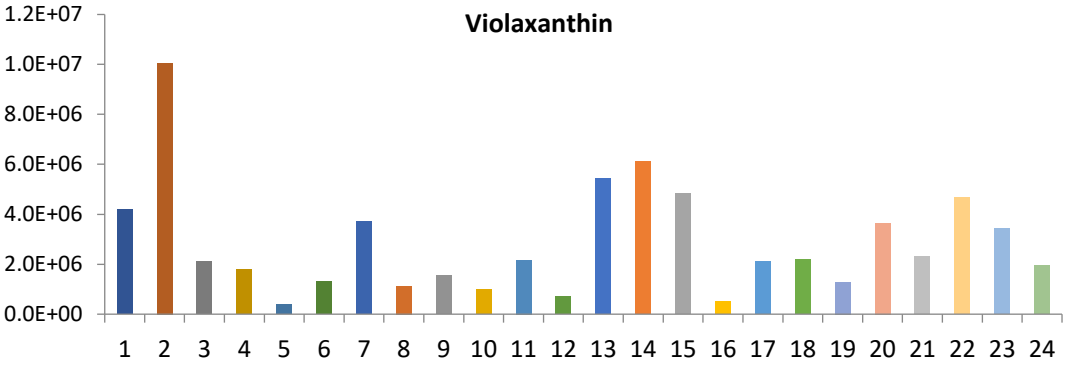

Lycopene
4.5E+07
4.0E+07
3.5E+07
3.0E+07
2.5E+07
2.0E+07
1.5E+07
1.0E+07
5.0E+06
0.0E+00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24


β-Carotene
3.0E+07
2.5E+07
2.0E+07
1.5E+07
1.0E+07
5.0E+06
0.0E+00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24


Lutein
1.2E+07
1.0E+07
8.0E+06
6.0E+06
4.0E+06
2.0E+06
0.0E+00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24


γ- Carotene
4.5E+06
4.0E+06
3.5E+06
3.0E+06
2.5E+06
2.0E+06
1.5E+06
1.0E+06
5.0E+05
0.0E+00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24


(6R)-δ-Carotene
1.2E+06
1.0E+06
8.0E+05
6.0E+05
4.0E+05
2.0E+05
0.0E+00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24


Violaxanthin
1.2E+07
1.0E+07
8.0E+06
6.0E+06
4.0E+06
2.0E+06
0.0E+00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

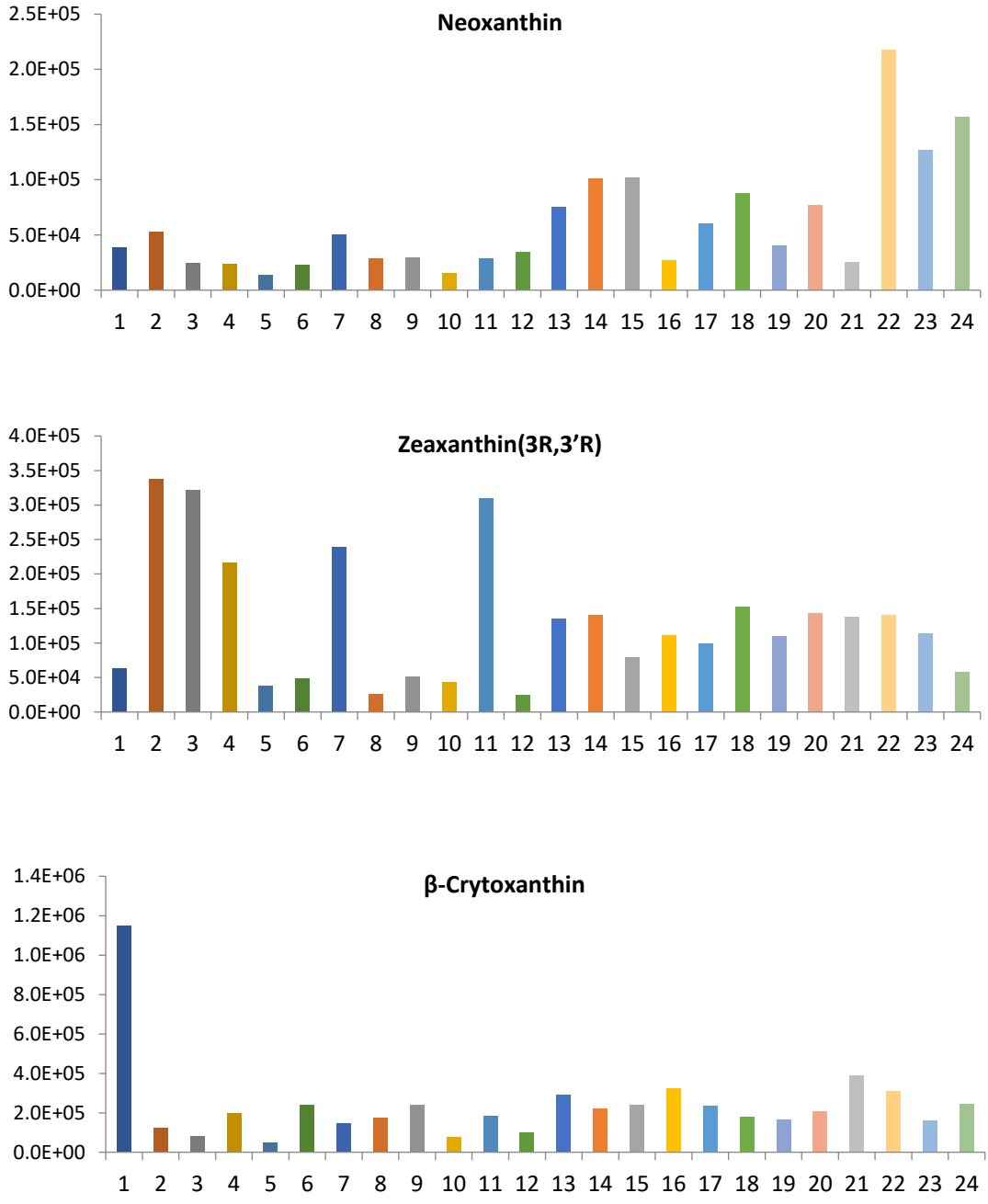


Figure 6.2 The contents of 11 carotenoids within the 24 randomly selected tomato accessions. *y* axis shows the relative abundance of each carotenoid based on the signal intensity captured in the analysis. The *x* axis corresponds to same order of the x axis in Figure 6.3, ID referred to Appendix 4 (the column of 'Individual Code').

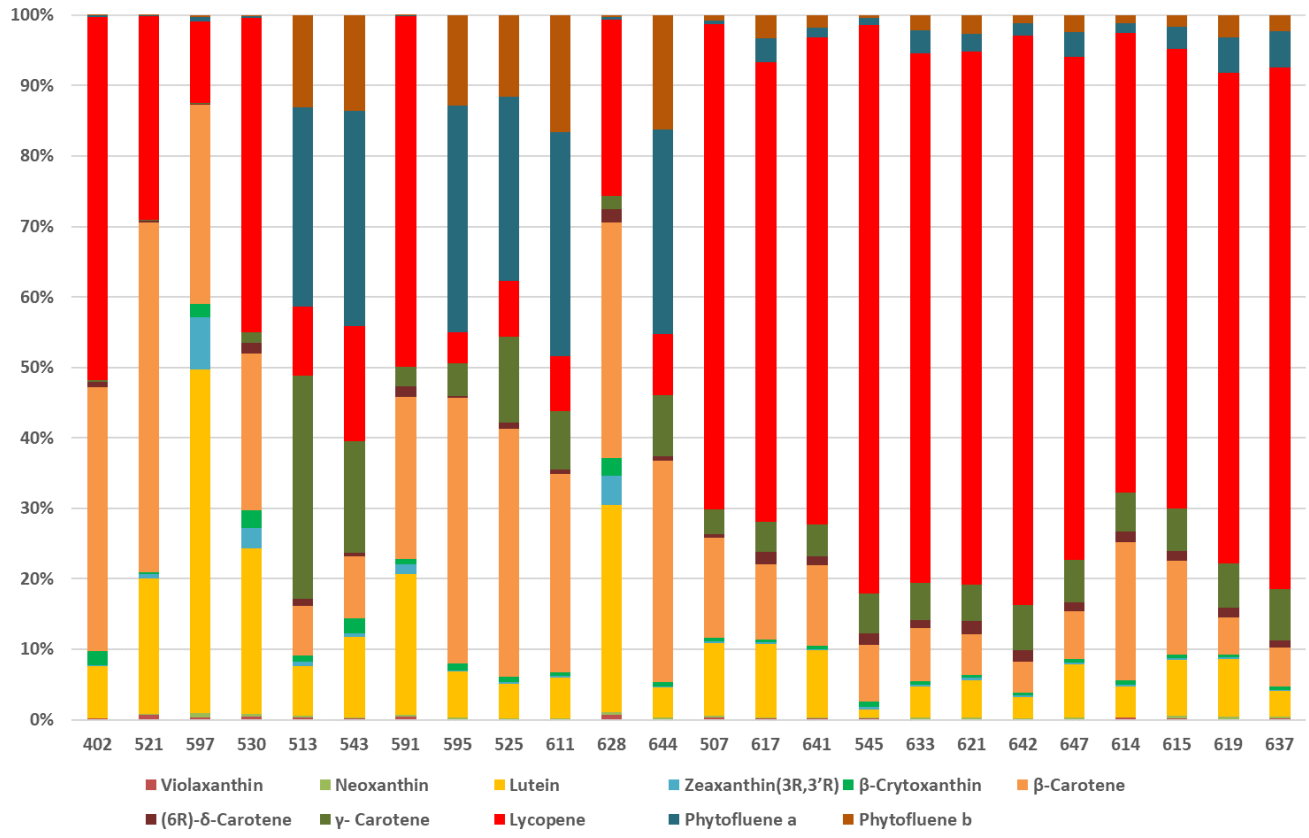


Figure 6.3 The percentage of each compound within the total carotenoids in each tomato accession analysed. The *x* axis shows the accession ID referred to Appendix 4 (the column of 'Individual Code'). The *y* axis represents the percentage of each carotenoid within the total carotenoids in each line tested.

then performed for the entire 154 tomato accessions. Five carotenoids were profiled across the population, including the three compounds detected with high abundance and variability in the running test, which were lycopene, β-carotene and lutein (Table 6.1). SNPs were identified using SOAPsnp (Li et al., 2009a). SOAP programme was used to align the short reads to the reference, with the allowance of a maximum of two mismatches. Two reads from a pair were aligned in the correct orientation with a proper coverage on the genome according to the reference. Only unique and ungapped reads aligned were used for SNP detection and consensus calling (Li et al., 2009a). After removing the low-quality SNPs with a missing rate (the fraction of missing calls per SNP over sample) >25% and those with a minor allele frequency (MAF) <0.05, the imputed genotypes of the association panel resulted 68,068 SNPs. A linear mixed model (LLM) that returned fewer false-positive results by considering the genome-wide patterns of genetic relatedness was used for the analysis, and the genome-wide significance threshold, P_{LLM}, was set to 1.44E-6 after Bonferroni correction. From the results of the mGWAS, the targeted compounds were visualized by Manhattan plots with genomic coordinates (chromosome 1-12) displayed along the *x* axis and the negative logarithm of the associated *P* value for each SNP displayed on the *y* axis (Gibson, 2010). This analysis identified a total of 212 lead SNPs (P_{LLM} < 1.44E-6) corresponding to the three major carotenoids. Lead SNPs, which were significantly associated with the levels of corresponding metabolites, were identified for lycopene, β-carotene, lutein and γ-carotene, the numbers of which were 67, 57,178 and 15 respectively, whereas were no lead SNPs were identified in this population for (6R)-delta-6-carotene. This result was consistent with the contents and variance of carotenoids observed in the test analysis for the selected 24 accessions described in Section 6.3.1.

6.3.3 Seven carotenoid biosynthetic genes were identified in the mGWAS

mGWAS makes it possible to understand the genetic basis of the diversity in metabolites and the relevance of specific loci to the complex traits, through screening a very large number of accessions simultaneously (Riedelsheimer et al., 2012; Chen et al., 2014). Among the significant loci, carotenoid biosynthetic genes were mapped. Firstly, *SlPSY1*, encoding phytoene synthase 1, was co-located with two lead SNPs, 16 kb downstream of

Table 6.1 Summary of results from analysis of mGWAS in 154 accessions

Name	AU (nm)	RT	Lead SNPs	TFs	Structural genes	BEE complex
Lycopene	450	54.2	67	33	*ISPF or MCS*	*SlbHLH020*
					PSY1	
					ZISO	
Y- Carotene	450	48.1	15	1	n.a	n.a
β-Carotene	480.6	36.2	57	20	*DXS*	*SlbHLH020*
					ISPG or HDS	
					GGPPS	
					ZISO	
(6R)-delta-6-Carotene	450	42.31	0	0	n.a	n.a
Lutein	446.7	23.8	178	48	*GGPPS*	*SlHONG*
					HYDB	

AU: Absorption Units(nm); RT: Retention Time; TF: Transcription Factor.

sf0304309387 and 21 kb upstream of sf0304351704, and significantly associated with the content of lycopene (*P*= 3.06E-08 and *P*= 1.25E-06 respectively), which is in line with the key role of SlPSY1 in controlling lycopene biosynthesis (Fray and Grierson, 1993). Moreover, there were six other structural genes in the carotenoid biosynthetic pathway were co-located with SNPs associated with lycopene, β-carotene and lutein. As lycopene is the dominant carotenoid in the ripening tomato fruit, there were another two genes, *SlISPF* (*Solyc08g081570*), and *SlZISO* (*Solyc12g098710*) which mapped to loci affecting lycopene accumulation. It has been reported that the major QTGs (Quantitative trait genes) revealed by mGWAS are closely associated with the structure of screening population. In other words, the metabolic properties of certain samples within the population could contribute substantially to the mapping results. Although the majority of the accessions analysed were red-fruited lines, an orange-fruited wild species, *S. cheesmaniae*, was also included. Its high accumulation of β-carotene in fruit contributed to the sensitivity of compounds detected. Furthermore, green-fruit traits were also covered by the inclusion of *S. peruvianum* in the population. Due to the elevated diversity of carotenoids of the population, three genes, *SlDXS* (*Solyc11g010850*), *SlGGPPS* (*Solyc11g011240*) and *SlZISO* (*Solyc12g098710*) were identified as QTGs for β-carotene levels. *SlHYDB* (*Solyc04g051190*), encoding a β-carotene hydroxylase, taking part in the formation of lutein and apocarotenoids, such as zeaxanthin (Galpaz, 2006), was significantly associated with the lutein content, located 92 kb downstream of SNP sf0449954079 with a very low P value, 9.09E-10. Interestingly, *SlHYDB* was also identified in the eQTL mapping in Chapter 3, for which the expression level was more than 90% reduced in IL4-2, IL4-3 and IL 4-3-2 compared to M82, making these three ILs strong *cis*-eQTLs of *SlHYDB* (Figure 3.6).

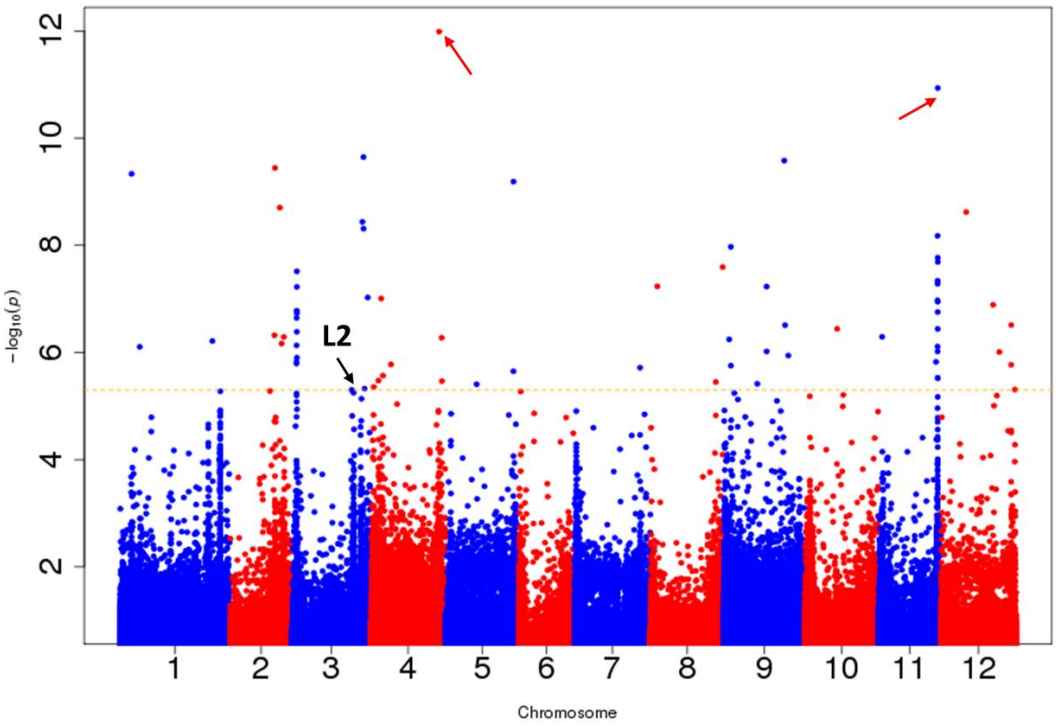
6.3.4 *SlHONG* and its homolog *SlbHLH020* (*Solyc03g034000*) were identified as candidates for two major QTGs associated with carotenoid accumulation.

Apart from biosynthetic genes identified with lead SNPs in the mGWAS screening, a wide range of transcription factors were also identified as candidate regulators. Based on the contents of lycopene, β-carotene and lutein, 33, 20 and 48 transcription factors lay within ± 100 kb of the lead SNP respectively, among which ten transcription factor candidates were

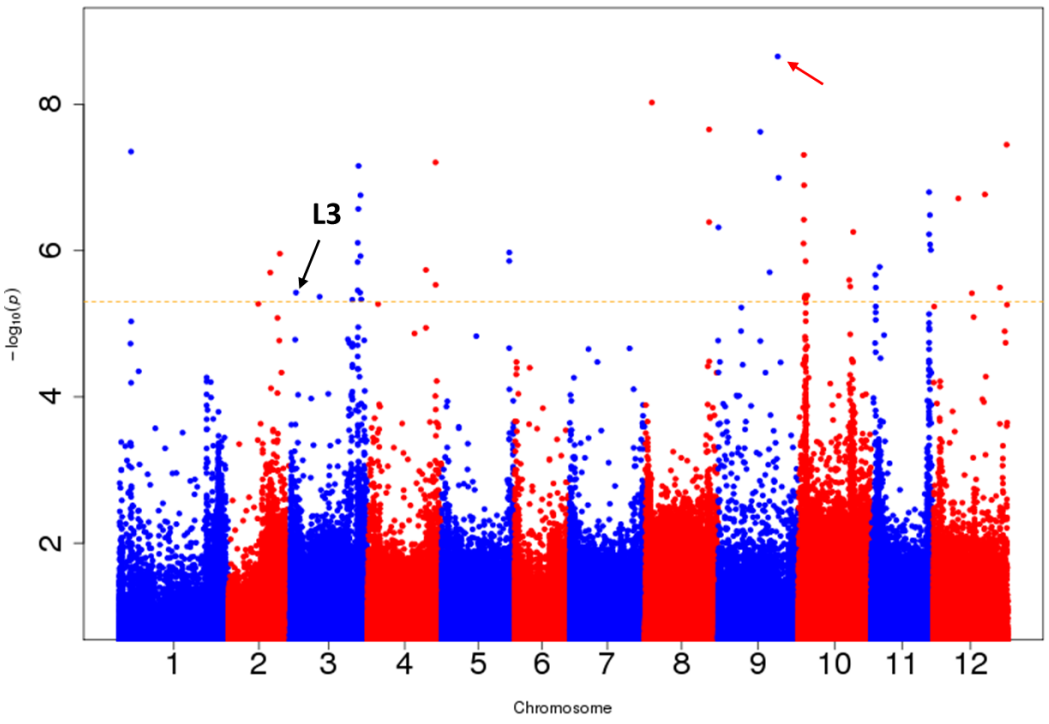
co-located by two or more carotenoids. For example, *Solyc05g050790*, *Solyc05g050830* and *Solyc09g059510* were found associated with variation in all the three major carotenoids in tomato fruit, making them strong candidates, worthy of further functional characterisation. Ethylene plays a crucial role in tomato fruit ripening, through which many transcription factors regulate carotenoid accumulation (Vrebalov et al., 2002; Chung et al., 2010; Liu et al., 2014b). Several transcription factors involved in ethylene signalling were identified in the screening, such as ETR4, ETR.2a and EIN3, confirming the regulatory role of ethylene in carotenoid biosynthesis.

Mining of these loci revealed that 16 were associated with genes encoding basic helix-loop-helix transcription factors. *SlHONG* was found located 76 kb downstream of a lead SNP, sf0234335151, and significantly associated with the level of lutein (*P*= 4.78E-06), suggesting SlHONG may play a role in lutein accumulation in fruit (Figure 6.4 C, L1). However, *SlHONG* was not detected as a QTG associated with lycopene levels, which may be because of the relatively low variability in lycopene contents within the population analysed. A model was proposed in Chapter 5 that SlHONG might work with another three bHLH proteins to form a complex involved in BR signalling in the same way as the BEE complex in Arabidopsis, to regulate carotenoid biosynthesis in tomato fruit. To understand the possible regulatory roles of the complex in carotenoid accumulation by mGWAS, we explore the candidate gene associated significant SNPs. This investigation revealed that besides *SlHONG*, its homolog *SlbHLH020* (*Solyc03g034000*) was significantly associated with lycopene and β-carotene accumulation, for which the *P* values were 3.73E-06 and 4.78E-06 respectively, co-localized with two lead SNPs, sf0305670883 and sf0234335151 (Figrue 6.4 A, L2; Figure 6.3 B L3). Taking the detection of *SlHONG* by mGWAS into consideration, as well as the interactions observed among the components within the complex (Chapter 5, Figure 5.17), the mapping of *SlbHLH020* by two major carotenoids supported the idea that different components of the BEE complex in tomato fruit may play important roles in regulation of the transcription of genes involved in carotenoid biosynthesis.

A


L2
−log10(p)
2
4
6
8
10
12
1
2
3
4
5
6
7
8
9
10
11
12
Chromosome

B


L3
−log10(p)
2
4
6
8
1
2
3
4
5
6
7
8
9
10
11
12
Chromosome

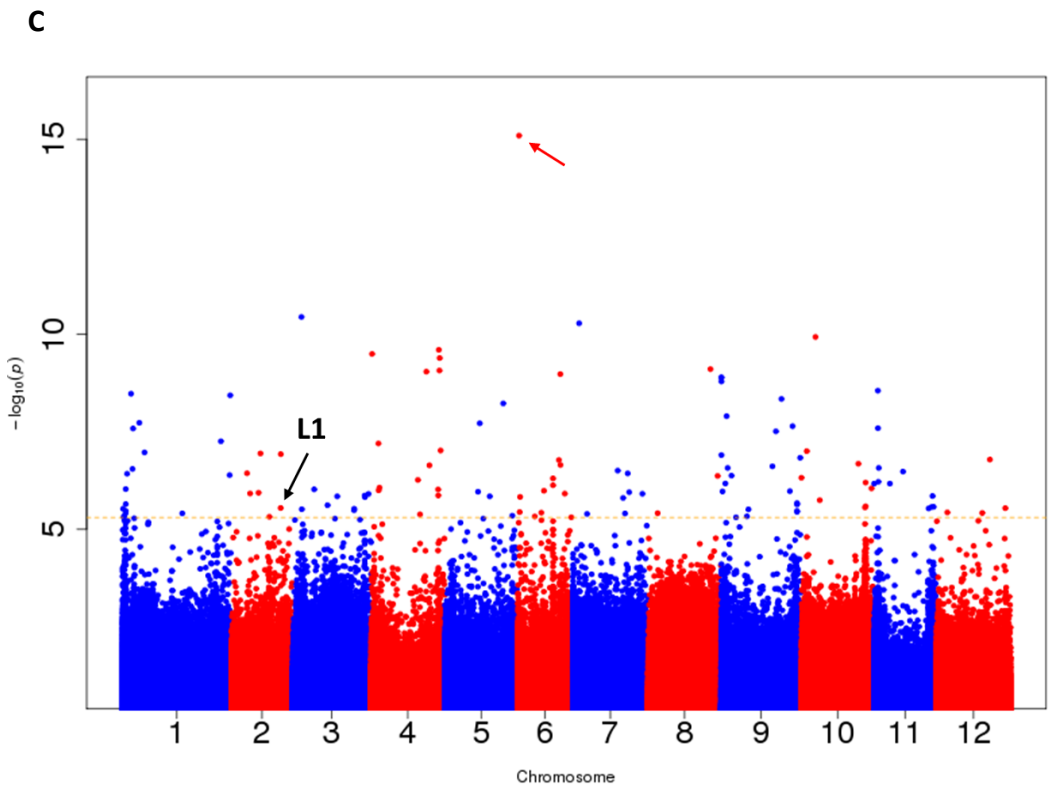


Figure 6.4 Manhattan plots displaying the GWAS results for the content of three major carotenoids. The horizontal dashed line indicates the thresholds set to P = 1.44E-6 by the LMM. A: lycopene; B: β-carotene; C: lutein. Red arrows indicate lead SNPs with significantly low *P* values. L1, L2 and L3 indicate significant SNPs sf0234335151 sf0305670883 and sf0234335151.

6.4 Discussion

Recently, the combination of high-throughput sequencing and advanced metabolomic profiling methods has offered a very comprehensive way to understand the natural variation in metabolism and the genetic basis for its regulation. GWAS have been conducted for the mapping of genetic loci contributing to important metabolic traits (Sauvage et al., 2014). A total of 154 accessions of tomato and its wild relatives were analysed as a natural population in a targeted metabolite-based genome-wide association study to elucidate the genetic basis of the regulation underlying carotenoid biosynthesis in tomato fruit. Seven structural genes were detected associated with three major carotenoids in ripening fruit: lycopene, β-carotene and lutein. *SlPSY1*, which is regarded as the key step in the lycopene biosynthetic pathway in tomato fruit, was detected with *P* = 3.06E-08. Another gene involved in lycopene biosynthesis, *SlZISO*, was also identified associated with the levels of lycopene and β-carotene. The identification of these two structural genes verified the use of targeted mGWAS for investigating the genetic basis of the carotenoid metabolic pathway, as has been used in other studies (Schwahn et al., 2014). The power of mGWAS is largely dependent on the properties of the metabolites of interest and the structure of the mapping population. Based on the complexity and diversity of different metabolic pathways, the difficulties in precisely detecting the related metabolites vary a lot, which affect the accuracy and efficiency of mGWAS directly, and are also closely related the plant species of interest. For example, anthocyanin biosynthetic pathway and carotenoid biosynthetic pathway are very well characterised in tomato and Arabidopsis, for which targeted mGWAS is sufficient to dissect the genetic basis as well as the regulatory mechanism. However, the situation could be far more complicated of other metabolic pathway or the studies in other plant species. Apart from the properties of the metabolites of interest, the structure of the mapping population also largely contribute to mapping result. There are two kinds of populations regularly used in mGWAS, those are natural population and structured population (or synthesised population), with distinguished methods applied. In general, genetic variation in a natural population is the result of several alleles with diverse effects on the phenotype of interest, for which the non-targeted mGWAS is normally used to capture the variance among the population. Whereas the structured population is established to study the differences between parental lines, such as the introgression

population used in this thesis, targeted mGWAS will be more efficient and direct, but with an increased risk of missing some information delivered by the population. The specific components of each kind of mapping population may also make a difference in the results dependent on the metabolitic pathway of interest.

The accumulation of carotenoids is a highly dynamic process closely related to fruit ripening, and is also influenced by a wide range of internal and external factors. The content of certain carotenoids, such as lycopene, varies between different species, even between accessions of the same species. Therefore, the structure of the screening population could affect the sensitivity and accuracy of the results obtained by mGWAS. In this study, only two non-red fruit species, *S. cheesmaniae* and *S. peruvianum* were included, while all the rest of the accessions produced lycopene in fruit when ripening. This relative lack of natural variation in the population for lycopene accumulation, may have impeded the identification of structural genes controlling lycopene biosynthesis, even if, for *SlPSY1*, the association was based on a relatively high *P* value (P_{LLM} = 3.06E-08). After thorough exploration of all the structural genes controlling carotenoid biosynthesis, as well as their paralogs, *SlDXS* (*Solyc11g010850*) and *SlGGPPS* (*Solyc11g011240*) were identified as key variable loci based on the contents of β-carotene and β-carotene and lutein respectively. However, among the three *DXS* genes and five *GGPS* genes annotated in tomato genome, the two paralogs identified by GWAS here are not those known to be functional in the fruit tissues. Indeed *SlDXS1* (*Solyc01g067890*) and *SlGGPP2* (*Soly04g079960*) are the functional gene, which was also verified by eQTL mapping as described in Chapter 3 (Lois et al., 2000; Paetzold et al., 2010; Ament et al., 2006). This limitation of mGWAS indicated that special attention should be paid to the composition of the mapping population when investigating the production of the metabolites with broad natural variation in accumulation, in a tissue- or stage-specific manner.

In analysis of mGWAS, metabolite levels are considered as quantitative traits with moderate inheritance. Although mGWAS has been proved to be a powerful tool to investigate the genetic basis of certain metabolic pathways in a range of plant species, it has limitations when applied to studying underlying transcriptional regulatory mechanisms. Unlike

complex traits such as grain yield and flowering time, which are controlled by many loci with small effects, levels of metabolites, especially secondary metabolites, are generally determined by a relatively small number of loci with relatively large effects (Chen et al., 2014). In general, transcription factors have broader but smaller effects in the regulation of metabolism, explaining why transcriptional regulators may be more difficult to characterise by mutant characterisation than structural genes. For example, according to previous analysis, *SlHONG* was not only expressed in fruits, but also in vegetative tissues (Figure 4.1). It appears to function as a positive regulator directly controling the transcript levels of biosynthetic genes in the lycopene biosynthetic pathway, as well as being involved in BR signaling in tomato plants, working together with three bHLH proteins, which may also influence carotenoid accumulation. It is very likely that SlHONG has pleiotropic effects throughout the growth and development of tomato plants. Although *SlHONG* and its homolog *SlbHLH020* were identified in the mGWAS analysis based on levels of selected carotenoids, the *P* value only just passed the threshold. Therefore, the effects caused by transcription factors like SlHONG may be masked by the large-effects of the polymorphisms underlying the structural genes in metabolite-based genome-wide associated studies. In addition, there may be many other regulatory steps influencing the final production of metabolites, such as post-transcriptional, translational and post-translational regulation, which could affect the sensitivity and accuracy of mGWAS in the identification of transcriptional regulators. In previous studies, there are several genes identified through targeted or untargeted mGWAS, most of which encode enzymes responsible for the key catalytic steps of the pathways (Zhu et al., 2018; Chen et al., 2014; Peng et al., 2016). Therefore, the genetic variation of these genes is more likely to cause significant changes in the contents of metabolites, which are easier to be identified compared to those with minor effects. In the context of carotenoid biosynthesis, many polymorphisms associated with structural genes were detected above the threshold, such as these for *SlPSY1*, *SlZISO* and *SlDXS* (Figure), as the biosynthetic genes have a more direct impact on the corresponding metabolite production. However, even if a transcriptional regulator could influence the transcript abundance of the structural genes, its effect on the phenotype in mGWAS, that is the metabolite level, which still could be covered by other regulatory process as well as the structure of the mapping population.

As discussed in Chapter 3, eQTL mapping is a sensitive method for the identification of transcriptional regulators through genome-wide analysis. In addition to using -omics approaches separately, transcriptome analysis combined with metabolic profiling could be used to study metabolic regulation in plants, and has already been used to analyse pheylpropanoid, flavonoid and steroidal glycoalkaloid biosynthesis (Alseekh et al., 2015; Cárdenas et al., 2016; Itkin et al., 2013). Six *trans*-eQTL hotspots for carotenoid biosynthesis were identified by eQTL analysis in Chapter 3 (Figure 3.7). In the mGWAS analysis, lead SNPs with significantly low *P* values were also characterised associated with the contents of lycopene, β-carotene and lutein; these loci are shown by red arrows in Figure 6.4. The combination of these two sets of data at genome scale offers more reliable datasets to identify further regulators controlling carotenoid biosynthesis in tomato fruit.

Chapter 7

General Discussion and Outlook

In this study, I used expression quantitative trait loci (eQTL) analysis to obtain a comprehensive understanding of the genetic basis of the regulation of carotenoid biosynthesis in tomato fruit using the *S. lycopersicum* x *S. pennellii* introgression population. In total, 31 *cis*-eQTLs related to 18 carotenoid biosynthetic genes were identified, and the paralogs of some structural genes functional in fruit were also identified in the analysis (Table 7.1). 6 *trans*-eQTL hotspots were identified for lycopene biosynthesis. Co-expression analysis of one of the *trans*-eQTL candidate, IL2-1, revealed that a basic helix-loop-helix transcription factor, SlHONG may act as a positive regulator of lycopene accumulation in fruit, by activating the expression of the genes encoding enzymes in the pathway. SlRIN, the master regulator of fruit ripening, bound directly to two adjacent CArG motifs in the promoter of *SlHONG*, to control the function of SlHONG in an ethylene-independent manner (Zhong et al., 2013; Liu et al., 2015b). Genome-scale phylogenetic analysis of bHLH proteins in Arabidopsis and tomato found SlHONG was the most similar protein to BEE2 in Arabidopsis, which is involved in brassinosteroid signalling through formation of the BEE complex with another three bHLH transcription factors. bHLH transcription factor candidates potentially involved in the tomato BEE complex with SlHONG, were identified by phylogenetic analysis and the interactions between these were analysed. In this study, I identified a regulator of lycopene biosynthesis in tomato fruit, SlHONG, and characterised the multi-layered mechanism underlying the transcriptional regulation of lycopene biosynthesis in tomato fruit.

7.1 Expression QTL (eQTL) analysis is a powerful approach for characterising the regulation of complex metabolic pathways.

Carotenoid biosynthesis in plants is a complex and multifaceted pathway and its products participate in a wide range of physiological processes. Due to its central roles in plant growth and development, many internal and environmental regulators are involved in the regulation of carotenoid production under various conditions. Over the past 30 years, the enzymes involved in carotenoid biosynthesis in tomato have been elucidated, step by step, based mainly on mutant identification (Fray and Grierson, 1993; Ronen et al., 2000b; Fraser et al., 1999; Isaacson, 2002b; Ronen et al., 1999). Amongst these publications none have

Table 7.1 *cis*-eQTLs identified as functioning in carotenoid metabolism in fruit in this thesis

Gene ID	Gene name abbreviated	Metabolic Pathway	*cis*-eQTL (s)		
Solyc01g067890	DXS1	Isoprenoid biosynthesis	IL1-1		
Solyc01g109300	HDR	Isoprenoid biosynthesis	IL1-4		
Solyc04g056390	IPP1	Isoprenoid biosynthesis	IL4-2	IL4-3	IL4-3-2
Solyc05g055760	IPP2	Isoprenoid biosynthesis	IL5-5		
Solyc04g079960	GGPPS2	Isoprenoid biosynthesis	IL4-3	IL4-4	
Solyc03g031860	PSY1	Lycopene biosynthesis	IL3-2.B		
Solyc02g081330	PSY2	Lycopene biosynthesis	IL2-3.B	IL2-4	IL2-5
Solyc12g098710	ZISO	Lycopene biosynthesis	IL12-3	IL12-4	
Solyc10g081650	CrtISO	Lycopene biosynthesis	IL10-2	IL10-2-2	
Solyc06g074240	CYCB	Lycopene catabolism	IL6-3		
Solyc10g079480	LCYB2	Lycopene catabolism	IL10-2	IL10-2-2	
Solyc12g008980	LCYE	Lycopene catabolism	IL12-2.A		
Solyc03g007960	CHY2	Lycopene catabolism	IL3-2.A		
Solyc01g087250	CCD1A	Lycopene catabolism	IL1-1		
Solyc01g087260	CCD1B	Lycopene catabolism	IL1-1		
Solyc08g075480	CCD4A	Lycopene catabolism	IL8-2-1	IL8-2	
Solyc08g075490	CCD4B	Lycopene catabolism	IL8-2-1	IL8-2	
Solyc07g056570	NCED	Lycopene catabolism	IL7-2	IL7-3	

Table 7.2 Transcription factors studied in this thesis

Gene ID	Gene name abbreviated	TF family	Identification	IL location
Solyc02g062690	HONG	Basic helix-loop-helix	eQTL	IL2-1
Solyc12g036470	bHLH071	Basic helix-loop-helix	Phylogenetic analysis	IL12-3 IL12-2.A
Solyc03g119390	bHLH026	Basic helix-loop-helix	Phylogenetic analysis	IL3-5
Solyc03g034000	bHLH020	Basic helix-loop-helix	Phylogenetic analysis	IL3-2
Solyc03g120530	n.a.	Helix-loop-helix	eQTL	IL3-5
Solyc09g083220	bHLH060	Basic helix-loop-helix	Ye et al., 2015	IL9-3, IL9-3-1 IL9-3-2
Solyc08g076820	bHLH0146	Basic helix-loop-helix	eQTL	IL8-2 IL8-2-1

reported the identification of more than one carotenoid biosynthetic gene in tomato, in a single study. This is mainly because the discovery of the mutants related to carotenoid biosynthesis has been dependent on the effects of each gene on carotenoid accumulation and this analysis is time-consuming. For this reason, *SlPSY1*, which plays a decisive role in lycopene accumulation in fruit and encodes an isoform of the key enzyme of lycopene biosynthesis, was the first structural gene identified (Fray and Grierson, 1993). Furthermore, the pleiotropic roles of carotenoids and their intermediates produced by the pathway, make it often difficult to link phenotypes to the underlying genes accurately. For example, because of its central role in terpenoid metabolism, the isoprenoid biosynthetic pathway supplies precursors not only for carotenoid production, but also for the synthesis of gibberellins, tocopherols, isoprenes and diterpenes. The functional redundancy of some of the genes involved in isoprenoid biosynthesis increases further the difficulties in identification and functional characterisation of structural genes active in tomato fruit. In addition, metabolic units were identified in this pathway, and a single mutant of one member within the unit may not cause significant changes in the production of the corresponding carotenoids (Fantini et al., 2013). The function of *SlZISO* was not confirmed until 2013, because it works together with *SlPDS* as a single unit in lycopene biosynthesis, catalysing the synthesis of di-*cis*-ζ-carotene. The single mutant of *SlZISO* does not show a visible phenotype (Fantini et al., 2013).

In my study, the advantages of eQTL analysis in the characterisation of the regulation of complex metabolic pathways were demonstrated clearly. eQTL analysis using RNA-seq data from fruit of the *S. lycopersicum* x *S. pennellii* IL population, identified 31 *cis*-eQTLs encoding 18 structural genes involved in carotenoid biosynthesis in tomato fruit, including *SlPSY1*, *SlCrtISO*, *SlLCYE* and *SlCYCB* (Chapter 3, Figure 3.5 C, Figure 3.6 C), which were identified originally by mutant discovery, as well as *SlPDS* mentioned above (Figure 3.5 C). Furthermore, expression profiling following eQTL mapping was able to confirm the identity of functional isoforms operating in a tissue-specific manner; for example, *SlGGPPS2* was identified as functional in isoprenoid biosynthesis in fruit out of five genes encoding isoforms of GGPP (Chapter 3, Figure 3.4 C). This method is, of course, dependent on their being large differences in gene expression between the parents of the IL population, which

was the case in the *S. lycopersicum* x *S. pennellii* population because *S. pennellii* is green-fruited and produces no lycopene in its fruit.

Apart from the elucidation of the genetic foundation of a metabolic pathway, eQTL analysis can also be used for the identification of transcriptional regulators of secondary metabolic pathways. Several transcription factors have been identified that are involved in the regulation of carotenoid biosynthesis in fruit (Liu et al., 2014b; Chung et al., 2010; Toledo-Ortiz et al., 2010; Bou-Torrent et al., 2015; Powell et al., 2012). However, all of these transcription factors have broad effects coupled with other pathways, such as ethylene signalling, photomorphogenesis, and plastid formation. This is because carotenoid accumulation in tomato fruit is a dynamic process which is closely related to fruit development and ripening and is affected by diverse internal and external signals. Furthermore, any transcription factors which work through direct binding to the promoter of structural genes, likely regulate the transcript levels of only a very few genes. For example, SlRIN, a MADS transcription factor which acts as a master regulator of fruit ripening, directly activate the expression of *SlPSY1* to regulate the lycopene accumulation in fruit, but no other carotenoid biosynthetic genes have been identified as direct targets of SlRIN (Fujisawa et al., 2013). SlPIF1, a helix-loop-helix transcription factor involved in light-related responses, binds to the promoter of only *SlPSY1* (Llorente et al., 2016). eQTL analysis offers an approach to allow discovery of the regulators of a metabolic pathway, by the identification and characterisation of *trans*-eQTL hotspots. In this study, six *trans*-acting eQTL hotspots were identified for lycopene biosynthesis (Figure 3.7). SlHONG was identified in one of the trans-eQTL hotspots, IL2-1, where transcript abundance of most of the lycopene biosynthetic genes was positively associated with lycopene production (Figure 3.11). The activity of SlHONG was confirmed further by VIGS (Figure 4.8, Figure 4.9) and stable transformation (Figure 4.21). Direct binding between SlHONG and the promoters of genes involved in lycopene biosynthesis was detected (Figure 4.7). Therefore, SlHONG was identified as a regulator of lycopene biosynthesis in tomato fruit by eQTL analysis, an identification which was unlikely to have been achieved by other approaches.

A few factors could influence the accuracy and efficiency of eQTL analysis in the study of transcriptional regulation of secondary metabolism. Firstly, a clearly-elucidated pathway is

the necessary foundation to identify transcriptional regulators by eQTL analysis. The identification of SlHONG was based on understanding the expression of the well-studied genes encoding enzymes of the lycopene biosynthetic pathway. Secondly, for a complicated, multiply-branched metabolic pathway, proper selection of likely regulons in the pathway can facilitate analysis, such as for carotenoid biosynthesis, which is composed of more than 40 catalytic steps. It would be very difficult to develop any understanding of the underlying regulatory networks if this pathway were analysed in its entirety. Thirdly, the structure of the mapping population plays an important role in eQTL analysis. In this study, the *S. lycpersicum* x *S. pennellii* introgression population was used. One of its parental lines, *S. pennellii* does not make lycopene in its fruit, producing green fruit, with no lycopene, when ripe, making it a suitable population for establishing functionality in the regulation of carotenoid biosynthesis (Figure 3.1). According to the properties of the pathway of interest, the time point for sampling the tissues used for transcriptomic profiling can affect the efficiency and accuracy of eQTL analysis. The lycopene biosynthetic genes exhibit very big changes in expression during fruit ripening: most genes are expressed at very low levels until the breaker stage, increase to a maximum approximately three days later and then decrease (Figure 4.3 B-G). One concern about this introgression population is that the fruits used for the RNA-seq were harvested when 80 to 100% of M82 tomatoes, (one of the parental lines) were red, which might lead to differences in the stage of ripening among the ILs at harvest. The transcriptomic dataset might have been more informative if a specific time point had been selected to capture the biggest effects of each transcription factor candidate on the expression levels of structural genes, such as B+3 in this study.

7.2 The identification of SlHONG as a positive regulator of lycopene biosynthesis reveals an ethylene-independent regulatory mechanism

Carotenoid production in tomato fruit is a process dependent on fruit development. Lycopene accumulation initiates at the start of fruit ripening, and is entirely dependent on ethylene production and perception. Most of the transcription factors already identified as regulating carotenoid biosynthesis in tomato fruit are involved in ethylene signalling,

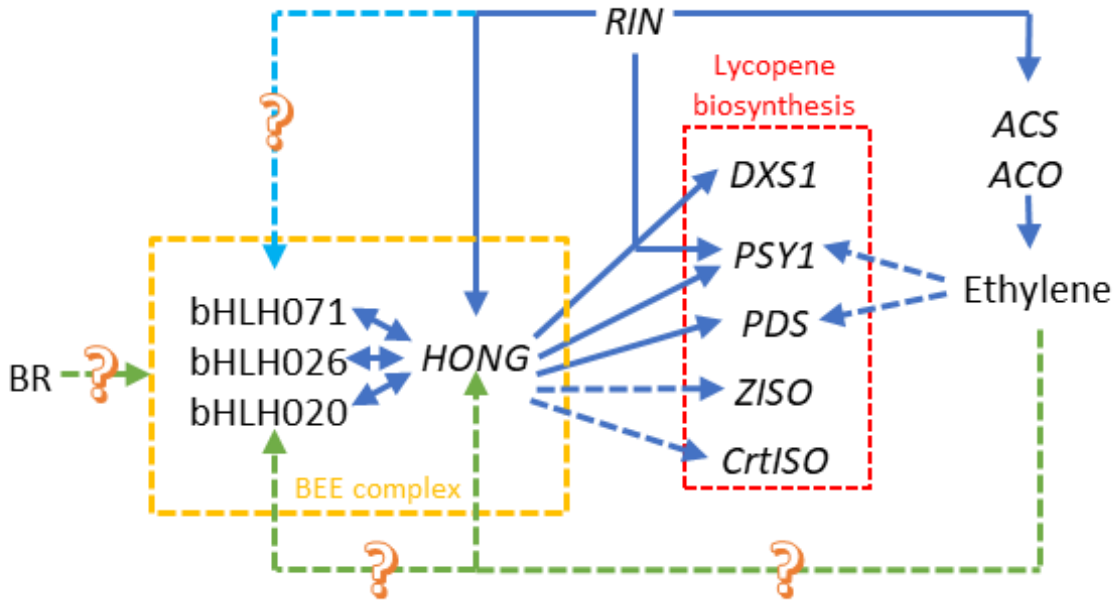


Figure 7.1 Schematic view of the proposed model for transcriptional regulation of lycopene biosynthesis developed in this study. Solid arrows, direct regulations. Dashed arrows, indirect regulations.

such as SlAP2a, SlERF6, SlERF.B3 and SlERF3 (Liu et al., 2014b; Chung et al., 2010; Lee et al., 2012; Lanahan, 1994). SlRIN, a master regulator of fruit ripening, plays a positive role in lycopene accumulation, primarily through its regulatory role in ethylene production where it interacts directly with the promoters of ethylene biosynthetic genes, *SlACS2* and *SlACS4* (Fujisawa et al., 2013, 2012) (Figure 7.1).

In my study, SlHONG was identified as a positive regulator of lycopene biosynthesis by eQTL analysis, and this was confirmed by transient and stable transformation experiments. The transactivation assays showed that SlHONG regulates the expression of carotenoid biosynthetic genes through direct transcriptional activation to their promoters, including *SlDXS1*, *SlPSY1* and *SlPDS* (Figure 5.4). Furthermore, *SlHONG* is regulated by the MADS-domain transcription factor, RIN, through a CArG motif in the promoter region of *SlHONG* (Figure 5.9, Figure 5.10, Zhong et al., 2013). This indicates that SlHONG links the transcriptional regulation of the lycopene biosynthetic pathway to SlRIN in an ethylene-independent manner (Figure 7.1) This is the first example which shows that the role of RIN in the regulation of lycopene biosynthesis in fruit can be uncoupled from ethylene signalling.

SpHONG, the orthologue of SlHONG in the green-fruited species, *S. pennellii*, has an identical bHLH domain to that of SlHONG, and very similar transcriptional activation abilities to SlHONG (Figure 5.7). Both of the CArG motifs found in the promoter of SlHONG were missing due to an A to G SNP in the SpHONG promoter (Figure 5.12). These differences in the presence of RIN binding sites may contribute to the distinct expression of the structural genes involved in lycopene biosynthesis, between *S. lycpersicum* and *S. pennellii*, even though both species can produce functional HONG proteins. Differences in the presence of CArG boxes in the promoters of *HONG* were also found in the analysis of other wild species in the *Lycopersicum* clade. In *S. harbrochites*, which is a green-fruited species, both CArG motifs were missing, caused by a C to T SNP (Figure 5.13). In another green-fruited species, *S. lycopersicoides*, one of the adjacent CArG boxes was missing while the other one was present, suggesting possible differences in the importance of the two adjacent RIN binding motifs (Figure 5.13). *S. pimpinellifolium* had the same CArG-boxes as *S. lycopersicum*, consistent with the accumulation of lycopene in fruit of this species (Figure 5.13, 5.14).

Changes in *cis*-elements, such as transcription factor binding motifs, can result in changes in gene expression patterns without distruption of their coding sequences or in their activity. The divergence in the promoters of *HONG* in tomato and its wild species may have played a role in the evolution of lycopene accumulation in fruits.

7.3 Brassinosteroids (BRs) play a role in the regulation of carotenoid biosynthesis in tomato fruit through SlHONG and the BEE complex

Brassinosteroids are a group of steroid compounds in plants, which participate in a wide range of physiological processes, regulating different aspects of plant growth, development and reproduction. It has been shown that brassinosteroids regulate carotenoid accumulation in tomato fruit (Vardhini and Rao, 2002), and increased brassinosteroid levels can elevate carotenoid production as well as the expression of the structural genes (Liu et al., 2014a; Nie et al., 2017; Zhu et al., 2015). Several key components involved in brassinosteroid biosynthesis and signalling have been characterised (Holton et al., 2007; Jones et al., 2002; Koka et al., 2000; Bishop et al., 1999). Dwarf, encodes a CYP85A1 P450 protein which catalyses two steps in BR biosynthesis involving C-6 oxidation, and was identified through the d^x mutant which is defective in BR biosynthesis (Bishop et al., 1999). The expression pattern of *Dwarf* reveals its important role in flower and fruit development (Montoya et al., 2005). A *dwarf* knock-out mutant has been generated by Javier Galdon Armero in the Martin Lab through CRISPR/Cas9 mediated genome editing. The *dwarf* knock-out plant was much shorter than wild type (Figure 7.2 A). The flowers of the mutant had branched anthers and were also much smaller than wild type (Figure 7.2 C & D). Many pollinated flowers failed to set fruit (Figure 7.2 E). Differences in leaf morphology were observed as well (Figure 7.2 B). Most interestingly, fruit ripening was greatly delayed in the *dwarf* knock-out mutant, such that the fruit started to ripen 70 days after anthesis (Figure 7.2 F) compared to 45days after anthesis in MoneyMaker (España et al., 2014). However fruit size in the *dwarf* knock-out plant was normal compared to wild type and no defects in lycopene accumulation were observed in ripe fruits. The phenotypes displayed in d^x mutant and *dwarf* knock-out

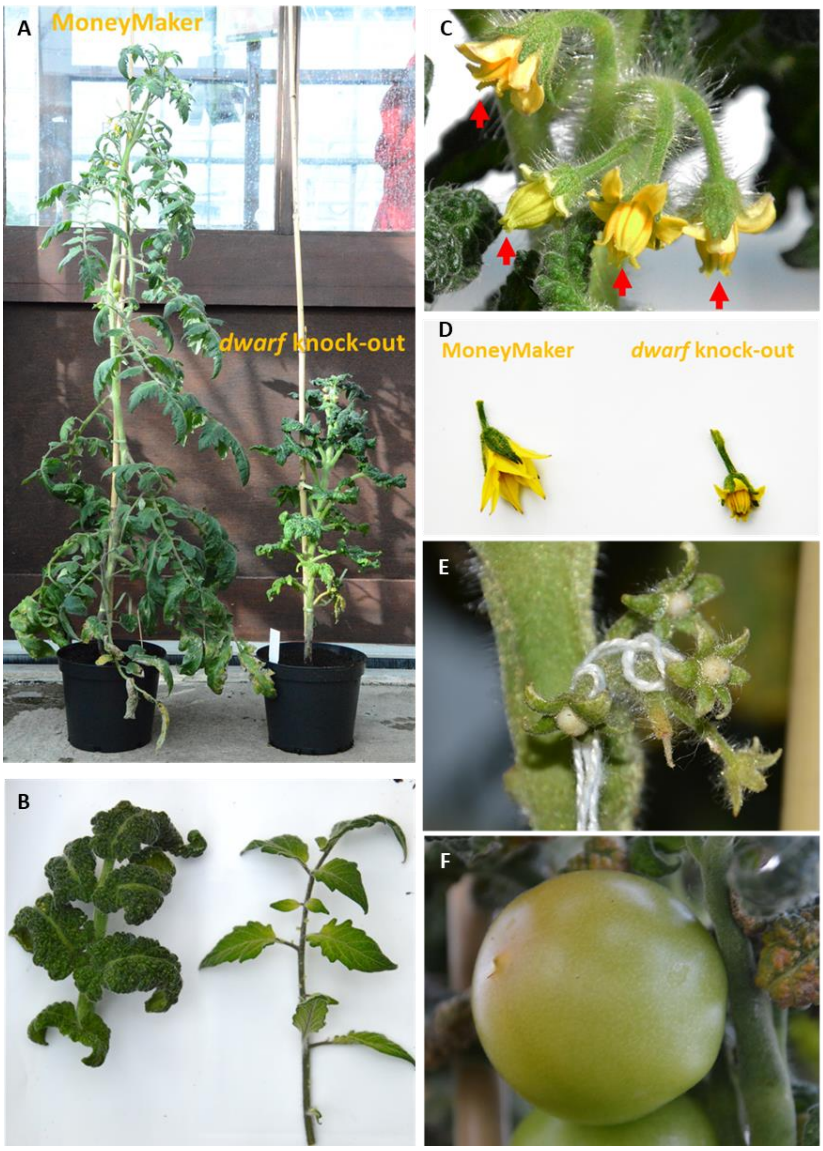


Figure 7.2 Phenotypes of *dwarf* knock-out mutant. A, Photograph of wild-type Money Maker plant and *dwarf* knock-out mutant. B, Photograph of leaves of wild-type Money Maker plant (right) and *dwarf* knock-out mutant (left). C, Flowers of the *dwarf* knock-out mutant have branched anthers. D, Photograph of flowers of wild-type Money Maker plant and the *dwarf* knock-out mutant which had much shorter petals. E, Fruits of *dwarf* knock-out mutant at 70 days after anthesis. Many fruit did not develop any further. F, Fruit of *dwarf* knock-out mutant which reached breaker stage at 70 days after anthesis compared to 45 days after anthesis for a wild type MoneyMaker plant. The *dwarf* CRISPR/Cad9 knock-out muant was kindly provided by Mr Javier Galdon Armero, John Innes Centre, UK.

mutants indicated that brassinosteroids have important, pleiotropic roles in plant growth, development and fruit maturation. A genome-wide phylogenetic analysis of basic helix-loop-helix transcription factors in Arabidopsis and tomato revealed that SlHONG was most similar to two Arabidopsis bHLH proteins, BEE2 (AT4G36540.1) and HBI1 (AT2G18300.3) (Figure 5.15). In Arabidopsis, BEE1 (AT1GT18400.1), BEE2 (AT4G36540.1), BEE3 (AT1G73830.1), and CESTA, another bHLH protein (AT1G25330.1), interact with BEE1, to form a complex that functions as a positive regulator of brassinosteroid signalling (Friedrichsen et al., 2002; Poppenberger et al., 2011). Another three bHLH proteins were identified by phylogenetic analysis as likely to form the BEE complex with SlHONG in tomato, SlbHLH071, SlbHLH06 and SlbHLH020 (Figure 5.15). The interactions among these four bHLH proteins were supported by yeast two-hybrid assays, suggesting that SlHONG plays a crucial role in forming the BEE complex in tomato (Figure 5.17). Because SlHONG might regulate lycopene biosynthesis in tomato fruit directly, brassinosteroids may play a role in the regulation of carotenoid biosynthesis in tomato fruit through SlHONG and the BEE complex, providing an explanation for the effects of BRs on carotenoid accumulation in tomato.

7.4 A model of transcriptional regulation underlying lycopene biosynthesis with SlHONG as a central component

Based on my study and the knowledge of current regulatory mechanisms modulating carotenoid accumulation in tomato fruit, I propose a model for the transcriptional regulation controlling carotenoid biosynthesis with SlHONG as a core component (Figure 7.1). In this model, SlHONG activates expression of the genes involved in lycopene biosynthesis, *SlDXS1*, *SlPSY1* and *SlPDS*, through direct transcriptional activation. The levels of *SlZISO* and *SlCrtISO* transcripts were also positively associated with those of *SlHONG*. Therefore, SlHONG acts as a direct regulator of lycopene biosynthesis in fruit. The expression of *SlHONG* is controlled by different factors at different levels. *SlHONG* is regulated by the MADS transcription factor, SlRIN, by direct transcriptional activation of two adjacent CArG motifs in the promoter of *SlHONG*, in a mechanism independent of ethylene signalling. Furthermore,

SlHONG interacts with another three bHLH proteins to form the BEE complex in tomato, which may be regulated by BR as it is in Arabidopsis. It is likely that SlHONG activates transcription of carotenoid biosynthetic genes as part of a complex of bHLH proteins, which normally serve in responses to BR signals. Due to the distinct expression patterns of the members of the BEE complex, the complex may achieve its regulatory activity in carotenoid biosynthesis in a tissue- and/or development dependent manner through the dynamic changes in transcript abundance of the different proteins within the complex.

There are many questions needing answers to complete the model. I have shown that SlHONG is regulated by SlRIN in an ethylene-independent way. The response of SlHONG to ethylene needs to be investigated as well as whether ethylene can regulate the other three bHLH transcription factors. SlbHLH071, SlbHLH06 and SlbHLH020 likely participate in the BEE complex. In Arabidopsis, the BEE complex serves a positive role in BR signalling and directly responds to BR levels. It needs to be tested whether the BEE complex in tomato functions in a similar way, and how the function of the complex in tomato fruit differs from its function in vegetative tissues. By addressing these questions, we will be able to develop a more comprehensive understanding of the transcriptional regulatory networks controlling carotenoid biosynthesis in tomato fruit. Of course, SlHONG probably participates in other biological processes in other tissues, and Δ*SlHONG* knock-out mutants displayed significantly shorter hypocotyls compared to wildtype plants, consistent with the phenotype observed in mutants of *BEE2* in Arabidopsis (Friedrichsen et al., 2002; Malinovsky et al., 2014). In addition, the identification of regulatory mechanisms, particularly the transcriptional control of biosynthesis of phytonutrients, such as carotenoids in this study, may facilitate metabolic engineering of nutritionally improved tomato varieties in the future.

References

Adams-Phillips, L., Barry, C., Kannan, P., Leclercq, J., Bouzayen, M., and Giovannoni, J. (2004). Evidence that CTR1-mediated ethylene signal transduction in tomato is encoded by a multigene family whose members display distinct regulatory features. Plant Mol. Biol. **54**: 387–404.

Adato, A. et al. (2009). Fruit-surface flavonoid accumulation in tomato is controlled by a SLMYB12-regulated transcriptional network. PLoS Genet. **5**: 12.

Albrecht, M., Takaichi, S., Steiger, S., Wang, Z.Y., and Sandmann, G. (2000). Novel hydroxycarotenoids with improved antioxidative properties produced by gene combination in Escherichia coli. Nat. Biotechnol. **18**: 843–846.

Alexander, L. (2002). Ethylene biosynthesis and action in tomato: a model for climacteric fruit ripening. J. Exp. Bot. **53**: 2039–2055.

Alseekh, S. et al. (2015). Identification and Mode of Inheritance of Quantitative Trait Loci for Secondary Metabolite Abundance in Tomato. Plant Cell **27**: 485–512.

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

Ament, K., Van Schie, C.C., Bouwmeester, H.J., Haring, M.A., and Schuurink, R.C. (2006). Induction of a leaf specific geranylgeranyl pyrophosphate synthase and emission of (E,E)-4,8,12-trimethyltrideca-1,3,7,11-tetraene in tomato are dependent on both jasmonic acid and salicylic acid signaling pathways. Planta **224**: 1197–1208.

Anderson, S.M. and Krinsky, N.I. (1973). PROTECTIVE ACTION OF CAROTENOID PIGMENTS AGAINST PHOTODYNAMIC DAMAGE TO LIPOSOMES*. Photochem. Photobiol. **18**: 403–408.

Angelovici, R., Lipka, A.E., Deason, N., Gonzalez-Jorge, S., Lin, H., Cepela, J., Buell, R., Gore, M. a, and DellaPenna, D. (2013). Genome-Wide Analysis of Branched-Chain Amino Acid Levels in Arabidopsis Seeds. Plant Cell Online **25**: 4827–4843.

Asíns, M.J. (2002). Review Present and future of quantitative trait locus analysis in plant

breeding. Plant Breed. **121**: 281–291.

Atchley, W.R., Terhalle, W., and Dress, A. (1999). Positional dependence, cliques, and predictive motifs in the bHLH protein domain. J. Mol. Evol. **48**: 501–516.

Aust, O., Stahl, W., Sies, H., Tronnier, H., and Heinrich, U. (2005). Supplementation with tomato-based products increases lycopene, phytofluene, and phytoene levels in human serum and protects against UV-light-induced erythema. Int. J. Vitam. Nutr. Res. **75**: 54–60.

Bai, M.-Y., Fan, M., Oh, E., and Wang, Z.-Y. (2012). A Triple Helix-Loop-Helix/Basic Helix-Loop-Helix Cascade Controls Cell Elongation Downstream of Multiple Hormonal and Environmental Signaling Pathways in Arabidopsis. Plant Cell **24**: 4917–4929.

Bailey, P.C., Martin, C., Toledo-Ortiz, G., Quail, P.H., Huq, E., Heim, M.A., Jakoby, M., Werber, M., and Weisshaar, B. (2003). Update on the basic helix-loop-helix transcription factor gene family in Arabidopsis thaliana. Plant Cell Online **15**: 2497.

Ballester, A.-R. et al. (2010). Biochemical and Molecular Analysis of Pink Tomatoes: Deregulated Expression of the Gene Encoding Transcription Factor SlMYB12 Leads to Pink Tomato Fruit Color. PLANT Physiol. **152**: 71–84.

Barickman, T.C., Kopsell, D.A., and Sams, C.E. (2014). Abscisic acid increases carotenoid and chlorophyll concentrations in leaves and fruit of two tomato genotypes. J. Am. Soc. Hortic. Sci. **139**: 261–266.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science (80-.). **315**: 1709–1712.

Barry, C.S., Blume, B., Bouzayen, M., Cooper, W., Hamilton, A.J., and Grierson, D. (1996). Differential expression of the 1-aminocyclopropane-1-carboxylate oxidase gene family of tomato. Plant J. **9**: 525–535.

Barry, C.S. and Giovannoni, J.J. (2006). From The Cover: Ripening in the tomato Green-ripe mutant is inhibited by ectopic expression of a protein that disrupts ethylene signaling. Proc. Natl. Acad. Sci. **103**: 7923–7928.

Barry, C.S., Llop-Tous, M.I., and Grierson, D. (2000). The Regulation of 1-Aminocyclopropane-1-Carboxylic Acid Synthase Gene Expression during the Transition from System-1 to System-2 Ethylene Synthesis in Tomato. Plant Physiol. **123**: 979–986.

Bartlett, M.E. (2017). Changing MADS-Box transcription factor protein-protein interactions as a mechanism for generating floral morphological diversity. In Integrative and Comparative Biology, pp. 1312–1321.

Bartley, G.E. and Scolnik, P.A. (1993). cDNA cloning, expression during development, and genome mapping of PSY2, a second tomato gene encoding phytoene synthase. J. Biol. Chem. **268**: 25718–25721.

Baudry, A., Heim, M.A., Dubreucq, B., Caboche, M., Weisshaar, B., and Lepiniec, L. (2004). TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in Arabidopsis thaliana. Plant J. **39**: 366–380.

Baulcombe, D.C. (1999). Fast forward genetics based on virus-induced gene silencing. Curr. Opin. Plant Biol. **2**: 109–113.

Beck, J.J., Smith, L., and Baig, N. (2014). An overview of plant volatile metabolomics, sample treatment and reporting considerations with emphasis on mechanical damage and biological control of weeds. Phytochem. Anal. **25**: 331–341.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., and Nekrasov, V. (2013). Plant genome editing made easy: Targeted mutagenesis in model and crop plants using the CRISPR/Cas system. Plant Methods **9**: 39.

Beyer, P., Nievelstein, V., Albabili, S., Bonk, M., and Kleinig, H. (1994). Biochemical aspects of carotene desaturation and cyclization in chromoplast membranes from Narcissus-pseudonarcissus. Pure Appl. Chem. **66**: 1047–1056.

Bhuvaneswari, V., Velmurugan, B., Balasenthil, S., Ramachandran, C.R., and Nagini, S. (2001). Chemopreventive efficacy of lycopene on 7,12-dimethylbenz[a]anthracene-induced hamster buccal pouch carcinogenesis. Fitoterapia **72**: 865–874.

Bishop, G.J., Nomura, T., Yokota, T., Harrison, K., Noguchi, T., Fujioka, S., Takatsuto, S., Jones, J.D., and Kamiya, Y. (1999). The tomato DWARF enzyme catalyses C-6 oxidation in brassinosteroid biosynthesis. Proc. Natl. Acad. Sci. U. S. A. **96**: 1761–1766.

Bjelakovic, G., Nikolova, D., Gluud, L.L., Simonetti, R.G., and Gluud, C. (2015). Antioxidant supplements for prevention of mortality in healthy participants and patients with various diseases. Sao Paulo Med. J. **133**: 164.

Bohs, L. (2007). Phylogeny of the Cyphomandra clade of the genus Solanum (Solanaceae) based on ITS sequence data. Taxon **56**: 1012–1026.

Bolger, A. et al. (2014). The genome of the stress-tolerant wild tomato species Solanum pennellii. Nat. Genet. **46**: 1034–1038.

Borevitz, J.O., Xia, Y., Blount, J., Dixon, R.A., and Lamb, C. (2000). Activation Tagging Identifies a Conserved MYB Regulator of Phenylpropanoid Biosynthesis. Plant Cell **12**: 2383.

Botella-Pavía, P., Besumbes, Ó., Phillips, M.A., Carretero-Paulet, L., Boronat, A., and Rodríguez-Concepción, M. (2004). Regulation of carotenoid biosynthesis in plants: Evidence for a key role of hydroxymethylbutenyl diphosphate reductase in controlling the supply of plastidial isoprenoid precursors. Plant J. **40**: 188–199.

Bou-Torrent, J., Toledo-Ortiz, G., Ortiz-Alcaide, M., Cifuentes-Esquivel, N., Halliday, K.J., Martinez-Garcia, J.F., and Rodriguez-Concepcion, M. (2015). Regulation of carotenoid biosynthesis by shade relies on specific subsets of antagonistic transcription factors and co-factors. Plant Physiol.: pp.00552.2015.

Bowman, J.L. et al. (2017). Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome. Cell **171**: 287–304.e15.

Brandi, F., Bar, E., Mourgues, F., Horváth, G., Turcsi, E., Giuliano, G., Liverani, A., Tartarini, S., Lewinsohn, E., and Rosati, C. (2011). Study of "Redhaven" peach and its white-fleshed mutant suggests a key role of CCD4 carotenoid dioxygenase in carotenoid and norisoprenoid volatile metabolism. BMC Plant Biol. **11**: 24.

Brem, R.B. and Kruglyak, L. (2005). The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proc. Natl. Acad. Sci. **102**: 1572–1577.

Briviba, K., Schnabele, K., Rechkemmer, G., and Bub, A. (2004). Supplementation of a Diet Low in Carotenoids with Tomato or Carrot Juice Does Not Affect Lipid Peroxidation in Plasma and Feces of Healthy Men. J. Nutr. **134**: 1081–1083.

Brooks, C., Nekrasov, V., Lippman, Z.B., and Van Eck, J. (2014). Efficient Gene Editing in Tomato in the First Generation Using the Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-Associated9 System. PLANT Physiol. **166**: 1292–1297.

Brummell, D.A. and Harpster, M.H. (2001). Cell wall metabolism in fruit softening and quality and its manipulation in transgenic plants. Plant Mol. Biol. **47**: 311–340.

Buck, M.J. and Atchley, W.R. (2003). Phylogenetic analysis of plant basic helix-loop-helix proteins. J. Mol. Evol. **56**: 742–750.

Buckler, E.S. et al. (2009). The genetic architecture of maize flowering time. Science (80-.). **325**: 714–718.

Buiatti, E., Palli, D., Decarli, A., Amadori, D., Avellini, C., Bianchi, S., Biserni, R., Cipriani, F., Cocco, P., and Giacosa, A. (1989). A case-control study of gastric cancer and diet in Italy. Int. J. cancer **44**: 611–6.

Burger, Y., Paris, H.S., Cohen, R., Katzir, N., Tadmor, Y., Lewinsohn, E., and Schaffer, A.A. (2009). Genetic diversity of Cucumis Melo. In Horticultural Reviews, pp. 165–198.

Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., Reforgiato-Recupero, G., and Martin, C. (2012). Retrotransposons Control Fruit-Specific, Cold-Dependent Accumulation of Anthocyanins in Blood Oranges. Plant Cell **24**: 1242–1255.

Butelli, E., Titta, L., Giorgio, M., Mock, H.P., Matros, A., Peterek, S., Schijlen, E.G.W.M., Hall, R.D., Bovy, A.G., Luo, J., and Martin, C. (2008). Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nat. Biotechnol. **26**: 1301–1308.

Caldana, C., Scheible, W.R., Mueller-Roeber, B., and Ruzicic, S. (2007). A quantitative RT-PCR platform for high-throughput expression profiling of 2500 rice transcription factors. Plant Methods **3**.

Cárdenas, P.D. et al. (2016). GAME9 regulates the biosynthesis of steroidal alkaloids and upstream isoprenoids in the plant mevalonate pathway. Nat. Commun. **7**.

Cardoso, C. et al. (2014). Natural variation of rice strigolactone biosynthesis is associated with the deletion of two *MAX1* orthologs. Proc. Natl. Acad. Sci. **111**: 2379–2384.

Carrari, F. and Fernie, A.R. (2006). Metabolic regulation underlying tomato fruit development. In Journal of Experimental Botany, pp. 1883–1897.

Carretero-Paulet, L., Cairó, A., Botella-Pavía, P., Besumbes, O., Campos, N., Boronat, A., and Rodríguez-Concepción, M. (2006). Enhanced flux through the methylerythritol 4-phosphate pathway in Arabidopsis plants overexpressing deoxyxylulose 5-phosphate reductoisomerase. Plant Mol. Biol. **62**: 683–695.

Celenza, J.L. (2005). The Arabidopsis ATR1 Myb Transcription Factor Controls Indolic Glucosinolate Homeostasis. PLANT Physiol. **137**: 253–262.

Chan, E.K.F., Rowe, H.C., Corwin, J.A., Joseph, B., and Kliebenstein, D.J. (2011). Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in Arabidopsis thaliana. PLoS Biol. **9**.

Chappell, J., Wolf, F., Proulx, J., Cuellar, R., and Saunders, C. (1995). Is the Reaction Catalyzed by 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase a Rate-Limiting Step for Isoprenoid Biosynthesis in Plants? Plant Physiol. **109**: 1337–1343.

Chayut, N. et al. (2015). A bulk segregant transcriptome analysis reveals metabolic and cellular processes associated with Orange allelic variation and fruit β-carotene accumulation in melon fruit. BMC Plant Biol. **15**: 274.

Chayut, N. et al. (2017). Distinct Mechanisms of the ORANGE Protein in Controlling Carotenoid Flux. Plant Physiol. **173**: 376–389.

Chen, W. et al. (2014). Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. Nat. Genet. **46**: 714–721.

Cheng, C.Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S., and Town, C.D. (2017). Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. Plant J. **89**: 789–804.

Cherian, S., Figueroa, C.R., and Nair, H. (2014). "Movers and shakers" in the regulation of fruit ripening: A cross-dissection of climacteric versus non-climacteric fruit. J. Exp. Bot. **65**: 4705–4722.

Cho, E., Spiegelman, D., Hunter, D.J., Chen, W.Y., Zhang, S.M., Colditz, G.A., and Willett,

W.C. (2003). Premenopausal intakes of vitamins A, C, and E, folate, and carotenoids, and risk of breast cancer. Cancer Epidemiol. Biomarkers Prev. **12**: 713–720.

Chung, M.Y., Vrebalov, J., Alba, R., Lee, J., McQuinn, R., Chung, J.D., Klein, P., and Giovannoni, J. (2010). A tomato (Solanum lycopersicum) APETALA2/ERF gene, SlAP2a, is a negative regulator of fruit ripening. Plant J. **64**: 936–947.

Cocaliadis, M.F., Fernández-Muñoz, R., Pons, C., Orzaez, D., and Granell, A. (2014). Increasing tomato fruit quality by enhancing fruit chloroplast function. A double-edged sword? J. Exp. Bot. **65**: 4589–4598.

Collins, a R., Olmedilla, B., Southon, S., Granado, F., and Duthie, S.J. (1998). Serum carotenoids and oxidative DNA damage in human lymphocytes. Carcinogenesis **19**: 2159–2162.

Cominelli, E., Galbiati, M., and Tonelli, C. (2010). Transcription factors controlling stomatal movements and drought tolerance. Transcription **1**: 41–45.

Cong, L. et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science **339**: 819–23.

Cook-Mozaffari, P.J., Azordegan, F., Day, N.E., Ressicaud, A., Sabai, C., and Aramesh, B. (1979). Oesophageal cancer studies in the caspian littoral of iran: Results of a case-control study. Br. J. Cancer **39**: 293–309.

Cook, N.R., Le, I.M., Manson, J.E., Buring, J.E., and Hennekens, C.H. (2000). Effects of beta-carotene supplementation on cancer incidence by baseline characteristics in the Physicians' Health Study (United States). Cancer Causes Control **11**: 617–626.

Cunningham, F.X. and Gantt, E. (1998). GENES AND ENZYMES OF CAROTENOID BIOSYNTHESIS IN PLANTS. Annu. Rev. Plant Physiol. Plant Mol. Biol. **49**: 557–583.

Cunningham, F.X. and Gantt, E. (2001). One ring or two? Determination of ring number in carotenoids by lycopene ε-cyclases. Proc. Natl. Acad. Sci. U. S. A. **98**: 2905–2910.

Cvetkovic, D., Fiedor, L., Fiedor, J., Wisniewska-Becker, A., and Markovic, D. (2013). Molecular base for carotenoids antioxidant activity in model and biological systems: the health-related effects. In Carotenoids, pp. 93–126.

Davuluri, G.R. et al. (2005). Fruit-specific RNAi-mediated suppression of DET1 enhances carotenoid and flavonoid content in tomatoes. Nat. Biotechnol. **23**: 890–895.

Davuluri, G.R., Van Tuinen, A., Mustilli, A.C., Manfredonia, A., Newman, R., Burgess, D., Brummell, D.A., King, S.R., Palys, J., Uhlig, J., Pennings, H.M.J., and Bowler, C. (2004). Manipulation of DET1 expression in tomato results in photomorphogenic phenotypes caused by post-transcriptional gene silencing. Plant J. **40**: 344–354.

Davuluri, R. V., Sun, H., Palaniswamy, S.K., Matthews, N., Molina, C., Kurtz, M., and Grotewold, E. (2003). AGRIS: Arabidopsis Gene Regulatory Information Server, an information resource of Arabidopsis cis-regulatory elements and transcription factors. BMC Bioinformatics **4**.

Dixon, Z.R. et al. (1994). Effects of a carotene-deficient diet on measures of oxidative susceptibility and superoxide dismutase activity in adult women. Free Radic. Biol. Med. **17**: 537–544.

Dixon, Z.R., Shie, F.S., Warden, B.A., Burri, B.J., and Neidlinger, T.R. (1998). The effect of a low carotenoid diet on malondialdehyde-thiobarbituric acid (mda-tba) concentrations in women: A placebo-controlled double-blind study. J. Am. Coll. Nutr. **17**: 54–58.

Domonkos, I., Kis, M., Gombos, Z., and Ughy, B. (2013). Carotenoids, versatile components of oxygenic photosynthesis. Prog. Lipid Res. **52**: 539–561.

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

Eddy, S.R. (2011). Accelerated profile HMM searches. PLoS Comput. Biol. **7**.

Eliassen, A.H., Liao, X., Rosner, B., Tamimi, R.M., Tworoger, S.S., and Hankinson, S.E. (2015). Plasma carotenoids and risk of breast cancer over 20 y of follow-up. Am. J. Clin. Nutr. **101**: 1197–1205.

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

España, L., Heredia-Guerrero, J.A., Segado, P., Benítez, J.J., Heredia, A., and Domínguez, E. (2014). Biomechanical properties of the tomato (Solanum lycopersicum) fruit cuticle during development are modulated by changes in the relative amounts of its components. New Phytol. **202**: 790–802.

Estévez, J.M., Cantero, A., Reindl, A., Reichler, S., and León, P. (2001). 1-Deoxy-D-xylulose-5-phosphate Synthase, a Limiting Enzyme for Plastidic Isoprenoid Biosynthesis in Plants. J. Biol. Chem. **276**: 22901–22909.

Fairman, R., Beran-Steed, R.K., Anthony-Cahill, S.J., Lear, J.D., Stafford, W.F., DeGrado, W.F., Benfield, P.A., Brenner, S.L., and Brenner, S.L. (1993). Multiple oligomeric states regulate the DNA binding of helix-loop-helix peptides. Proc. Natl. Acad. Sci. U. S. A. **90**: 10429–33.

Fan, M., Bai, M.-Y., Kim, J.-G., Wang, T., Oh, E., Chen, L., Park, C.H., Son, S.-H., Kim, S.-K., Mudgett, M.B., and Wang, Z.-Y. (2014). The bHLH Transcription Factor HBI1 Mediates the Trade-Off between Growth and Pathogen-Associated Molecular Pattern-Triggered Immunity in Arabidopsis. Plant Cell **26**: 828–841.

Fantini, E., Falcone, G., Frusciante, S., Giliberto, L., and Giuliano, G. (2013). Dissection of Tomato Lycopene Biosynthesis through Virus-Induced Gene Silencing. PLANT Physiol. **163**: 986–998.

Feng, Z. et al. (2014). Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. Proc. Natl. Acad. Sci. **111**: 4632–4637.

Ferré-D'Amaré, A.R., Prendergast, G.C., Ziff, E.B., and Burley, S.K. (1993). Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. Nature **363**: 38–45.

Franco-Zorrilla, J.M., López-Vidriero, I., Carrasco, J.L., Godoy, M., Vera, P., and Solano, R. (2014). DNA-binding specificities of plant transcription factors and their potential to define target genes. Proc. Natl. Acad. Sci. **111**: 2367–2372.

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

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

Friedrichsen, D.M., Nemhauser, J., Muramitsu, T., Maloof, J.N., Alonso, J., Ecker, J.R., Furuya, M., and Chory, J. (2002). Three redundant brassinosteroid early response genes encode putative bHLH transcription factors required for normal growth. Genetics **162**: 1445–1456.

Fromme, P. (2008). Photosynthetic Protein Complexes: A Structural Approach.

Frusciante, S., Diretto, G., Bruno, M., Ferrante, P., Pietrella, M., Prado-Cabrero, A., Rubio-Moraga, A., Beyer, P., Gomez-Gomez, L., Al-Babili, S., and Giuliano, G. (2014). Novel carotenoid cleavage dioxygenase catalyzes the first dedicated step in saffron crocin biosynthesis. Proc. Natl. Acad. Sci. **111**: 12246–12251.

Fujisawa, M., Nakano, T., and Ito, Y. (2011). Identification of potential target genes for the tomato fruit-ripening regulator RIN by chromatin immunoprecipitation. BMC Plant Biol. **11**.

Fujisawa, M., Nakano, T., Shima, Y., and Ito, Y. (2013). A Large-Scale Identification of Direct Targets of the Tomato MADS Box Transcription Factor RIPENING INHIBITOR Reveals the Regulation of Fruit Ripening. Plant Cell **25**: 371–386.

Fujisawa, M., Shima, Y., Higuchi, N., Nakano, T., Koyama, Y., Kasumi, T., and Ito, Y. (2012). Direct targets of the tomato-ripening regulator RIN identified by transcriptome and chromatin immunoprecipitation analyses. Planta **235**: 1107–1122.

Fulton, T.M., Beck-Bunn, T., Emmatty, D., Eshed, Y., Lopez, J., Petiard, V., Uhlig, J., Zamir, D., and Tanksley, S.D. (1997). QTL analysis of an advanced backcross of Lycopersicon peruvianum to the cultivated tomato and comparisons with QTLs found in other wild species. Theor. Appl. Genet. **95**: 881–894.

Gady, A.L.F., Vriezen, W.H., Van de Wal, M.H.B.J., Huang, P., Bovy, A.G., Visser, R.G.F., and Bachem, C.W.B. (2012). Induced point mutations in the phytoene synthase 1 gene cause differences in carotenoid content during tomato fruit ripening. Mol. Breed. **29**: 801–812.

Galpaz, N. (2006). A Chromoplast-Specific Carotenoid Biosynthesis Pathway Is Revealed by Cloning of the Tomato white-flower Locus. PLANT CELL ONLINE **18**: 1947–1960.

Galpaz, N., Wang, Q., Menda, N., Zamir, D., and Hirschberg, J. (2008). Abscisic acid deficiency in the tomato mutant high-pigment 3 leading to increased plastid number and higher fruit lycopene content. Plant J. **53**: 717–730.

García-Limones, C., Schnäbele, K., Blanco-Portales, R., Bellido, M.L., Caballero, J.L., Schwab, W., and Muñoz-Blanco, J. (2008). Functional characterization of FaCCD1: A carotenoid cleavage dioxygenase from strawberry involved in lutein degradation during fruit ripening. J. Agric. Food Chem. **56**: 9277–9285.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. **109**: E2579–E2586.

Gey, K.F., Stähelin, H.B., and Eichholzer, M. (1993). Poor plasma status of carotene and vitamin C is associated with higher mortality from ischemic heart disease and stroke Basel Prospective Study. Clin. Investig. **71**: 3–6.

Gibson, G. (2010). Hints of hidden heritability in GWAS. Nat. Genet. **42**: 558–560.

Gietz, R.D. and Woods, R.A. (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. Methods Enzymol. **350**: 87–96.

Giovannoni, J.J., Noensie, E.N., Ruezinsky, D.M., Lu, X., Tracy, S.L., Ganal, M.W., Martin, G.B., Pillen, K., Albert, K., and Tankslev, S.D. (1995). 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. MGG Mol. Gen. Genet. **248**: 195–206.

Giovannoni, J.J.J. (2004). Genetic regulation of fruit development and ripening. Plant Cell Online.

Gong, L., Chen, W., Gao, Y., Liu, X., Zhang, H., Xu, C., Yu, S., Zhang, Q., and Luo, J. (2013). Genetic analysis of the metabolome exemplified using a rice population. Proc. Natl. Acad. Sci. **110**: 20320–20325.

Gonzalez, A., Zhao, M., Leavitt, J.M., and Lloyd, A.M. (2008). Regulation of the anthocyanin

biosynthetic pathway by the TTG1 / bHLH / Myb transcriptional complex in Arabidopsis seedlings. Regulation **53**: 814–827.

Green, B.R. and Parson, W.W. (2003). Light-harvesting antennas in photosynthesis. Adv. Photosynth. Respir. ; **13**: 513.

Greenberg, E.R., Baron, J.A., Karagas, M.R., Stukel, T.A., Nierenberg, D.W., Stevens, M.M., Mandel, J.S., and Haile, R.W. (1996). Mortality associated with low plasma concentration of beta carotene and the effect of oral supplementation. JAMA **275**: 699–703.

Guo, A., He, K., Liu, D., Bai, S., Gu, X., Wei, L., and Luo, J. (2005). DATF: A database of Arabidopsis transcription factors. Bioinformatics **21**: 2568–2569.

Guzman, I., Hamby, S., Romero, J., Bosland, P.W., and O'Connell, M.A. (2010). Variability of carotenoid biosynthesis in orange colored Capsicum spp. Plant Sci. **179**: 49–59.

Ha, S.H., Kim, J.B., Park, J.S., Lee, S.W., and Cho, K.J. (2007). A comparison of the carotenoid accumulation in Capsicum varieties that show different ripening colours: Deletion of the capsanthin-capsorubin synthase gene is not a prerequisite for the formation of a yellow pepper. J. Exp. Bot. **58**: 3135–3144.

Hachez, C., Ohashi-Ito, K., Dong, J., and Bergmann, D.C. (2011). Differentiation of Arabidopsis guard cells: analysis of the networks incorporating the basic helix-loop-helix transcription factor, FAMA. Plant Physiol. **155**: 1458–72.

Hamauzu, Y., Chachin, K., and Ueda, Y. (1998). Effect of postharvest storage temperature on the conversion of 14C-Mevalonic acid to carotenes in tomato fruit. J. Japanese Soc. Hortic. Sci. **67**: 549–855.

Hamilton, A.J., Lycett, G.W., and Grierson, D. (1990). Antisense gene that inhibits synthesis of the hormone ethylene in transgenic plants. Nature **346**: 284–287.

Heim, M.A., Jakoby, M., Werber, M., Martin, C., Weisshaar, B., and Bailey, P.C. (2003). The basic helix-loop-helix transcription factor family in plants: A genome-wide study of protein structure and functional diversity. Mol. Biol. Evol. **20**: 735–747.

Helzlsouer, K.J., Comstock, G.W., and Morris, J.S. (1989). Selenium, lycopene, alpha-

tocopherol, beta-carotene, retinol, and subsequent bladder cancer. Cancer Res. **49**: 6144–8.

Hennekens, C.H., Buring, J.E., Manson, J.E., Stampfer, M., Rosner, B., Cook, N.R., Belanger, C., LaMotte, F., Gaziano, J.M., Ridker, P.M., Willett, W., and Peto, R. (1996). Lack of Effect of Long-Term Supplementation with Beta Carotene on the Incidence of Malignant Neoplasms and Cardiovascular Disease. N. Engl. J. Med. **334**: 1145–1149.

Hermoso, A. (2004). TrSDB: a proteome database of transcription factors. Nucleic Acids Res. **32**: 171D–173.

Le Hir, R., Castelain, M., Chakraborti, D., Moritz, T., Dinant, S., and Bellini, C. (2017). AtbHLH68 transcription factor contributes to the regulation of ABA homeostasis and drought stress tolerance in Arabidopsis thaliana. Physiol. Plant. **160**: 312–327.

Hirschberg, J. (2001). Carotenoid biosynthesis in flowering plants. Curr Opin Plant Biol **4**: 210–218.

Holton, N., Cano-Delgado, A., Harrison, K., Montoya, T., Chory, J., and Bishop, G.J. (2007). Tomato BRASSINOSTEROID INSENSITIVE1 Is Required for Systemin-Induced Root Elongation in Solanum pimpinellifolium but Is Not Essential for Wound Signaling. PLANT CELL ONLINE **19**: 1709–1717.

Huang, S. et al. (2009). The genome of the cucumber, Cucumis sativus L. Nat. Genet. **41**: 1275–1281.

Huang, X. et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Genet. **42**: 961–967.

Hughes, K.A., Ayroles, J.F., Reedy, M.M., Drnevich, J.M., Rowe, K.C., Ruedi, E.A., Cáceres, C.E., and Paige, K.N. (2006). Segregating variation in the transcriptome: Cis regulation and additivity of effects. Genetics **173**: 1347–1355.

Ilg, A., Bruno, M., Beyer, P., and Al-Babili, S. (2014). Tomato carotenoid cleavage dioxygenases 1A and 1B: Relaxed double bond specificity leads to a plenitude of dialdehydes, mono-apocarotenoids and isoprenoid volatiles. FEBS Open Bio **4**: 584–593.

Immink, R.G.H., Tonaco, I. a N., de Folter, S., Shchennikova, A., van Dijk, A.D.J., Busscher-Lange, J., Borst, J.W., and Angenent, G.C. (2009). SEPALLATA3: the “glue” for MADS box transcription factor complex formation. Genome Biol. **10**: R24.

Isaacson, T. (2002a). Cloning of tangerine from Tomato Reveals a Carotenoid Isomerase Essential for the Production of beta-Carotene and Xanthophylls in Plants. PLANT CELL ONLINE **14**: 333–342.

Isaacson, T. (2002b). Cloning of tangerine from Tomato Reveals a Carotenoid Isomerase Essential for the Production of beta-Carotene and Xanthophylls in Plants. PLANT CELL ONLINE **14**: 333–342.

Itkin, M. et al. (2013). Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. Science (80-.). **341**: 175–179.

Itkin, M., Seybold, H., Breitel, D., Rogachev, I., Meir, S., and Aharoni, A. (2009). TOMATO AGAMOUS-LIKE 1 is a component of the fruit ripening regulatory network. Plant J. **60**: 1081–1095.

Ito, Y., Kitagawa, M., Ihashi, N., Yabe, K., Kimbara, J., Yasuda, J., Ito, H., Inakuma, T., Hiroi, S., and Kasumi, T. (2008). DNA-binding specificity, transcriptional activation potential, and the rin mutation effect for the tomato fruit-ripening regulator RIN. Plant J. **55**: 212–223.

Ito, Y., Nishizawa-Yokoi, A., Endo, M., Mikami, M., Shima, Y., Nakamura, N., Kotake-Nara, E., Kawasaki, S., and Toki, S. (2017). Re-evaluation of the rin mutation and the role of RIN in the induction of tomato ripening. Nat. Plants **3**: 866–874.

Jahns, P. and Holzwarth, A.R. (2012). The role of the xanthophyll cycle and of lutein in photoprotection of photosystem II. Biochim. Biophys. Acta - Bioenerg. **1817**: 182–193.

Janik, E., Bednarska, J., Zubik, M., Sowinski, K., Luchowski, R., Grudzinski, W., Matosiuk, D., and Gruszecki, W.I. (2016). The xanthophyll cycle pigments, violaxanthin and zeaxanthin, modulate molecular organization of the photosynthetic antenna complex LHCII. Arch. Biochem. Biophys. **592**: 1–9.

Jiang, C., Wright, R.J., Woo, S.S., DelMonte, T.A., and Paterson, A.H. (2000). QTL analysis of leaf morphology in tetraploid Gossypium (cotton). Theor. Appl. Genet. **100**: 409–

418.

Jiang, W.-B., Huang, H.-Y., Hu, Y.-W., Zhu, S.-W., Wang, Z.-Y., and Lin, W.-H. (2013a). Brassinosteroid Regulates Seed Size and Shape in Arabidopsis. PLANT Physiol. **162**: 1965–1977.

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B., and Weeks, D.P. (2013b). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucleic Acids Res. **41**.

Jiang, Y., Zeng, B., Zhao, H., Zhang, M., Xie, S., and Lai, J. (2012). Genome-wide Transcription Factor Gene Prediction and their Expressional Tissue-Specificities in Maize. J. Integr. Plant Biol. **54**: 616–630.

Jimenez, A., Creissen, G., Kular, B., Firmin, J., Robinson, S., Verhoeyen, M., and Mullineaux, P. (2002). Changes in oxidative processes and components of the antioxidant system during tomato fruit ripening. Planta **214**: 751–758.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptice Bacterial Immunity. Science **337**: 816–822.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. Elife **2013**.

Jones, B., Frasse, P., Olmos, E., Zegzouti, H., Li, Z.G., Latché, A., Pech, J.C., and Bouzayen, M. (2002). Down-regulation of DR12, an auxin-response-factor homolog, in the tomato results in a pleiotropic phenotype including dark green and blotchy ripening fruit. Plant J. **32**: 603–613.

Jones, S. (2004). An overview of the basic helix-loop-helix proteins. Genome Biol. **5**.

Joosen, R.V.L., Arends, D., Li, Y., Willems, L. a J., Keurentjes, J.J.B., Ligterink, W., Jansen, R.C., and Hilhorst, H.W.M. (2013). Identifying Genotype-by-Environment Interactions in the Metabolism of Germinating Arabidopsis Seeds Using Generalized Genetical Genomics. Plant Physiol. **162**: 553–566.

Kachanovsky, D.E., Filler, S., Isaacson, T., and Hirschberg, J. (2012). Epistasis in tomato

color mutations involves regulation of phytoene synthase 1 expression by cis-carotenoids. Proc. Natl. Acad. Sci. **109**: 19021–19026.

Karlova, R., Rosin, F.M., Busscher-Lange, J., Parapunova, V., Do, P.T., Fernie, A.R., Fraser, P.D., Baxter, C., Angenent, G.C., and de Maagd, R.A. (2011). Transcriptome and Metabolite Profiling Show That APETALA2a Is a Major Regulator of Tomato Fruit Ripening. Plant Cell **23**: 923–941.

Kato, M. (2012). Mechanism of carotenoid accumulation in citrus fruit. J. Japanese Soc. Hortic. Sci. **81**: 219–33.

Kawahara, Y. et al. (2013). Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. Rice **6**: 1–10.

Keurentjes, J.J.B., Fu, J., Terpstra, I.R., Garcia, J.M., van den Ackerveken, G., Snoek, L.B., Peeters, A.J.M., Vreugdenhil, D., Koornneef, M., and Jansen, R.C. (2007). Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. Proc. Natl. Acad. Sci. U. S. A. **104**: 1708–13.

Kevany, B.M., Tieman, D.M., Taylor, M.G., Cin, V.D., and Klee, H.J. (2007). Ethylene receptor degradation controls the timing of ripening in tomato fruit. Plant J. **51**: 458–467.

Kim, S. et al. (2014). Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. Nat. Genet. **46**: 270–278.

Kim, Y.B., Kim, S.M., Kang, M.K., Kuzuyama, T., Lee, J.K., Park, S.C., Shin, S.C., and Kim, S.U. (2009). Regulation of resin acid synthesis in Pinus densiflora by differential transcription of genes encoding multiple 1-deoxy-d-xylulose 5-phosphate synthase and 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase genes. Tree Physiol. **29**: 737–749.

Klee, H.J. and Giovannoni, J.J. (2011). Genetics and Control of Tomato Fruit Ripening and Quality Attributes. Annu. Rev. Genet. **45**: 41–59.

Knapp, S., Bohs, L., Nee, M., and Spooner, D.M. (2004). Solanaceae - A model for linking genomics with biodiversity. Comp. Funct. Genomics **5**: 285–291.

Koenig, D. et al. (2013). Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. Proc. Natl. Acad. Sci. U. S. A. **110**: E2655-62.

Koes, R., Verweij, W., and Quattrocchio, F. (2005). Flavonoids: A colorful model for the regulation and evolution of biochemical pathways. Trends Plant Sci. **10**: 236–242.

Koka, C. V, Cerny, R.E., Gardner, R.G., Noguchi, T., Fujioka, S., Takatsuto, S., Yoshida, S., and Clouse, S.D. (2000). A putative role for the tomato genes DUMPY and CURL-3 in brassinosteroid biosynthesis and response. Plant Physiol. **122**: 85–98.

Krieger-Liszkay, A. (2005). Singlet oxygen production in photosynthesis. In Journal of Experimental Botany, pp. 337–346.

Kump, K.L., Bradbury, P.J., Wisser, R.J., Buckler, E.S., Belcher, A.R., Oropeza-Rosas, M.A., Zwonitzer, J.C., Kresovich, S., McMullen, M.D., Ware, D., Balint-Kurti, P.J., and Holland, J.B. (2011). Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat. Genet. **43**: 163–168.

Laguna, L., Casado, C.G., and Heredia, A. (1999). Flavonoid biosynthesis in tomato fruit cuticles after in vivo incorporation of 3H-phenylalanine precursor. Physiol. Plant. **105**: 491–498.

Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell **172**: 650–665.

Lanahan, M.B. (1994). The Never Ripe Mutation Blocks Ethylene Perception in Tomato. PLANT CELL ONLINE **6**: 521–530.

Lashbrook, C.C., Tieman, D.M., and Klee, H.J. (1998). Differential regulation of the tomato ETR gene family throughout plant development. Plant J. **15**: 243–252.

Lee, A., Thurnham, D.I., and Chopra, M. (2000). Consumption of tomato products with olive oil but not sunflower oil increases the antioxidant activity of plasma. Free Radic. Biol. Med. **29**: 1051–1055.

Lee, I.M., Cook, N.R., Manson, J.E., Buring, J.E., and Hennekens, C.H. (1999). Beta-carotene

supplementation and incidence of cancer and cardiovascular disease: the Women's Health Study. J Natl Cancer Inst **91**: 2102–2106.

Lee, J.M., Joung, J.G., McQuinn, R., Chung, M.Y., Fei, Z., Tieman, D., Klee, H., and Giovannoni, J. (2012). Combined transcriptome, genetic diversity and metabolite profiling in tomato fruit reveals that the ethylene response factor SlERF6 plays an important role in ripening and carotenoid accumulation. Plant J. **70**: 191–204.

Lepage, G., Champagne, J., Ronco, N., Lamarre, A., Osberg, I., Sokol, R.J., and Roy, C.C. (1996). Supplementation with carotenoids corrects increased lipid peroxidation in children with cystic fibrosis. Am. J. Clin. Nutr. **64**: 87–93.

Letunic, I. and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. **44**: W242–W245.

Li, F., Vallabhaneni, R., and Wurtzel, E.T. (2008a). PSY3, a New Member of the Phytoene Synthase Gene Family Conserved in the Poaceae and Regulator of Abiotic Stress-Induced Root Carotenogenesis. PLANT Physiol. **146**: 1333–1345.

Li, F., Vallabhaneni, R., and Wurtzel, E.T. (2008b). PSY3, a New Member of the Phytoene Synthase Gene Family Conserved in the Poaceae and Regulator of Abiotic Stress-Induced Root Carotenogenesis. PLANT Physiol. **146**: 1333–1345.

Li, L., Paolillo, D.J., Parthasarathy, M. V., DiMuzio, E.M., and Garvin, D.F. (2001). A novel gene mutation that confers abnormal patterns of β-carotene accumulation in cauliflower (Brassica oleracea var. botrytis). Plant J. **26**: 59–67.

Li, L. and Yuan, H. (2013). Chromoplast biogenesis and carotenoid accumulation. Arch. Biochem. Biophys. **539**: 102–109.

Li, R., Li, Y., Fang, X., Yang, H., Wang, J., Kristiansen, K., and Wang, J. (2009a). SNP detection for massively parallel whole-genome resequencing. Genome Res. **19**: 1124–1132.

Li, R., Yu, C., Li, Y., Lam, T.W., Yiu, S.M., Kristiansen, K., and Wang, J. (2009b). SOAP2: An improved ultrafast tool for short read alignment. Bioinformatics **25**: 1966–1967.

Li, X., Duan, X., Jiang, H., Sun, Y., Tang, Y., Yuan, Z., and Guo, J. (2006). Genome-Wide Analysis of Basic / Helix-Loop-Helix Transcription Factor Family in Rice and Arabidopsis.

Plant Physiol. **141**: 1167–1184.

Libault, M., Joshi, T., Benedito, V.A., Xu, D., Udvardi, M.K., and Stacey, G. (2009). Legume Transcription Factor Genes: What Makes Legumes So Special? PLANT Physiol. **151**: 991–1001.

Lichtenthaler, H.K. (1999). THE 1-DEOXY-D-XYLULOSE-5-PHOSPHATE PATHWAY OF ISOPRENOID BIOSYNTHESIS IN PLANTS. Annu. Rev. Plant Physiol. Plant Mol. Biol. **50**: 47–65.

Lichtenthaler, H.K., Schwender, J., Disch, A., and Rohmer, M. (1997). Biosynthesis of isoprenoids in higher plant chloroplasts proceeds via a mevalonate-independent pathway. FEBS Lett. **400**: 271–274.

Lieberman, M. and Kunishi, a (1966). Stimulation of ethylene production in apple tissue slices by methionine. Plant Physiol. **41**: 376–382.

Liebler, D.C., Stratton, S.P., and Kaysen, K.L. (1997). Antioxidant actions of β-carotene in liposomal and microsomal membranes: Role of carotenoid-membrane incorporation and α-tocopherol. Arch. Biochem. Biophys. **338**: 244–250.

von Lintig, J. and Vogt, K. (2000). Filling the gap in vitamin A research. Molecular identification of an enzyme cleaving beta-carotene to retinal. J. Biol. Chem. **275**: 11915–11920.

Lipka, A.E., Gore, M. a., Magallanes-Lundback, M., Mesberg, A., Lin, H., Tiede, T., Chen, C., Buell, C.R., Buckler, E.S., Rocheford, T., and DellaPenna, D. (2013). Genome-Wide Association Study and Pathway Level Analysis of Tocochromanol Levels in Maize Grain. G3 Genes|Genomes|Genetics **3**: g3.113.006148.

Lippert, C., Listgarten, J., Liu, Y., Kadie, C.M., Davidson, R.I., and Heckerman, D. (2011). FaST linear mixed models for genome-wide association studies. Nat. Methods **8**: 833–835.

Lippman, Z.B., Semel, Y., and Zamir, D. (2007). An integrated view of quantitative trait variation using tomato interspecific introgression lines. Curr. Opin. Genet. Dev. **17**: 545–552.

Liu, L., Jia, C., Zhang, M., Chen, D., Chen, S., Guo, R., Guo, D., and Wang, Q. (2014a). Ectopic expression of a BZR1-1D transcription factor in brassinosteroid signalling enhances carotenoid accumulation and fruit quality attributes in tomato. Plant Biotechnol. J. **12**: 105–115.

Liu, L., Shao, Z., Zhang, M., and Wang, Q. (2015a). Regulation of carotenoid metabolism in tomato. Mol. Plant **8**: 28–39.

Liu, L., Wei, J., Zhang, M., Zhang, L., Li, C., and Wang, Q. (2012). Ethylene independent induction of lycopene biosynthesis in tomato fruits by jasmonates. J. Exp. Bot. **63**: 5751–5762.

Liu, M., Diretto, G., Pirrello, J., Roustan, J.P., Li, Z., Giuliano, G., Regad, F., and Bouzayen, M. (2014b). The chimeric repressor version of an Ethylene Response Factor (ERF) family member, Sl-ERF.B3, shows contrasting effects on tomato fruit ripening. New Phytol. **203**: 206–218.

Liu, M., Pirrello, J., CHERVIN, C., Roustan, J.-P., and Bouzayen, M. (2015b). Ethylene control of fruit ripening: revisiting the complex network of transcriptional regulation. Plant Physiol.: pp.01361.2015.

Liu, M., Pirrello, J., Kesari, R., Mila, I., Roustan, J.P., Li, Z., Latch, A., Pech, J.C., Bouzayen, M., and Regad, F. (2013). A dominant repressor version of the tomato Sl-ERF.B3 gene confers ethylene hypersensitivity via feedback regulation of ethylene signaling and response components. Plant J. **76**: 406–419.

Liu, Y.-S., Gur, A., Ronen, G., Causse, M., Damidaux, R., Buret, M., Hirschberg, J., and Zamir, D. (2003). There is more to tomato fruit colour than candidate carotenoid genes. Plant Biotechnol. J. **1**: 195–207.

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

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

Llorente, B., D'Andrea, L., Ruiz-Sola, M.A., Botterweg, E., Pulido, P., Andilla, J., Loza-Alvarez, P., and Rodriguez-Concepcion, M. (2016). Tomato fruit carotenoid biosynthesis is adjusted to actual ripening progression by a light-dependent mechanism. Plant J. **85**: 107–119.

Lockhart, J. (2013). A Quantitative Genetic Basis for Leaf Morphology is Revealed in a Set of Precisely Defined Tomato Introgression Lines. Plant Cell **25**: 2379–2379.

Lois, L.M., Rodríguez-Concepción, M., Gallego, F., Campos, N., and Boronat, A. (2000). Carotenoid biosynthesis during tomato fruit development: Regulatory role of 1-deoxy-D-xylulose 5-phosphate synthase. Plant J. **22**: 503–513.

Lopez, A.B., Van Eck, J., Conlin, B.J., Paolillo, D.J., O'Neill, J., and Li, L. (2008). Effect of the cauliflower or transgene on carotenoid accumulation and chromoplast formation in transgenic potato tubers. J. Exp. Bot. **59**: 213–223.

Lorenzo, Y., Azqueta, A., Luna, L., Bonilla, F., Domínguez, G., and Collins, A.R. (2009). The carotenoid β-cryptoxanthin stimulates the repair of DNA oxidation damage in addition to acting as an antioxidant in human cells. Carcinogenesis **30**: 308–314.

Lu, S. et al. (2006). The cauliflower Or gene encodes a DnaJ cysteine-rich domain-containing protein that mediates high levels of beta-carotene accumulation. Plant Cell **18**: 3594–3605.

Lv, P., Li, N., Liu, H., Gu, H., and Zhao, W.E. (2015). Changes in carotenoid profiles and in the expression pattern of the genes in carotenoid metabolisms during fruit development and ripening in four watermelon cultivars. Food Chem. **174**: 52–59.

Lyzenga, W.J. and Stone, S.L. (2012). Abiotic stress tolerance mediated by protein ubiquitination. J. Exp. Bot. **63**: 599–616.

Malinovsky, F.G., Batoux, M., Schwessinger, B., Youn, J.H., Stransfeld, L., Win, J., Kim, S.-K., and Zipfel, C. (2014). Antagonistic Regulation of Growth and Immunity by the Arabidopsis Basic Helix-Loop-Helix Transcription Factor HOMOLOG OF BRASSINOSTEROID ENHANCED EXPRESSION2 INTERACTING WITH INCREASED LEAF INCLINATION1 BINDING bHLH1. PLANT Physiol. **164**: 1443–1455.

Manning, K., Tör, M., Poole, M., Hong, Y., Thompson, A.J., King, G.J., Giovannoni, J.J., and

Seymour, G.B. (2006). A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nat. Genet. **38**: 948–952.

Martel, C., Vrebalov, J., Tafelmeyer, P., and Giovannoni, J.J. (2011). The Tomato MADS-Box Transcription Factor RIPENING INHIBITOR Interacts with Promoters Involved in Numerous Ripening Processes in a COLORLESS NONRIPENING-Dependent Manner. PLANT Physiol. **157**: 1568–1579.

Marty, I., Bureau, S., Sarkissian, G., Gouble, B., Audergon, J.M., and Albagnac, G. (2005). Ethylene regulation of carotenoid accumulation and carotenogenic gene expression in colour-contrasted apricot varieties (Prunus armeniaca). J. Exp. Bot. **56**: 1877–1886.

Massari, M.E. and Murre, C. (2000). Helix-Loop-Helix Proteins: Regulators of Transcription in Eucaryotic Organisms. Mol. Cell. Biol. **20**: 429–440.

MATHEWS-ROTH, M.M. (1993). Carotenoids in Erythropoietic Protoporphyria and Other Photosensitivity Diseases. Ann. N. Y. Acad. Sci. **691**: 127–138.

Mathews-Roth, M.M. (1983). CAROTENOID PIGMENT ADMINISTRATION and DELAY IN DEVELOPMENT OF UV-B-INDUCED TUMORS. Photochem. Photobiol. **37**: 509–511.

Matos, H.R., Capelozzi, V.L., Gomes, O.F., Mascio, P.D., and Medeiros, M.H.G. (2001). Lycopene inhibits DNA damage and liver necrosis in rats treated with ferric nitrilotriacetate. Arch. Biochem. Biophys. **396**: 171–177.

Matsuda, F., Nakabayashi, R., Yang, Z., Okazaki, Y., Yonemaru, J.I., Ebana, K., Yano, M., and Saito, K. (2015). Metabolome-genome-wide association study dissects genetic architecture for generating natural variation in rice secondary metabolism. Plant J. **81**: 13–23.

Mayne, S.T. (1996). Beta-carotene, carotenoids, and disease prevention in humans. FASEB J. **10**: 690–701.

Meraji, S., Ziouzenkova, O., Resch, U., Khoschsorur, a, Tatzber, F., and Esterbauer, H. (1997). Enhanced plasma level of lipid peroxidation in Iranians could be improved by antioxidants supplementation. Eur. J. Clin. Nutr. **51**: 318–25.

Milborrow, B. V and Lee, H.-S. (1998). Endogenous biosynthetic precursors of (+)-abscisic

acid. VI* Carotenoids and ABA are formed by the `non-mevalonate´ triose-pyruvate pathway in chloroplasts. Funct. Plant Biol. **25**: 507–512.

Mintz-Oron, S., Mandel, T., Rogachev, I., Feldberg, L., Lotan, O., Yativ, M., Wang, Z., Jetter, R., Venger, I., Adato, A., and Aharoni, A. (2008). Gene Expression and Metabolism in Tomato Fruit Surface Tissues. PLANT Physiol. **147**: 823–851.

Montoya, T., Nomura, T., Yokota, T., Farrar, K., Harrison, K., Jones, J.G.D., Kaneta, T., Kamiya, Y., Szekeres, M., and Bishop, G.J. (2005). Patterns of Dwarf expression and brassinosteroid accumulation in tomato reveal the importance of brassinosteroid synthesis during fruit development. Plant J. **42**: 262–269.

Moyer, V.A. (2014). Vitamin, mineral, and multivitamin supplements for the primary prevention of cardiovascular disease and cancer: U.S. preventive services task force recommendation statement. Ann. Intern. Med. **160**: 558–564.

Moyle, L.C. (2008). Ecological and evolutionary genomics in the wild tomatoes (Solanum Sect. Lycopersicon). Evolution (N. Y). **62**: 2995–3013.

Muller, P. (2001). Non-Photochemical Quenching. A Response to Excess Light Energy. PLANT Physiol. **125**: 1558–1566.

Murre, C., Bain, G., van Dijk, M.A., Engel, I., Furnari, B.A., Massari, M.E., Matthews, J.R., Quong, M.W., Rivera, R.R., and Stuiver, M.H. (1994). Structure and function of helix-loop-helix proteins. BBA - Gene Struct. Expr. **1218**: 129–135.

Mustilli, A.C. (1999). Phenotype of the Tomato high pigment-2 Mutant Is Caused by a Mutation in the Tomato Homolog of DEETIOLATED1. PLANT CELL ONLINE **11**: 145–158.

Nadakuduti, S.S., Holdsworth, W.L., Klein, C.L., and Barry, C.S. (2014). KNOX genes influence a gradient of fruit chloroplast development through regulation of GOLDEN2-LIKE expression in tomato. Plant J. **78**: 1022–1033.

Nair, S.K. and Burley, S.K. (2000). Recognizing DNA in the library. Nature **404**: 715–718.

Nakatsuka, A., Murachi, S., Okunishi, H., Shiomi, S., Nakano, R., Kubo, Y., and Inaba, A. (1998). Differential expression and internal feedback regulation of 1-aminocyclopropane-1-carboxylate synthase, 1-aminocyclopropane-1-carboxylate

oxidase, and ethylene receptor genes in tomato fruit during development and ripening. Plant Physiol. **118**: 1295–1305.

Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J.D.G., and Kamoun, S. (2013). Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. Nat. Biotechnol. **31**: 691–693.

Nemri, a, Atwell, S., Tarone, a M., Huang, Y.S., Zhao, K., Studholme, D.J., Nordborg, M., and Jones, J.D.G. (2010). Genome-wide survey of Arabidopsis natural variation in downy mildew resistance using combined association and linkage mapping. Pnas **107**: 10302–10307.

Nesi, N., Jond, C., Debeaujon, I., Caboche, M., and Lepiniec, L. (2001). The Arabidopsis TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. Plant Cell **13**: 2099–2114.

Nie, S., Huang, S., Wang, S., Cheng, D., Liu, J., Lv, S., Li, Q., and Wang, X. (2017). Enhancing Brassinosteroid Signaling via Overexpression of Tomato (Solanum lycopersicum) SlBRI1 Improves Major Agronomic Traits. Front. Plant Sci. **8**.

Nishino, H., Murakoshi, M., Tokuda, H., and Satomi, Y. (2009). Cancer prevention by carotenoids. Arch. Biochem. Biophys. **483**: 165–168.

Niyogi, K.K., Grossman, A.R., and Björkman, O. (1998). Arabidopsis mutants define a central role for the xanthophyll cycle in the regulation of photosynthetic energy conversion. Plant Cell **10**: 1121–34.

North, H.M., Almeida, A. De, Boutin, J.P., Frey, A., To, A., Botran, L., Sotta, B., and Marion-Poll, A. (2007). The Arabidopsis ABA-deficient mutant aba4 demonstrates that the major route for stress-induced ABA accumulation is via neoxanthin isomers. Plant J. **50**: 810–824.

Oguz, M. (2017). Biochemistry and Antioxidant Properties of Carotenoids. In Carotenoids, pp. 51–66.

Omenn, G.S. et al. (1996). Risk factors for lung cancer and for intervention effects in CARET, the beta-carotene and retinol efficacy trial. J. Natl. Cancer Inst. **88**: 1550–1559.

Orzaez, D. (2005). Agroinjection of Tomato Fruits. A Tool for Rapid Functional Analysis of Transgenes Directly in Fruit. PLANT Physiol. **140**: 3–11.

Orzaez, D., Medina, A., Torre, S., Fernandez-Moreno, J.P., Rambla, J.L., Fernandez-del-Carmen, A., Butelli, E., Martin, C., and Granell, A. (2009). A Visual Reporter System for Virus-Induced Gene Silencing in Tomato Fruit Based on Anthocyanin Accumulation. PLANT Physiol. **150**: 1122–1134.

Osganian, S.K., Stampfer, M.J., Rimm, E., Spiegelman, D., Manson, J.E., and Willett, W.C. (2003). Dietary carotenoids and risk of coronary artery disease in women. Am. J. Clin. Nutr. **77**: 1390–1399.

Paetzold, H., Garms, S., Bartram, S., Wieczorek, J., Urós-Gracia, E.M., Rodríguez-Concepción, M., Boland, W., Strack, D., Hause, B., and Walter, M.H. (2010). The isogene 1-deoxy-D-xylulose 5-phosphate synthase 2 controls isoprenoid profiles, precursor pathway allocation, and density of tomato trichomes. Mol. Plant **3**: 904–916.

Palozza, P. (1998). Prooxidant actions of carotenoids in biologic systems. Nutr. Rev. **56**: 257–265.

Palozza, P., Calviello, G., Emilia De Leo, M., Serini, S., and Bartoli, G.M. (2000). Canthaxanthin supplementation alters antioxidant enzymes and iron concentration in liver of balb/c mice. J. Nutr. **130**: 1303–1308.

Pan, I.L., McQuinn, R., Giovannoni, J.J., and Irish, V.F. (2010). Functional diversification of AGAMOUS lineage genes in regulating tomato flower and fruit development. J. Exp. Bot. **61**: 1795–1806.

Pan, Y. et al. (2013). Network Inference Analysis Identifies an APRR2-Like Gene Linked to Pigment Accumulation in Tomato and Pepper Fruits. PLANT Physiol. **161**: 1476–1485.

Park, S.Y. et al. (2009). Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. Science (80-.). **324**: 1068–1071.

Pawelkowicz, M., Zielinski, K., Zielinska, D., Plader, W., Yagi, K., Wojcieszek, M., Siedlecka, E., Bartoszewski, G., Skarzynska, A., and Przybecki, Z. (2015). Next generation sequencing and omics in cucumber (Cucumis sativus L.) breeding directed research. Plant Sci. **242**: 77–88.

Peng, M. et al. (2016). Evolutionarily distinct BAHD N-acyltransferases are responsible for natural variation of aromatic amine conjugates in Rice. Plant Cell **28**: 1533–50.

Peralta, I., Spooner, D., and Knapp, S. (2008). Taxonomy of wild tomatoes and their relatives (Solanum sect. Lycopersicoides, sect. Juglandifolia, sect. Lycopersicon; Solanaceae). In Systematic Botany Monographs, p. 186.

Peterhansel, C. and Maurino, V.G. (2011). Photorespiration Redesigned. PLANT Physiol. **155**: 49–55.

Pires, N. and Dolan, L. (2010). Origin and diversification of basic-helix-loop-helix proteins in plants. Mol. Biol. Evol. **27**: 862–874.

Pogson, B. (1996). Arabidopsis Carotenoid Mutants Demonstrate That Lutein Is Not Essential for Photosynthesis in Higher Plants. PLANT CELL ONLINE **8**: 1627–1639.

Poppenberger, B., Rozhon, W., Khan, M., Husar, S., Adam, G., Luschnig, C., Fujioka, S., and Sieberer, T. (2011). CESTA, a positive regulator of brassinosteroid biosynthesis. EMBO J. **30**: 1149–1161.

Potokina, E., Druka, A., Luo, Z., Wise, R., Waugh, R., and Kearsey, M. (2008). Gene expression quantitative trait locus analysis of 16 000 barley genes reveals a complex pattern of genome-wide transcriptional regulation. Plant J. **53**: 90–101.

Powell, A.L.T. et al. (2012). Uniform ripening encodes a Golden 2-like transcription factor regulating tomato fruit chloroplast development. Science (80-.). **336**: 1711–1715.

Quadrana, L. et al. (2014). Natural occurring epialleles determine vitamin e accumulation in tomato fruits. Nat. Commun. **5**.

Quail, P.H. (2002). Phytochrome photosensory signalling networks. Nat. Rev. Mol. Cell Biol. **3**: 85–93.

Quattrocchio, F. (1993). Regulatory Genes Controlling Anthocyanin Pigmentation Are Functionally Conserved among Plant Species and Have Distinct Sets of Target Genes. PLANT CELL ONLINE **5**: 1497–1512.

Ray, J., Moureau, P., Bird, C., Bird, A., Grierson, D., Maunders, M., Truesdale, M., Bramley, P., and Schuch, W. (1992). Cloning and characterization of a gene involved in phytoene

synthesis from tomato. Plant Mol. Biol. **19**: 401–404.

Riaño-Pachón, D.M., Ruzicic, S., Dreyer, I., and Mueller-Roeber, B. (2007). PlnTFDB: An integrative plant transcription factor database. BMC Bioinformatics **8**.

Rice-Evans, C.A., Sampson, J., Bramley, P.M., and Holloway, D.E. (1997). Why do we expect carotenoids to be antioxidants in vivo? Free Radic. Res. **26**: 381–398.

Riechmann, J.L. et al. (2000). Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. Science (80-.). **290**: 2105–2110.

Riedelsheimer, C., Lisec, J., Czedik-eysenberg, A., Sulpice, R., Flis, A., Grieder, C., Altmann, T., Stitt, M., Willmitzer, L., and Melchinger, A.E. (2012). Genome-wide association mapping of leaf metabolic pro fi les for dissecting complex traits in maize. Proc. Natl. Acad. Sci. **109**: 8872–8877.

Robinson, K. a, Koepke, J.I., Kharodawala, M., and Lopes, J.M. (2000). A network of yeast basic helix-loop-helix interactions. Nucleic Acids Res. **28**: 4460–6.

Rodrigo, M.J., Alquézar, B., Alós, E., Medina, V., Carmona, L., Bruno, M., Al-Babili, S., and Zacarías, L. (2013). A novel carotenoid cleavage activity involved in the biosynthesis of Citrus fruit-specific apocarotenoid pigments. J. Exp. Bot. **64**: 4461–4478.

Rodriguez-Concepcion, M. (2002). Elucidation of the Methylerythritol Phosphate Pathway for Isoprenoid Biosynthesis in Bacteria and Plastids. A Metabolic Milestone Achieved through Genomics. PLANT Physiol. **130**: 1079–1089.

Rodríguez-Concepción, M. (2010). Supply of precursors for carotenoid biosynthesis in plants. Arch. Biochem. Biophys. **504**: 118–122.

Rodriguez-Concepcion, M. and Stange, C. (2013). Biosynthesis of carotenoids in carrot: An underground story comes to light. Arch. Biochem. Biophys. **539**: 110–116.

Rodriguez-Uribe, L., Guzman, I., Rajapakse, W., Richins, R.D., and O'Connell, M.A. (2012). Carotenoid accumulation in orange-pigmented Capsicum annuum fruit, regulated at multiple levels. J. Exp. Bot. **63**: 517–526.

Ronen, G., Carmel-Goren, L., Zamir, D., and Hirschberg, J. (2000a). 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. **97**: 11102–11107.

Ronen, G., Carmel-Goren, L., Zamir, D., and Hirschberg, J. (2000b). 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. **97**: 11102–11107.

Ronen, G., Cohen, M., Zamir, D., and Hirschberg, J. (1999). 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. **17**: 341–351.

Rose, J.K.C., Bashir, S., Giovannoni, J.J., Jahn, M.M., and Saravanan, R.S. (2004). Tackling the plant proteome: Practical approaches, hurdles and experimental tools. Plant J. **39**: 715–733.

Rosenwasser, S., Mausz, M.A., Schatz, D., Sheyn, U., Malitsky, S., Aharoni, A., Weinstock, E., Tzfadia, O., Ben-Dor, S., Feldmesser, E., Pohnert, G., and Vardi, A. (2014). Rewiring Host Lipid Metabolism by Large Viruses Determines the Fate of Emiliania huxleyi, a Bloom-Forming Alga in the Ocean. Plant Cell **26**: 2689–2707.

Rottmann, W.H., Peter, G.F., Oeller, P.W., Keller, J.A., Shen, N.F., Nagy, B.P., Taylor, L.P., Campbell, A.D., and Theologis, A. (1991). 1-Aminocyclopropane-1-carboxylate synthase in tomato is encoded by a multigene family whose transcription is induced during fruit and floral senescence. J. Mol. Biol. **222**: 937–961.

Rousseaux, M.C., Jones, C.M., Adams, D., Chetelat, R., Bennett, A., and Powell, A. (2005). QTL analysis of fruit antioxidants in tomato using *Lycopersicon pennellii* introgression lines. Theor. Appl. Genet. **111**: 1396–1408.

Rushton, P.J., Bokowiec, M.T., Han, S., Zhang, H., Brannock, J.F., Chen, X., Laudeman, T.W., and Timko, M.P. (2008). Tobacco Transcription Factors: Novel Insights into Transcriptional Regulation in the Solanaceae. PLANT Physiol. **147**: 280–295.

Sagar, M. et al. (2013). SlARF4, an auxin response factor involved in the control of sugar metabolism during tomato fruit development. Plant Physiol. **161**: 1362–74.

Sarrion-Perdigones, A., Falconi, E.E., Zandalinas, S.I., Juárez, P., Fernández-del-Carmen, A., Granell, A., and Orzaez, D. (2011). GoldenBraid: An iterative cloning system for

standardized assembly of reusable genetic modules. PLoS One **6**.

Sarrion-Perdigones, A., Vazquez-Vilar, M., Palaci, J., Castelijns, B., Forment, J., Ziarsolo, P., Blanca, J., Granell, A., and Orzaez, D. (2013). GoldenBraid 2.0: A Comprehensive DNA Assembly Framework for Plant Synthetic Biology. PLANT Physiol. **162**: 1618–1631.

Satia, J.A., Littman, A., Slatore, C.G., Galanko, J.A., and White, E. (2009). Long-term use of β-carotene, retinol, lycopene, and lutein supplements and lung cancer risk: Results from the vitamins and lifestyle (vital) study. Am. J. Epidemiol. **169**: 815–828.

Schauer, N. et al. (2006). Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nat. Biotechnol. **24**: 447–454.

Schauer, N., Semel, Y., Balbo, I., Steinfath, M., Repsilber, D., Selbig, J., Pleban, T., Zamir, D., and Fernie, A.R. (2008). Mode of Inheritance of Primary Metabolic Traits in Tomato. PLANT CELL ONLINE **20**: 509–523.

Scheible, W.R. and Pauly, M. (2004). Glycosyltransferases and cell wall biosynthesis: Novel players and insights. Curr. Opin. Plant Biol. **7**: 285–295.

Seetang-Nun, Y., Sharkey, T.D., and Suvachittanont, W. (2008). Molecular cloning and characterization of two cDNAs encoding 1-deoxy-d-xylulose 5-phosphate reductoisomerase from Hevea brasiliensis. J. Plant Physiol. **165**: 991–1002.

Semel, Y., Nissenbaum, J., Menda, N., Zinder, M., Krieger, U., Issman, N., Pleban, T., Lippman, Z., Gur, A., and Zamir, D. (2006). Overdominant quantitative trait loci for yield and fitness in tomato. Proc. Natl. Acad. Sci. **103**: 12981–12986.

Sesso, H.D., Buring, J.E., Norkus, E.P., and Gaziano, J.M. (2004). Plasma lycopene, other carotenoids, and retinol and the risk of cardiovascular disease in women. Am. J. Clin. Nutr. **79**: 47–53.

Seymour, G.B., Østergaard, L., Chapman, N.H., Knapp, S., and Martin, C. (2013). Fruit Development and Ripening. Annu. Rev. Plant Biol. **64**: 219–241.

Shah, P., Powell, A.L.T., Orlando, R., Bergmann, C., and Gutierrez-Sanchez, G. (2012). Proteomic analysis of ripening tomato fruit infected by botrytis cinerea. J. Proteome

Res. **11**: 2178–2192.

Sharlach, M. et al. (2013). Fine genetic mapping of RXopJ4, a bacterial spot disease resistance locus from Solanum pennellii LA716. Theor. Appl. Genet. **126**: 601–609.

Sharma, S.K. et al. (2013). Construction of Reference Chromosome-Scale Pseudomolecules for Potato: Integrating the Potato Genome with Genetic and Physical Maps. G3: Genes|Genomes|Genetics **3**: 2031–2047.

Shima, Y., Fujisawa, M., Kitagawa, M., Nakano, T., Kimbara, J., Nakamura, N., Shiina, T., Sugiyama, J., Nakamura, T., Kasumi, T., and Ito, Y. (2014). Tomato FRUITFULL homologs regulate fruit ripening via ethylene biosynthesis. Biosci. Biotechnol. Biochem. **78**: 231–237.

Singh, K.B., Foley, R.C., and Oñate-Sánchez, L. (2002). Transcription factors in plant defense and stress responses. Curr. Opin. Plant Biol. **5**: 430–436.

Singh, M., Gupta, A., Singh, D., Khurana, J.P., and Laxmi, A. (2017). Arabidopsis RSS1 Mediates Cross-Talk between Glucose and Light Signaling during Hypocotyl Elongation Growth. Sci. Rep. **7**.

Slavov, G.T., Nipper, R., Robson, P., Farrar, K., Allison, G.G., Bosch, M., Clifton-Brown, J.C., Donnison, I.S., and Jensen, E. (2014). Genome-wide association studies and prediction of 17 traits related to phenology, biomass and cell wall composition in the energy grass Miscanthus sinensis. New Phytol. **201**: 1227–1239.

Song, X.-M., Huang, Z.-N., Duan, W.-K., Ren, J., Liu, T.-K., Li, Y., and Hou, X.-L. (2014). Genome-wide analysis of the bHLH transcription factor family in Chinese cabbage (Brassica rapa ssp. pekinensis). Mol. Genet. Genomics **289**: 77–91.

Species, R.O., Finkel, T., and Species, R.O. (2001). Reactive oxygen species and signal transduction. IUBMB Life **52**: 3–6.

Spelt, C., Quattrocchio, F., Mol, J., and Koes, R. (2002). Anthocyanin1 of petunia controls pigment synthesis, vacuolar pH, and seed coat development by genetically distinct mechanisms. Plant Cell **14**: 2121–2135.

Stahl, W., Heinrich, U., Wiseman, S., Eichler, O., Sies, H., and Tronnier, H. (2001). Dietary

tomato paste protects against ultraviolet light-induced erythema in humans. J. Nutr. **131**: 1449–51.

Stahl, W. and Sies, H. (2003). Antioxidant activity of carotenoids. Mol. Aspects Med. **24**: 345–351.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30**: 1312–1313.

Stevens, M.A. and Rick, C.M. (1986). Genetics and breeding. In The tomato crop: A scientific basis for improvement., pp. 35–109.

Stigliani, A.L., Giorio, G., and D'Ambrosio, C. (2011). Characterization of P450 Carotenoid β- And -hydroxylases of tomato and transcriptional regulation of xanthophyll biosynthesis in root, leaf, petal and fruit. Plant Cell Physiol. **52**: 851–865.

Stracke, R., Werber, M., and Weisshaar, B. (2001). The R2R3-MYB gene family in Arabidopsis thaliana. Curr. Opin. Plant Biol. **4**: 447–456.

Su, L., Diretto, G., Purgatto, E., Danoun, S., Zouine, M., Li, Z., Roustan, J.P., Bouzayen, M., Giuliano, G., and Chervin, C. (2015). Carotenoid accumulation during tomato fruit ripening is modulated by the auxin-ethylene balance. BMC Plant Biol. **15**.

Svistoonoff, S., Creff, A., Reymond, M., Sigoillot-Claude, C., Ricaud, L., Blanchet, A., Nussaume, L., and Desnos, T. (2007). Root tip contact with low-phosphate media reprograms plant root architecture. Nat. Genet. **39**: 792–796.

Tanumihardjo, S.A. (2013). Carotenoids and human health.

Teng, S. (2005). Sucrose-Specific Induction of Anthocyanin Biosynthesis in Arabidopsis Requires the MYB75/PAP1 Gene. PLANT Physiol. **139**: 1840–1852.

Tian, L. and Dellapenna, D. (2001). Characterization of a second carotenoid β-hydroxylase gene from Arabidopsis and its relationship to the LUT1 locus. Plant Mol. Biol. **47**: 379–388.

Tieman, D.M. and Klee, H.J. (1999). Differential expression of two novel members of the tomato ethylene-receptor family. Plant Physiol. **120**: 165–172.

Tieman, D.M., Zeigler, M., Schmelz, E.A., Taylor, M.G., Bliss, P., Kirst, M., and Klee, H.J.

(2006). Identification of loci affecting flavour volatile emissions in tomato fruits. J. Exp. Bot. **57**: 887–896.

Todesco, M. et al. (2010). Natural allelic variation underlying a major fitness trade-off in Arabidopsis thaliana. Nature **465**: 632–636.

Toledo-Ortiz, G. (2003). The Arabidopsis Basic/Helix-Loop-Helix Transcription Factor Family. PLANT CELL ONLINE **15**: 1749–1770.

Toledo-Ortiz, G., Huq, E., and Rodríguez-Concepción, M. (2010). Direct regulation of phytoene synthase gene expression and carotenoid biosynthesis by phytochrome-interacting factors. Proc. Natl. Acad. Sci. U. S. A. **107**: 11626–31.

Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485**: 635–41.

Torbergsen, a C. and Collins, a R. (2000). Recovery of human lymphocytes from oxidative DNA damage; the apparent enhancement of DNA repair by carotenoids is probably simply an antioxidant effect. Eur. J. Nutr. **39**: 80–85.

Tzuri, G. et al. (2015). A "golden" SNP in CmOr governs the fruit flesh color of melon (Cucumis melo). Plant J. **82**: 267–279.

Ulmasov, T., Hagen, G., and Guilfoyle, T.J. (1999). Dimerization and DNA binding of auxin response factors. Plant J. **19**: 309–319.

Uozumi, A., Ikeda, H., Hiraga, M., Kanno, H., Nanzyo, M., Nishiyama, M., Kanahama, K., and Kanayama, Y. (2012). Tolerance to salt stress and blossom-end rot in an introgression line, IL8-3, of tomato. Sci. Hortic. (Amsterdam). **138**: 1–6.

Upadhyay, S.K., Kumar, J., Alok, a., and Tuli, R. (2013). RNA Guided Genome Editing for Target Gene Mutations in Wheat. G3 Genes|Genomes|Genetics **3**: 2233–2238.

Valero, M.A., Vidal, A., Burgos, R., Calvo, F.L., Martinez, C., Luengo, L.M., and Cuerda, C. (2011). [Meta-analysis on the role of lycopene in type 2 diabetes mellitus]. Nutr Hosp **26**: 1236–1241.

Vardhini, B.V. and Rao, S.S.R. (2002). Acceleration of ripening of tomato pericarp discs by brassinosteroids. Phytochemistry **61**: 843–847.

Virtamo, J., Pietinen, P., Huttunen, J.K., Korhonen, P., Malila, N., Virtanen, M.J., Albanes, D., Taylor, P.R., Albert, P., and ATBC Study Group (2003). Incidence of cancer and mortality following alpha-tocopherol and beta-carotene supplementation: a postintervention follow-up. JAMA **290**: 476–85.

Vogel, J.T., Tieman, D.M., Sims, C.A., Odabasi, A.Z., Clark, D.G., and Klee, H.J. (2010). Carotenoid content impacts flavor acceptability in tomato (Solanum lycopersicum). J. Sci. Food Agric. **90**: 2233–2240.

Vrebalov, J., Pan, I.L., Arroyo, A.J.M., McQuinn, R., Chung, M., Poole, M., Rose, J., Seymour, G., Grandillo, S., Giovannoni, J., and Irish, V.F. (2009). Fleshy Fruit Expansion and Ripening Are Regulated by the Tomato SHATTERPROOF Gene TAGL1. Plant Cell **21**: 3041–3062.

Vrebalov, J., Ruezinsky, D., Padmanabhan, V., White, R., Medrano, D., Drake, R., Schuch, W., and Giovannoni, J. (2002). A MADS-box gene necessary for fruit ripening at the tomato ripening-inhibitor (rin) locus. Science (80-.). **296**: 343–346.

Walter, M.H. and Strack, D. (2011). Carotenoids and their cleavage products: Biosynthesis and functions. Nat. Prod. Rep. **28**: 663.

Wang, C.Y. and Chen, B.H. (2006). Tomato pulp as source for the production of lycopene powder containing high proportion of cis-isomers. Eur. Food Res. Technol. **222**: 347–353.

Wang, H., Nussbaum-Wagler, T., Li, B., Zhao, Q., Vigouroux, Y., Faller, M., Bomblies, K., Lukens, L., and Doebley, J.F. (2005). The origin of the naked grains of maize. Nature **436**: 714–719.

Wang, J., Hu, Z., Zhao, T., Yang, Y., Chen, T., Yang, M., Yu, W., and Zhang, B. (2015). Genome-wide analysis of bHLH transcription factor and involvement in the infection by yellow leaf curl virus in tomato (Solanum lycopersicum). BMC Genomics **16**.

Wang, L., Liu, S., Manson, J.E., Gaziano, J.M., Buring, J.E., and Sesso, H.D. (2006). The consumption of lycopene and tomato-based food products is not associated with the risk of type 2 diabetes in women. J Nutr **136**: 620–625.

Wei, Y., Wan, H., Wu, Z., Wang, R., Ruan, M., Ye, Q., Li, Z., Zhou, G., Yao, Z., and Yang, Y.

(2016). A Comprehensive Analysis of Carotenoid Cleavage Dioxygenases Genes in Solanum Lycopersicum. Plant Mol. Biol. Report. **34**: 512–523.

Wen, W., Li, D., Li, X., Gao, Y., Li, W., Li, H., Liu, J., Liu, H., Chen, W., Luo, J., and Yan, J. (2014). Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. Nat. Commun. **5**.

Werner, J.D., Borevitz, J.O., Warthmann, N., Trainer, G.T., Ecker, J.R., Chory, J., and Weigel, D. (2005). Quantitative trait locus mapping and DNA array hybridization identify an FLM deletion as a cause for natural flowering-time variation. Proc. Natl. Acad. Sci. **102**: 2460–2465.

West, K.P. (2003). Vitamin A deficiency disorders in children and women. Food Nutr. Bull. **24**.

West, M.A., Kim, K., Kliebenstein, D.J., van Leeuwen, H., Michelmore, R.W., Doerge, R.W., and St Clair, D.A. (2007). Global eQTL Mapping Reveals the Complex Genetic Architecture of Transcript-Level Variation in Arabidopsis. Genetics **175**: 1441–1450.

Wilkinson, J.Q., Lanahan, M.B., Yen, H.-C., Giovannoni, J.J., and Klee, H.J. (1995). An Ethylene-Inducible Component of Signal Transduction Encoded by Never-ripe. Science (80-.). **270**: 1807–1809.

Winklhofer-Roob, B.M., Puhl, H., Khoschsorur, G., van't Hof, M.A., Esterbauer, H., and Shmerling, D.H. (1995). Enhanced resistance to oxidation of low density lipoproteins and decreased lipid peroxide formation during beta-carotene supplementation in cystic fibrosis.

Wurtzel, E.T. and Kutchan, T.M. (2016). Plant metabolism, the diverse chemistry set of the future. Science (80-.). **353**: 1232–1236.

Xie, K. and Yang, Y. (2013). RNA-Guided genome editing in plants using a CRISPR-Cas system. Mol. Plant **6**: 1975–1983.

Yamagishi, M., Kishimoto, S., and Nakayama, M. (2010). Carotenoid composition and changes in expression of carotenoid biosynthetic genes in tepals of Asiatic hybrid lily. Plant Breed. **129**: 100–107.

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

Yip, W.K., Moore, T., and Yang, S.F. (1992). Differential accumulation of transcripts for four tomato 1-aminocyclopropane-1-carboxylate synthase homologs under various conditions. Proc. Natl. Acad. Sci. **89**: 2475–2479.

Young, A.J. and Lowe, G.M. (2001). Antioxidant and prooxidant properties of carotenoids. Arch. Biochem. Biophys. **385**: 20–27.

Yuan, J.-M., Stram, D.O., Arakawa, K., Lee, H.-P., and Yu, M.C. (2003). Dietary Cryptoxanthin and Reduced Risk of Lung Cancer. Cancer Epidemiol. Biomarkers Prev. **12**: 890–898.

Zanor, M.I. et al. (2009). RNA Interference of LIN5 in Tomato Confirms Its Role in Controlling Brix Content, Uncovers the Influence of Sugars on the Levels of Fruit Hormones, and Demonstrates the Importance of Sucrose Cleavage for Normal Fruit Development and Fertility. Plant Physiol. **150**: 1204–1218.

Zhang, L., Fetch, T., Nirmala, J., Schmierer, D., Brueggeman, R., Steffenson, B., and Kleinhofs, A. (2006). Rpr1, a gene required for Rpg1-dependent resistance to stem rust in barley. Theor. Appl. Genet. **113**: 847–855.

Zhang, P. and Omaye, S.. (2001). Antioxidant and prooxidant roles for β-carotene, α-tocopherol and ascorbic acid in human lung cells. Toxicol. Vitr. **15**: 13–24.

Zhang, P. and Omaye, S.T. (2000). Beta-carotene and protein oxidation: effects of ascorbic acid and alpha-tocopherol. Toxicology **146**: 37–47.

Zhang, S., Tang, G., Russell, R.M., Mayzel, K.A., Stampfer, M.J., Willett, W.C., and Hunter, D.J. (1997). Measurement of retinoids and carotenoids in breast adipose tissue and a comparison of concentrations in breast cancer cases and control subjects. Am. J. Clin. Nutr. **66**: 626–632.

Zhang, Z., Liu, L., Zhang, M., Zhang, Y., and Wang, Q. (2014). Effect of carbon dioxide enrichment on health-promoting compounds and organoleptic properties of tomato fruits grown in greenhouse. Food Chem. **153**: 157–163.

Zhong, S., Fei, Z., Chen, Y.R., Zheng, Y., Huang, M., Vrebalov, J., McQuinn, R., Gapper, N., Liu, B., Xiang, J., Shao, Y., and Giovannoni, J.J. (2013). Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. Nat. Biotechnol. **31**: 154–159.

Zhong, S., Joung, J.G., Zheng, Y., Chen, Y.R., Liu, B., Shao, Y., Xiang, J.Z., Fei, Z., and Giovannoni, J.J. (2011). High-throughput illumina strand-specific RNA sequencing library preparation. Cold Spring Harb. Protoc. **6**: 940–949.

Zhou, X., Welsch, R., Yang, Y., Álvarez, D., Riediger, M., Yuan, H., Fish, T., Liu, J., Thannhauser, T.W., and Li, L. (2015). *Arabidopsis* OR proteins are the major posttranscriptional regulators of phytoene synthase in controlling carotenoid biosynthesis. Proc. Natl. Acad. Sci. **112**: 3558–3563.

Zhu, C., Bai, C., Sanahuja, G., Yuan, D., Farré, G., Naqvi, S., Shi, L., Capell, T., and Christou, P. (2010). The regulation of carotenoid pigmentation in flowers. Arch. Biochem. Biophys. **504**: 132–141.

Zhu, G. et al. (2018). Rewiring of the Fruit Metabolome in Tomato Breeding. Cell **172**: 249–255.e12.

Zhu, T., Tan, W.R., Deng, X.G., Zheng, T., Zhang, D.W., and Lin, H.H. (2015). Effects of brassinosteroids on quality attributes and ethylene synthesis in postharvest tomato fruit. Postharvest Biol. Technol. **100**: 196–204.

Appendix 1 Vectors and Plasmids

Name	Supplier/Reference	Selective Antibiotics	Description
pDONR207	Invitrogen	$Gent^R$	Donor vector for Gateway cloning
pDONR221	Invitrogen	Kan^R, Cm^R	Donor vector for Gateway cloning
pDEST32	Invitrogen	$Gent^R$, Cm^R	Destination vecotr with Gateway LR site
pDEST22	Invitrogen	Amp^R,Cm^R	Destination vecotr with Gateway LR site
pTRV1	Orzaez et al., 2009	Kan^R	VIGS
pTRV2	Orzaez et al., 2009	Kan^R	VIGS
pTRV2-GW	Orzaez et al., 2009	Kan^R, Cm^R	VIGS
pTRV2-Del/Ros1-GW	Orzaez et al., 2009	Kan^R, Cm^R	VIGS
pBin19-35S-GW	Lab stock	Kan^R, Cm^R	Plant Stable Transformation
pBin19-E8-GW	In this thesis	Kan^R, Cm^R	Plant Stable Transformation
pDEST-GADT-GW	Huang et al., 2014	Kan^R, Cm^R	Yeast two-hybrid assay, AD
pDEST-GBKT-GW	Huang et al., 2014	Kan^R, Cm^R	Yeast two-hybrid assay, BD
pTRV2-Del/Ros1	Orzaez et al., 2009	Kan^R	VIGS
pTRV2-Del/Ros1-SlHONG	This thesis	Kan^R	VIGS
pBin19-35S-SlHONG	This thesis	Kan^R	Tomato stable transformation
pBin19-E8-SlHONG	This thesis	Kan^R	Tomato stable transformation

pDEST-GADT-Solyc03g120530	This thesis	Amp^R	Yeast two-hybrid assay
pDEST-GADT-SlbHLH146	This thesis	Amp^R	Yeast two-hybrid assay
pDEST-GADT-SlbHLH147	This thesis	Amp^R	Yeast two-hybrid assay
pDEST-GADT-SlbHLH090	This thesis	Amp^R	Yeast two-hybrid assay
pDEST-GADT-Solyc08g082940	This thesis	Amp^R	Yeast two-hybrid assay
pDEST-GADT-Solyc01g096050	This thesis	Amp^R	Yeast two-hybrid assay
pDEST-GADT-SlbHLH060	This thesis	Amp^R	Yeast two-hybrid assay
pDEST-GADT-SlbHLH043	This thesis	Amp^R	Yeast two-hybrid assay
pDEST-GADT-SlHONG	This thesis	Amp^R	Yeast two-hybrid assay
pDEST-GADT-SlTCP2	This thesis	Amp^R	Yeast two-hybrid assay
pDEST-GADT-SpHONG	This thesis	Amp^R	Yeast two-hybrid assay
pDEST-GADT-SlbHLH071	This thesis	Amp^R	Yeast two-hybrid assay
pDEST-GADT-SlbHLH026	This thesis	Amp^R	Yeast two-hybrid assay
pDEST-GADT-SlbHLH020	This thesis	Amp^R	Yeast two-hybrid assay
pDEST-GBKT-Solyc03g120530	This thesis	Kan^R	Yeast two-hybrid assay
pDEST-GBKT-SlbHLH146	This thesis	Kan^R	Yeast two-hybrid assay
pDEST-GBKT-SlbHLH147	This thesis	Kan^R	Yeast two-hybrid assay
pDEST-GBKT-SlbHLH090	This thesis	Kan^R	Yeast two-hybrid assay
pDEST-GBKT-Solyc08g082940	This thesis	Kan^R	Yeast two-hybrid assay
pDEST-GBKT-Solyc01g096050	This thesis	Kan^R	Yeast two-hybrid assay
pDEST-GBKT-SlbHLH060	This thesis	Kan^R	Yeast two-hybrid assay

pDEST-GBKT-SlbHLH043	This thesis	Kan^R	Yeast two-hybrid assay
pDEST-GBKT-SlHONG	This thesis	Kan^R	Yeast two-hybrid assay
pDEST-GBKT-SlTCP2	This thesis	Kan^R	Yeast two-hybrid assay
pDEST-GBKT-SpHONG	This thesis	Kan^R	Yeast two-hybrid assay
pDEST-GBKT-SlbHLH071	This thesis	Kan^R	Yeast two-hybrid assay
pDEST-GBKT-SlbHLH026	This thesis	Kan^R	Yeast two-hybrid assay
pDEST-GBKT-SlbHLH020	This thesis	Kan^R	Yeast two-hybrid assay
pDGB2_alpha1	GB Database	Kan^R	Golden Braid Cloning
pDGB2_alpha2	GB Database	Kan^R	Golden Braid Cloning
pDGB2_omega1	GB Database	$Spec^R$	Golden Braid Cloning
pDGB2_omega2	GB Database	$Spec^R$	Golden Braid Cloning
pccdB	GB Database	Amp^R	Transactivation assay
pRenilla	GB Database	Amp^R	Transactivation assay
pLuciferase	GB Database	Amp^R	Transactivation assay
pP35S	GB Database	Amp^R	Transactivation assay
pTnos	GB Database	Amp^R	Transactivation assay
pP19	GB Database	Amp^R	Transactivation assay
pEGB2-Ω1-35s:SlHONG:Tnos-proSlDXS1:LUC:Tnos	This thesis	$Spec^R$	Transactivation assay
pEGB2-Ω1-35s:SlHONG:Tnos-proSlPDS:LUC:Tnos	This thesis	$Spec^R$	Transactivation assay
pEGB2-Ω1-35s:SlHONG:Tnos-proSlPSY1:LUC:Tnos	This thesis	$Spec^R$	Transactivation assay
pEGB2-Ω1-35s:SlHONG:Tnos-proSlZISO:LUC:Tnos	This thesis	$Spec^R$	Transactivation assay

pEGB2-Ω1-35s:SlHONG:Tnos-proSlCRTISO:LUC:Tnos	This thesis	Spec^R	Transactivation assay
pEGB2-Ω1-35s:SlHONG:Tnos-proSlCYCB:LUC:Tnos	This thesis	Spec^R	Transactivation assay
pEGB2-Ω1-35s:SpHONG:Tnos-proSlDXS1:LUC:Tnos	This thesis	Spec^R	Transactivation assay
pEGB2-Ω1-35s:SpHONG:Tnos-proSlPDS:LUC:Tnos	This thesis	Spec^R	Transactivation assay
pEGB2-Ω1-35s:SpHONG:Tnos-proSlPSY1:LUC:Tnos	This thesis	Spec^R	Transactivation assay
pEGB2-Ω1-35s:SpHONG:Tnos-proSlZISO:LUC:Tnos	This thesis	Spec^R	Transactivation assay
pEGB2-Ω1-35s:SpHONG:Tnos-proSlCRTISO:LUC:Tnos	This thesis	Spec^R	Transactivation assay
pEGB2-Ω1-35s:SpHONG:Tnos-proSlCYCB:LUC:Tnos	This thesis	Spec^R	Transactivation assay
pEGB2-Ω1-35s:SlRIN:Tnos-proSlPSY1:LUC:Tnos	This thesis	Spec^R	Transactivation assay
pEGB2-Ω1-35s:SlRIN:Tnos-proSlHONG:LUC:Tnos	This thesis	Spec^R	Transactivation assay
pEGB2-Ω1-35s:SlHONG:Tnos-proSlPSY1-f1:LUC:Tnos	This thesis	Spec^R	Transactivation assay
pEGB2-Ω1-35s:SlHONG:Tnos-proSlPSY1-f2:LUC:Tnos	This thesis	Spec^R	Transactivation assay
pEGB2-Ω1-35s:SlHONG:Tnos-proSlHONG:LUC:Tnos	This thesis	Spec^R	Transactivation assay
pEGB2-Ω1-35s:SlRIN:Tnos-proSlRIN:LUC:Tnos	This thesis	Spec^R	Transactivation assay
pEGB2-Ω1-35s:SlHONG:Tnos-proSlRIN:LUC:Tnos	This thesis	Spec^R	Transactivation assay
pEGB2-Ω1-35s:SlRIN:Tnos-proSlPDS:LUC:Tnos	This thesis	Spec^R	Transactivation assay
pEGB2-Ω1-35s:SlRIN:Tnos-proSlHONG-f1:LUC:Tnos	This thesis	Spec^R	Transactivation assay
pEGB2-Ω12-3xSF-35S:Renilla:Tnos-35S:P19:Tnos	This thesis	Spec^R	Transactivation assay
ec15029-PNOS-KAN	TSL SynBio	Carb^R	CRISPR/Cas9
ec41744-ELE2	TSL SynBio	Spec^R	Golden Gate Cloning
ec41766-ELE3	TSL SynBio	Spec^R	Golden Gate Cloning

ec41780-ELE4	TSL SynBio	Spec^{R}	Golden Gate Cloning
ec47732-pL1-F1	TSL SynBio	Carb^{R}	Golden Gate Cloning
ec47742-pL1-F2	TSL SynBio	Carb^{R}	Golden Gate Cloning
ec47751-pL1-F3	TSL SynBio	Carb^{R}	Golden Gate Cloning
ec47761-pL1-F4	TSL SynBio	Carb^{R}	Golden Gate Cloning
pL1V-F2-TU-Cas9-pICSL11050	Carbenicilin^{R}	Carb^{R}	CRISPR/Cas9
pL2V-sgRNA4A-sgRNA4B-Cas9	This thesis	Kan^{R}	CRISPR/Cas9

Appendix 2 Recipes of media used in this thesis

1. Luria-Bertani (LB) medium

1 litre LB medium contains

Tryptone	10 g
Yeast Extract	5 g
NaCl	15 g
For solid medium:	
Agar	10 g

2. Media for tomato stable transformation

1 litre seed germination medium contains:

pH 5.8	(KOH)
Ms+vitamins medium	4.4 g
Agarose	6 g

1 litre regeneration medium contains

pH 6.0	(KOH)
MS+Nitch's vitamins	4.4 g
myo-inositol	100 mg
Sucrose	20 g
Agargel	4 g
Zeatin Riboside (trans isomer)	2 mg
Timentin	320 mg
Kanamycin	50-100 mg

1 litre rooting medium contains

pH 6.0	(KOH)
MS+vitamins	2.2 g
Sucrose	5 g
Gelrite	2.25 g
Timentin	320 mg

Kanamycin	50 mg

3. PDA medium

1 litre PDA medium contains:

Potatoes, Infusion	200 g
Dextrose	20 g
Agar	15 g

4. TY medium

1 litre TY medium contains

Typtone	5 g
Yeast Extract	3 g
$CaCl_2 \cdot 6H_20$	1.325 g
For solid medium:	
Agar	10 g

5. Media for yeast two-hybrid assays

1 litre YPD medium contains:

pH 6.5	
Difco pepetone	20 g
Yeast extract	10 g
For solid medium:	
Agar	20 g

1 litre SD medium contains:

pH 5.8	
Yeast nitrogen base without animo acids	6.7 g
3 AT	Depends on yeast strain
For solid medium:	
Agar	20 g

Appendix 3 Primers used in this thesis

Gene/Plasmid	Locus	Name	Purpose	Sequences
pDONR207	pDONR207	pDONR207-seq-F	Sequencing	TCGCGTTAACGCTAGCATGGATCTC
		pDONR207-seq-R	Sequencing	GTAACATCAGAGATTTTGAGACAC
pUPD	pUPD	pUPD-seq-F	Sequencing	AATTGGGCCCGACGTCGCAT
		pUPD-seq-R	Sequencing	AGCTATGCATCCAACGCGTT
pDONR221	pDONR221	pDONR221-seq-F	Sequencing	GTTTTCCCAGTCACGAC
		pDONR221-seq-R	Sequencing	CAGGAAACAGCTATGAC
SlHONG	*Solyc02g062690*	SlHONG-q-F	RT-q-PCR	CCGGGGAATGATTCAAAGGA
		SlHONG-q-R	RT-q-PCR	GCTATCAGTAGCTTGGCCACGA
		SlHONG-V-attB1	GW Cloning	GGGGACAAGTTTGTACAAAAAAGCAGGCTTATGAATCCTAGGCTTGATTTGAA
		SlHONG-V-attB2	GW Cloning	GGGGACCACTTTGTACAAGAAAGCTGGGTAAGATCAGGCTCAAAAGTTGGTT
SlDXS1	*Solyc01g067890*	SlDXS1-q-F	RT-q-PCR	CGGCTTACAAGTAACAGTTGCAGAT
		SlDXS1-q-R	RT-q-PCR	TGATCCTTCTTCGACAGTGATTAGC
SlPSY1	*Solyc03g031860*	SlPSY1-q-F	RT-q-PCR	TGGCCCAAACGCATCATATA
		SlPSY1-q-R	RT-q-PCR	CACCATCGAGCATGTCAAATG
SlPDS	*Solyc03g123760*	SlPDS-q-F	RT-q-PCR	CTCACTGCTCAGTGTGTATGCTGAC
		SlPDS-q-R	RT-q-PCR	CGAGATATCCACTCTTCTGCAGGT
SlCYCB	*Solyc06g074240*	SlCYCB-q-F	RT-q-PCR	GATCCTAAATACTGGCAAGGGTTCC
		SlCYCB-q-R	RT-q-PCR	ACCTAGTCATGTTTGAGCCATGTCC
SlCRTISO	*Solyc10g081650*	SlCRTISO-q-F	RT-q-PCR	AGGGACTCTCTCCGAAAGACTATGA
		SlCRTISO-q-R	RT-q-PCR	TAGATGACTTAAGCCCTGGGAAGAG

SlZISO	*Solyc12g098710*	SlZISO-q-F	RT-q-PCR	TTTACTAGAGGTAGCGGCTGTTGAC
		SlZISO-q-R	RT-q-PCR	CGTGTGAGCTAAGCACCATATAACC
ACTIN	*Solyc03g078400*	SlACT-q-F	RT-q-PCR	GGGGGCTATGAATGCACGGT
		SlACT-q-F	RT-q-PCR	GGCAATGCATCAGGCACCTC
SlPSY1	*Solyc03g031860*	proSlPSY1-g-F	PCR	CAACGACGATAACAAGGATTAGG
		proSlPSY1-g-R	PCR	CACCACACCTATCATATGCTTCA
		proSlPSY1-attB1	GW Cloning	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGGCTAAATCGAAAATTGAATCGT
		proSlPSY1-attB2	GW Cloning	GGGGACCACTTTGTACAAGAAAGCTGGGTATCTGAGCAAGAAAACCTTGGTTGTC
SlDXS1	*Solyc01g067890*	proSlDXS1-g-F	PCR	CTCACATGCCATTAAATTTTTGA
		proSlDXS1-g-R	PCR	AACTGCAGATCTGTTCCATGAAT
		proSlDXS1-attB1	GW Cloning	GGGGACAAGTTTGTACAAAAAAGCAGGCTTATGTCCATAAATTTAAAGAAGATTTCA
		proSlDXS1-attB2	GW Cloning	GGGGACCACTTTGTACAAGAAAGCTGGGTAGATTAGTCAATTCAACTG
SlPDS	*Solyc03g123760*	proSlPDS-g-F	PCR	CGGCATACTATGTTGTCAGGAGT
		proSlPDS-g-F	PCR	CTTATGACCCATTGATTCGCTAC
		proSlPDS-attB1	GW Cloning	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGGTAAAAGCTTCAATGCCCTATT
		proSlPDS-attB2	GW Cloning	GGGGACCACTTTGTACAAGAAAGCTGGGTATTTACTGAAAAATAACAGT
SlHONG	*Solyc02g062690*	proSlHONG-g-F	PCR	CGGTGCTAAGTTCATCTGAACG
		proSlHONG-g-R	PCR	ATGTTAAAATTCCCCTGAACGC
		proSlHONG-attB1	GW Cloning	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGTATTGACACTGAACAAGATAATTC
		proSlHONG-attB2	GW Cloning	GGGGACCACTTTGTACAAGAAAGCTGGGTAGTTTGCAAACAACTGTAACGA
SlRIN	*Solyc05g012020*	proSlRIN-g-F	PCR	GGAGGAACCACTTATATTACAAAAG
		proSlRIN-g-R	PCR	GTTCATAAGCTTTCTTTAGGAGTCC
		proSlRIN-attB1	GW Cloning	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGGAACTCAAATTTTATATGTGGTAG
		proSlRIN-attB2	GW Cloning	GGGGACCACTTTGTACAAGAAAGCTGGGTAATTGTATGAAGAAAAAAAATGTAAAAT
		SlRIN-CDS-attB1	GW Cloning	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGGTAGAGGGAAAGTAGAAT
		SlRIN-CDS-attB2	GW Cloning	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAAAGCATCCATCCAGGTACA

SlZISO	*Solyc12g098710*	proSlZISO-g-F	PCR	ATGATTCTCAGTGATCATCGGC
		proSlZISO-g-R	PCR	GTTGGTGGAGTGATATGCTATGG
		proSlZISO-attB1	GW Cloning	GGGGACAAGTTTGTACAAAAAAGCAGGCTTA GGAAGAAGCATGGTATTCAGTC
		proSlZISO-attB2	GW Cloning	GGGGACCACTTTGTACAAGAAAGCTGGGTA TTTTACAAGATTATGTGAGCTCTAAG
SlHONG	*Solyc02g062690*	E8-HONG-F	Genotyping	TCTAATCATTCACAGTGCCATCACA
		E8-HONG-R	Genotyping	TCACTTTCTTTCTCTGGTGGCG
		35S-HONG-F	Genotyping	AGGAGCATCGTGGAAAAAGA
		35S-HONG-R	Genotyping	CAGCCGGTAATCAAATGAGC
		SlHONG-E1-F	Genotyping	TTTTAACATGTTTGGCGGGG
		SlHONG-E2-R	Genotyping	ATGGCTATCAGTAGCTTGGC
		SlHONG-E3-F	Genotyping	TGACCACTACTTTTCCACCA
		SlHONG-E4-R	Genotyping	CGACACTGAATAAGCTCTGC
		SlHONG-F	Genotyping	ATGTTACAAATGAGTGTTCTTGAAAGAC
		SlHONG-R	Genotyping	CTAGTTAAATCCGACACTGAATAAGCTC
		SlHONG-KO-F	Genotyping	AGTCGTGATTGCATGGTGCGTT
		SlHONG-KO-F	Genotyping	AATGGCTGGATGCAACTTTGGCTA
Solyc03g120530	*Solyc03g120530*	Solyc03g120530_CDS_attB1	GW Cloning	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGCATCCACATGAATCTCATG
		Solyc03g120530_CDS_attB2	GW Cloning	GGGGACCACTTTGTACAAGAAAGCTGGGTATTATAAATATCCTCCTCCTCCAA
Solyc08g076820	*Solyc08g076820*	SlbHLH146_CDS_attB1	GW Cloning	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCTTTAGAAGCACTTTCTA
		SlbHLH146_CDS_attB2	GW Cloning	GGGGACCACTTTGTACAAGAAAGCTGGGTATTAAAGTGTAGCTGCTTCCTCC
Solyc08g076930	*Solyc08g076930*	SlbHLH147_CDS_attB1	GW Cloning	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGACTGAATACAGCTTGCCCA
		SlbHLH147_CDS_attB2	GW Cloning	GGGGACCACTTTGTACAAGAAAGCTGGGTATTAGTGTGTTTCAGCAATTTTC
Solyc08g081140	*Solyc08g081140*	SlbHLH090_CDS_attB1	GW Cloning	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCTATGGGACACCAAGATC
		SlbHLH090_CDS_attB2	GW Cloning	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAAGATTTCCATACTACTCTCT

Solyc08g082940	*Solyc08g082940*	Solyc08g082940_CDS-attB1	GW Cloning	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGGATCTGTGATTGATGAGA
		Solyc08g082940_CDS-attB2	GW Cloning	GGGGACCACTTTGTACAAGAAAGCTGGGTACTATCTTACACATCGGATCGTT
Solyc01g096050	*Solyc01g096050*	Solyc01g096050_CDS_attB1	GW Cloning	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGTTACTGGGAATATGTTGT
		Solyc01g096050_CDS_attB2	GW Cloning	GGGGACCACTTTGTACAAGAAAGCTGGGTATTATTGCCCTACCGGTGAAAGTT
SlbHLH060	*Solyc09g083220*	SlbHLH060_CDS_attB1	GW Cloning	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCAGCAAACCAACCGGAAG
		SlbHLH060_CDS_attB2	GW Cloning	GGGGACCACTTTGTACAAGAAAGCTGGGTACTATGATGGGGCTGCTGCTTCA
SlbHLH043	*Solyc06g051260*	SlbHLH043_CDS_attB1	GW Cloning	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGCTAAGATCCATACAATTAA
		SlbHLH043_CDS_attB2	GW Cloning	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAGGTAGCAAGCAAGAAGAAC
pDONR221	pDONR221	pDONR221-seq-F	Sequencing	GTTTTCCCAGTCACGAC
		pDONR221-seq-R	Sequencing	CAGGAAACAGCTATGAC
pDEST32	pDEST32	pDEST32_Seq_F	Sequencing	GGAGACTGATATGCCTCTAAC
		pDEST32_Seq_R	Sequencing	TCGTTTTAAAACCTAAGAGTCA
SlbHLH071	*Solyc12g036470*	SlbHLH071_q_F1	RT-q-PCR	CAATGCAGAGAGCAAAGGCA
		SlbHLH071_q_R1	RT-q-PCR	GATAGCAGCCAAAAGTGCGG
		SlbHLH071_CDS_attB1	GW Cloning	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCTCATCATTATTTCAATCCA
		SlbHLH071_CDS_attB2	GW Cloning	GGGGACCACTTTGTACAAGAAAGCTGGGTATTAGGTGTTGTAAGACAACTTGG
SlbHLH026	*Solyc03g119390*	SlbHLH026_q_F1	RT-q-PCR	GGCATGGCAGGAATGTTGGA
		SlbHLH026_q_R1	RT-q-PCR	GTAGTATGAGCTTGCTGCGG
		SlbHLH026_CDS_attB1	GW Cloning	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCTGATCATCAATTCAACT
		SlbHLH026_CDS_attB2	GW Cloning	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAGGTGTTGTATGGCAACATT
SlbHLH020	*Solyc03g034000*	SlbHLH020_q_F1	RT-q-PCR	TGATGGAGAAGCAGGGGAAG
		SlbHLH020_q_R1	RT-q-PCR	GCGACGAGCCCTAACATGAA
		SlbHLH020_CDS_attB1	GW Cloning	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGAGCATTCTTGAAAGACAAA
		SlbHLH020_CDS_attB2	GW Cloning	GGGGACCACTTTGTACAAGAAAGCTGGGTATTACTTCTGAAACCGTACACCA

Appendix 4 Information about the tomato population used in the mGWAS analysis in Chapter 6

Individual code	Name	Categories	*Botanical variety*	TGRC#	PI CGN#	EA #	original reads	depth	coverage
TS-2	Moneymaker	Vintage Fresh Market	*S. lycopersicum*	LA2706	-	EA00587	89149358	9.8915	86.570%
TS-3	M-82	Modern Processing	*S. lycopersicum*	LA3475	-	-	59276558	5.9389	88.320%
TS-4	Hawaii 7998	Inbreed line	*S. lycopersicum*	LA3856	-	-	66970352	5.9195	69.820%
TS-6	San Marzano	Vintage Processing	*S. lycopersicum*	LA3008	-	-	49443886	5.4085	88.130%
TS-10	Marglobe	Vintage Fresh Market	*S. lycopersicum*	LA0502	-	-	68472300	5.8254	88.120%
TS-12	yoku improvement	Modern Fresh Market	*S. lycopersicum*	-	-	-	68472300	5.8254	88.120%
TS-16		Wild species	*S. pimpinellifolium*	LA1246	-	EA03028	75768318	6.3421	87.660%
TS-17	Culebras	Wild species	*S. pimpinellifolium*	LA0373	-	-	71531324	4.1554	85.700%
TS-18		Wild species	*S. pimpinellifolium*	LA1579	-	EA01027	136199312	9.9611	85.580%
TS-20	Sechin	Wild species	*S. pimpinellifolium*	LA0442	-	EA00587	89149358	9.8915	86.570%
TS-21		Wild species	*S. pimpinellifolium*	LA1375	-	-	123613310	9.0552	84.720%
TS-22	Pisiquillo	Wild species	*S. pimpinellifolium*	LA1269	-	-	61507622	2.4268	73.760%
TS-24		Wild species	*S. pimpinellifolium*	-	-	-	49228164	3.7429	84.770%
TS-26		cultivar	*S. lycopersicum var cerasiforme*	-	-	EA01712	49193058	5.6535	87.170%

TS-28	N 347 Yablochnyi		*S. lycopersicum var cerasiforme*	-	-	-	137607590	12.8504	70.580%
TS-33		Wild species	*S. lycopersicum var cerasiforme*	LA2137	-	-	81951812	9.0878	88.430%
TS-35	tomate Richters	uncertain	*S. lycopersicum var cerasiforme*	-	-	-	81951812	9.0878	88.430%
TS-42		Processing tomato	*S. lycopersicum*	-	-	-	63085572	6.9631	86.070%
TS-50		Wild species	*S. pimpinellifolium*	LA0417	-	-	219165344	24.4295	88.160%
TS-51			*S. lycopersicum*	-	-	-	53804108	5.8031	85.840%
TS-56		Wild species	*S. lycopersicum var cerasiforme*	LA1320	-	-	51502824	5.307	87.620%
TS-60	New Yorker	Vintage Fresh Market	*S. lycopersicum*	LA2009	-	-	49443886	5.4085	88.130%
TS-62			*S. lycopersicum var cerasiforme*	-	-	-	52289016	5.9484	87.750%
TS-67		Cocktail tomato	*S. lycopersicum var cerasiforme*	-	-	LYC1969/ EA02724	325460518	30.9891	87.490%
TS-69	Huachinango	Latin American cultivar	*S. lycopersicum*	LA1459	PI93302	EA04243	356071914	34.9219	89.150%
TS-78		Processing tomato	*S. lycopersicum*	-	-	-	165798036	13.0569	68.680%
TS-82		Processing tomato	*S. lycopersicum*	-	-	-	165798036	13.0569	68.680%
TS-83		cultivar	*S. lycopersicum var cerasiforme*	-	-	-	165798036	13.0569	68.680%
TS-90		Cocktail tomato	*S. lycopersicum*	-	-	-	165798036	13.0569	68.680%
TS-99	Celsior		*S. lycopersicum var cerasiforme*	-	-	-	165798036	13.0569	68.680%

TS-105		Wild species	*S. lycopersicum var cerasiforme*	-	-	-	68472300	5.8254	88.120%
TS-106		cultivar	*S. lycopersicum var cerasiforme*	-	-	-	68472300	5.8254	88.120%
TS-107	Turrialba	Wild species	*S. lycopersicum var cerasiforme*	LA1542	-	-	68472300	5.8254	88.120%
TS-108	Puno I	Processing tomato	*S. lycopersicum*	-	-	-	68472300	5.8254	88.120%
TS-109		Cocktail tomato	*S. lycopersicum var cerasiforme*	-	-	-	68472300	5.8254	88.120%
TS-110			*S. lycopersicum*	-	-	-	68472300	5.8254	88.120%
TS-111		Processing tomato	*S. lycopersicum*	-	-	-	68472300	5.8254	88.120%
TS-113		Processing tomato	*S. lycopersicum*	-	-	-	68472300	5.8254	88.120%
TS-114		Processing tomato	*S. lycopersicum*	-	-	-	68472300	5.8254	88.120%
TS-116	N1565	cultivar	*S. lycopersicum var cerasiforme*	-	-	-	68472300	5.8254	88.120%
TS-117	Scatolone di bolsena	Landrace	*S. lycopersicum*	-	-	-	68472300	5.8254	88.120%
TS-120			*S. lycopersicum var cerasiforme*	-	-	-	68472300	5.8254	88.120%
TS-123	Trujillo	Wild species	*S. pimpinellifolium*	LA0722	-	-	68472300	5.8254	88.120%
TS-124	Santa Rosa	Wild species	*S. pimpinellifolium*	LA1245	-	-	68472300	5.8254	88.120%
TS-130	Severianin	Modern Fresh Market	*S. lycopersicum*	LA2413	-	-	68472300	5.8254	88.120%
TS-133	Peto95-43	Modern Processing	*S. lycopersicum*	LA3528	-	-	68472300	5.8254	88.120%
TS-134	La Estancilla	Wild species	*S. lycopersicum var cerasiforme*	LA1429	-	-	68472300	5.8254	88.120%

TS-135	Hacienda Rosario	Landrace/Latin American cultivar	*S. lycopersicum*	LA0466	-	-	68472300	5.8254	88.120%
TS-137	Spagnoletta	Landrace	*S. lycopersicum*	-	-	-	68472300	5.8254	88.120%
TS-138	Malintka 101	cultivar	*S. lycopersicum var cerasiforme*	-	-	-	68472300	5.8254	88.120%
TS-145	Tumbes south	Wild species	*S. pimpinellifolium*	LA1617	-	-	68472300	5.8254	88.120%
TS-148	Pfacchayoc	Wild species	*S. lycopersicum var cerasiforme*	LA1323	-	-	68472300	5.8254	88.120%
TS-152	Santa Cruz B	Landrace/Latin American cultivar	*S. lycopersicum*	LA1021	-	-	68472300	5.8254	88.120%
TS-153	Xol Languna	Landrace/Latin American cultivar	*S. lycopersicum*	LA1544	-	-	76024050	7.8749	61.920%
TS-154	Muna	Wild species	*S. lycopersicum var cerasiforme*	LA1623	-	-	67258956	7.566	87.920%
TS-164	Jayanca to La Vina	Wild species	*S. pimpinellifolium*	LA1584	PI407541	-	165616234	12.6465	63.030%
TS-165	Veracruz	Latin American cultivar	*S. lycopersicum var cerasiforme*	LA1218	-	-	66263978	6.6964	88.560%
TS-166	Piura	Landrace/Latin American cultivar	*S. lycopersicum*	LA0404	-	-	59410570	6.2424	65.290%
TS-167	Tegucigalpa	Landrace/Latin American cultivar	*S. lycopersicum*	LA0147	-	-	63663508	5.2112	78.520%
TS-168	Da appendere	Landrace	*S. lycopersicum*	-	-	-	69754534	5.3999	86.930%
TS-174		Processing tomato	*S. lycopersicum*	-	-	EA00304	81109984	6.7697	87.480%
TS-175		Processing tomato	*S. lycopersicum*	-	-	EA03586	71666832	5.5541	83.910%

TS-176		Processing tomato	*S. lycopersicum*	-	-	EA02669	88868716	7.0343	69.870%
TS-178		Processing tomato	*S. lycopersicum*	-	PI513036	EA06485	82227160	3.9378	83.830%
TS-179		Processing tomato	*S. lycopersicum*	-	-	EA01027	136199312	9.9611	85.580%
TS-180		Processing tomato	*S. lycopersicum*	-	-	EA02728	59464058	4.1308	84.730%
TS-181		Wild species	*S. lycopersicum var cerasiforme*	LA1457	-	EA01438	167172210	11.3255	84.080%
TS-185	B-L-35	Vintage Fresh Market	*S. lycopersicum*	LA4347	-	-	130475878	8.1241	53.990%
TS-187	Monplaisir	cultivar	*S. lycopersicum var cerasiforme*	-	-	-	68708862	5.0868	86.060%
TS-188	Stone	Vintage Fresh Market	*S. lycopersicum*	LA1506	-	-	191316854	14.2023	75.430%
TS-192	Severianin	Vintage cultivar	*S. lycopersicum*	-	-	-	73420040	6.2453	87.970%
TS-196	N020212	Processing tomato	*S. lycopersicum*	-	-	EA00240	64236996	7.0696	64.070%
TS-197	Libanese	Landrace	*S. lycopersicum*	-	-	-	76457448	6.0806	86.960%
TS-198			*S. lycopersicum*	-	-	EA00512	61229566	4.8539	86.020%
TS-201	San Salvador	Landrace/Latin American cultivar	*S. lycopersicum*	LA1210	-	-	58173846	5.6949	87.700%
TS-203	Bell pepper-like	Landrace	*S. lycopersicum*	-	-	-	57041632	5.6665	87.270%
TS-205	8 bis	cultivar	*S. lycopersicum var cerasiforme*	-	-	-	85874472	4.8315	85.050%
TS-209	Da serbo	landrace	*S. lycopersicum var cerasiforme*	-	-	-	123613310	9.0552	84.720%
TS-210	NC 265-1 (93)-3-3	Modern Fresh Market	*S. lycopersicum*	LA3625	-	-	93539654	8.8113	87.350%
TS-211	NC 84173	Modern Fresh Market	*S. lycopersicum*	LA4354	-	-	56525612	6.2606	85.360%

TS-216	Phyra	cultivar	*S. lycopDersicum var cerasiforme*	-	-	-	53330592	5.5976	84.070%
TS-217			*S. cheesmaniae*	LA0429	-	EA00568	62621512	2.4883	65.980%
TS-218	Santa Clara 5800	cultivar	*S. lycopersicum*	-	-	-	48120346	5.41	86.800%
TS-219	Mobalcon	landrace	*S. lycopersicum var cerasiforme*	-	-	-	61507622	2.4268	73.760%
TS-221	Jalapa	Wild species	*S. lycopersicum var cerasiforme*	LA1569	-	-	57794504	5.7897	83.340%
TS-222	Wva 700	Wild species	*S. pimpinellifolium*	-	-	-	71034348	7.6353	70.090%
TS-238	Macas	Wild species	*S. lycopersicum var cerasiforme*	LA1228	PI379047	-	82695138	8.4851	74.600%
TS-239	NC EBR-5	Modern Fresh Market	*S. lycopersicum*	LA3845	-	-	49228164	3.7429	84.770%
TS-249	Merida	Landrace/Latin American cultivar	*S. lycopersicum*	LA1462	-	-	59304384	6.4582	87.610%
TS-250			*S. lycopersicum var cerasiforme*	-	-	EA03539	77614104	7.4811	88.510%
TS-252		cocktail tomato	*S. lycopersicum var cerasiforme*	-	-	EA01802	65682872	6.5933	88.080%
TS-255		Processing tomato	*S. lycopersicum*	-	-	EA03002	76063882	8.5891	88.380%
TS-257	Marpha N2	cultivar	*S. lycopersicum var cerasiforme*	-	-	-	129681962	6.7014	35.900%
TS-283	Cisterno	cultivar	*S. lycopersicum var cerasiforme*	-	-	-	54401908	5.8518	88.170%
TS-286	Allungato piccolo	landrace	*S. lycopersicum var cerasiforme*	-	-	-	71485374	2.0054	65.540%

TS-290	N 2257 Dikorastushii...	cultivar	*S. lycopersicum var cerasiforme*	-	-	-	74705530	7.9286	61.850%
TS-294	N 933	cultivar	*S. lycopersicum var cerasiforme*	-	-	-	90201026	3.5056	80.810%
TS-296	Droplet		*S. lycopersicum*	-	-	-	69494486	7.5893	87.400%
TS-297			*S. lycopersicum*	-	PI291344	EA05550	63579442	5.2764	54.110%
TS-298	San Francisco	Wild species	*S. lycopersicum var cerasiforme*	LA2308	-	-	78187158	7.6211	73.420%
TS-299	Bomboiza	Wild species	*S. lycopersicum var cerasiforme*	LA2131	-	-	59276558	5.9389	88.320%
TS-300	Tena	Wild species	*S. lycopersicum var cerasiforme*	LA1231	PI379049	-	72938256	7.7547	79.250%
TS-301	Santa Cruz near Shintuyo	Wild species	*S. lycopersicum var cerasiforme*	LA2688	-	-	68477920	7.2726	54.020%
TS-302	Upper Parana	Wild species	*S. lycopersicum var cerasiforme*	LA1543	-	-	71594600	7.1511	88.240%
TS-403			*S. peruvianum*	-	PI128650	-	56298762	5.8533	71.980%
TS-418		Wild species	*S. pimpinellifolium*	LA2147	-	-	58411394	6.5738	86.060%
TS-419		Wild species	*S. pimpinellifolium*	LA2173	-	-	63085572	6.9631	86.070%
TS-420		Wild species	*S. pimpinellifolium*	LA2184	-	-	58757118	6.5447	85.750%
TS-425		Wild species	*S. pimpinellifolium*	-	PI126925	-	61076076	6.6946	84.390%
TS-430	BL 587-S	cultivar	*S. lycopersicum var cerasiforme*	-	PI127807	-	72086752	8.0026	87.510%
TS-432		Wild species	*S. pimpinellifolium*	-	PI270449	-	50055308	5.7034	86.690%
TS-433		Wild species	*S. pimpinellifolium*	-	PI370093	-	50598418	5.7367	86.600%

TS-503	Black		*S. lycopersicum*	-	-	-	51356624	5.8984	86.090%
TS-505	Brandywine		*S. lycopersicum*	-	-	-	61266902	6.9775	87.850%
TS-506	Chadwick Cherry		*S. lycopersicum var cerasiforme*	-	-	-	56598208	6.4112	87.170%
TS-511	Crimson Sprinter		*S. lycopersicum*	-	-	-	54721218	5.86	87.400%
TS-515	Garden Peach		*S. lycopersicum*	-	-	-	56516904	5.6036	87.320%
TS-518	Giant Belgium		*S. lycopersicum*	-	-	-	54096346	6.0742	87.450%
TS-519	Giant Oxheart		*S. lycopersicum*	-	-	-	70555854	8.0022	87.860%
TS-524	Japanese Black Trifele		*S. lycopersicum*	-	-	-	59721694	5.8334	87.800%
TS-527	Legend		*S. lycopersicum*	-	-	-	56630724	6.5484	87.860%
TS-536	Matt's Wild Cherry		*S. lycopersicum var cerasiforme*	-	-	-	57429286	6.1687	87.600%
TS-537	Mexico Midget		*S. lycopersicum var cerasiforme*	-	-	-	68282214	6.7663	87.830%
TS-538	Micado Violettor		*S. lycopersicum*	-	-	-	49270028	5.5201	86.850%
TS-539	Mr. Stripey		*S. lycopersicum*	-	-	-	53117262	5.9007	87.500%
TS-541	Oaxacan Pink		*S. lycopersicum*	-	-	-	63804008	7.1416	88.230%
TS-544	Oregon Spring Bush		*S. lycopersicum*	-	-	-	46750616	5.3114	86.690%
TS-546	Peron Sprayless		*S. lycopersicum*	-	-	-	52150826	5.9178	87.580%
TS-547	Ponderosa Pink		*S. lycopersicum*	-	-	-	48139638	5.4494	86.680%
TS-549	Porter		*S. lycopersicum*	-	-	-	52030550	6.0107	86.800%

TS-551	Red House Free Standing		*S. lycopersicum*	-	-	-	55413226	6.3731	88.100%
TS-552	Red Pear		*S. lycopersicum var cerasiforme*	-	-	-	69612494	7.7233	88.120%
TS-554	Santiam		*S. lycopersicum*	-	-	-	58030670	6.6828	87.990%
TS-555	Skorospelka Red		*S. lycopersicum*	-	-	-	60174276	6.89	88.120%
TS-558	Sunray		*S. lycopersicum*	-	-	-	50632740	5.6933	87.080%
TS-559	Super Sioux		*S. lycopersicum*	-	-	-	51502824	5.307	87.620%
TS-560	Thai Pink Cherry		*S. lycopersicum var cerasiforme*	-	-	-	64329154	7.4311	87.890%
TS-564	Tommy Toe		*S. lycopersicum var cerasiforme*	-	-	-	52519908	5.9722	87.280%
TS-565	Valencia		*S. lycopersicum*	-	-	-	49417674	5.4818	87.050%
TS-567	Wisconsin 55		*S. lycopersicum*	-	-	-	59348748	6.7038	87.780%
TS-568	Yellow Perfection		*S. lycopersicum*	-	-	-	57449418	6.6052	88.030%
TS-572	Andrew Rahart's Jumbo Red		*S. lycopersicum*	-	-	-	54903780	6.1128	86.970%
TS-574	Arkansas Traveler		*S. lycopersicum*	-	-	-	54151242	6.0796	85.350%
TS-575	Azoychka Russian		*S. lycopersicum*	-	-	-	54121108	6.2	86.660%
TS-578	Black Krim		*S. lycopersicum*	-	-	-	51539152	5.8894	86.640%

TS-579	Black Prince		*S. lycopersicum*	-	-	-	52785472	5.962	87.610%
TS-582	Brown Berry		*S. lycopersicum var cerasiforme*	-	-	-	66134582	7.4825	86.720%
TS-583	Burbank		*S. lycopersicum*	-	-	-	53294210	6.1258	87.750%
TS-586	Christmas Grapes		*S. lycopersicum*	-	-	-	49044082	5.5785	87.130%
TS-588	Dr. Wyche's Yellow		*S. lycopersicum*	-	-	-	64084342	6.8444	88.070%
TS-592	German Johnson Pink		*S. lycopersicum*	-	-	-	53718272	6.0239	87.520%
TS-600	Kellogg's Breakfast		*S. lycopersicum*	-	-	-	48859948	5.4855	87.320%
TS-604	Mule Team		*S. lycopersicum*	-	-	-	76015826	8.603	88.520%
TS-622	Sub-Arctic Plenty		*S. lycopersicum*	-	-	-	63396644	7.2721	86.910%

Appendix 5 Publication & Copyright License Agreement