Genetic variations in tomato chromoplast specific lycopene beta cyclase gene (*CYCB*) and its promoter

Thesis submitted for the award of the degree of

DOCTOR OF PHILOSOPHY

By

VIJEE MOHAN

Supervisor

Prof. R. P. Sharma

Department of plant sciences

School of life sciences

University of Hyderabad

Hyderabad-500046, India

2014

Enrol. No. 07LPPH08

Dept. of Plant Sciences

School of Life Sciences

University of Hyderabad

DECLARATION

I hereby declare that the work described in this thesis entitled **"Genetic variations in tomato chromoplast specific lycopene beta cyclase gene (*CYCB*) and its promoter"** which is being done by me under the supervision of Professor R. P. Sharma, Department of Plant Sciences, School of Life Sciences, University of Hyderabad, Hyderabad 500046, has not been submitted for any degree or diploma of any other University.

Date: **Vijee Mohan**

Place: Hyderabad **Enrol. No. 07LPPH08**

Prof. R. P. Sharma

Supervisor

Dept. of Plant Sciences

School of Life Sciences

University of Hyderabad

CERTIFICATE

This is to certify that the thesis entitled **"Genetic variations in tomato chromoplast specific lycopene beta cyclase gene (*CYCB*) and its promoter"** is based on the results of the work done by **Ms. Vijee Mohan** for the degree of **Doctor of Philosophy** under my supervision. This work presented in this thesis is original and plagiarism free, no part has been submitted for any degree or diploma of any other University.

Prof. R. P. Sharma

Supervisor

Head

Dept of Plant Sciences

Dean

School of Life Sciences

Dedicated to

My Parents & Family

ACKNOWLEDGEMENTS

*First and above all, I thank '**The Almighty**' for giving me this great opportunity and strengthened me physically and mentally to successfully complete this journey.*

*I would like to express my sincere thanks to my supervisor **Prof. R. P. Sharma** for accepting me as a Ph.D student, for his valuable guidance to mould a researcher in me, for giving me enough freedom to think and work in my own way, for his unsurpassed knowledge and patience to solve the problems, for his constant encouragement and understanding in all my difficult situations and above all for his fatherly affection extended to me throughout my research work and during preparation of the manuscript.*

My special thanks to Dr. Y. Sreelakshmi for her valuable suggestions and encouragement in each step of my research work which helped me in fine tuning the ideas and its execution, for active discussions leading to fruitful conclusions, for the positive energy and for the friendly attitude.

I would like to thank Prof. A. S. Raghavendra, Dean, School of Life Sciences and Prof Ch.V. Ramana, Head, Dept. of Plant Sciences, for providing the necessary facilities and encouragement in pursuing my research work. I also thank Prof R. P. Sharma, Prof. Aprna Dutta Gupta, Prof M. Ramnadham, former Deans, School of Life sciences and Prof A. R Reddy, Prof A. R. Podille, former Heads, Department of Plant Sciences for all the facilities provided during my work.

My sincere thanks to Prof A. R. Reddy, my doctoral committee member for his guidance and timely suggestion throught the research.

I thank all the faculty members of the Department and School for extending their lab facilities and giving me valuable support needed for my research.

I thank Dr. Roger P. Hellens, The New Zealand Institute for Plant and Food Research, for kindly providing the construct for transient expression studies.

My heartful thanks to all the lab members of RTGR for their support and for making my stay at University a memorable one. My special thanks to Dr. Soni Gupta for her suggestions in standardizing conditions for LI-COR based SNP detection and Dr. Arun Kumar Pandey for his help in transient expression studies. I also acknowledge my seniors Dr. Vineeta, Dr. Santisree, Dr. Sulabha, Mr. Reddaiah, Dr. Osman, Dr. Vajir, Dr. Vineet and Dr. Bharti for their suggestions and encouragement. I would like to express my love to my batch mates Sherin and Mickey and my juniors Sapna, Chaitanya and Rachna who created a cheerful and enthusiastic atmosphere in our

old lab and helped me in many ways to face problems in research as well as in personal life. I also thank my juniors Suresh, Alka, Rakesh, Kapil, Supriya, Kamal, Prateek, Pallawi, Hima Bindu and Hyma for their help and positive attitude.

I am thankful to Zamil and Narasimha for their help in maintaining plants in green house and field. I also thank lab accountants Venkat and Anil for their help.

I would like to thank my husband Mr. Anil M.G. for his support, encouragement and suggestions during my research work. I also thank the two precious stones arrieved in my life, my daughter Revati and son Sachit for filling me up with positive emotions.

I have no words to put across my gratefulness to my parents Mr. P. V. Mohanan and Mrs. Rajamma Mohanan for their never ending encouragement and support especially during difficult situations. I am thankful to my dear sister Rajee for her encouraging words and affection. I would also like to thank my in-laws Mr. P. Madhusoodana Kurup and Mrs. Girija Devi for their support.

I would like to thank DBT, DST and IAEA for their funding to the lab. The financial support from University of Hyderabad, DBT-CREBB and CSIR during the period of my research work is greately acknowledged.

Vijee Mohan

2014

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

%.....	percent
°C……...	degree Celsius
µ…….	micro
A……	Absorbance
ABA……..	Abscisic acid
AV…….	Arka Vikas
B……	*Beta*
bp……...	base pair
CCD……...	carotenoid cleavage dioxygenase
CCS……...	capsanthin capsorubin synthase
cDNA……	Complementary DNA
cm……..	centimeter
Cnr……	*Colourless non-ripening*
CODDLE……..	Codons Optimized to Discover Deleterious Lesions
CRTISO……	carotenoid isomerase
CV…….	Coefficient of Variation
cv……...	cultivar
CYCB……	Lycopene cyclase, chromoplastic
DDB1……	DNA Damage-Binding Protein1
Del.……	*Delta*
DET1…….	De-etiolated1
DMAPP…….	dimethylallyl diphosphate
DNA……..	deoxyribonucleic acid
DNase……	deoxyribonuclease
dNTPs……	deoxynucleotide triphosphates
EcoTILLING….	Ecological TILLING
F…….	Forward
FLUC…….	Firefly luciferase
FW…….	Fresh Weight
g…….	gram
GGPP…….	geranylgeranyl pyrophosphate
gh……...	*ghost*
hp……..	*high-pigment*
hr……...	hour
IIVR……..	Indian Institute of Vegetable Research
IPP…….	isopentenyl pyrophosphate
IRDye……	infrared fluorescent dye
Kb……..	Kilobase
LCYB……	Lycopene cyclase, chloroplastic
LCYE……	lycopene epsilon cyclase
M……...	Molar
mA…….	milliampere
Mb…….	Megabase
MEP……...	2-C-methyl-D-erythritol 4-phosphate
mg…….	milligram
min……	minute

ml……..	millilitre
mM……	millimolar
NBPGR…….	National Bureau of Plant Genetic Resources
NCBI…….	National Center for Biotechnology Information
NCED……	9-cis-epoxycarotenoid dioxygenases
nm…….	nanometer
nor…….	*non-ripening*
not…….	*notabilis*
Nup……	Nuclear localized plastid
NXS……..	neoxanthin synthase
og.....	*old gold*
og^c…….	*old-gold-crimson*
ORF……..	Open Reading Frame
PARSESNP…..	Project Aligned Related Sequences and Evaluate SNPs
PCR……..	polymerase chain reaction
PDS……..	phytoene desaturase
PLACE……..	PLAnt Cis-acting regulatory DNA Elements
pmoles……..	picomoles
PSSM…….	Position Specific Scoring Matrix
PSY……..	phytoene synthase
PTOX…….	plastid terminal oxidase
r…….	*yellow-flesh*
R……	Reverse
rin…….	*ripening-inhibitor*
RLUC……	*Renilla* luciferase
RNA……..	ribonucleic acid
RNase……	ribonuclease
rpm……	rotations per minute
SD……..	standard deviation
sec……..	second
SGN……..	Solanaceae Genome Network
SIFT……..	Sorting Intolerant From Tolerant
SNP……	Single Nucleotide Polymorphism
t……..	*tangerine*
TGRC……	Tomato Genetic Resources Center
TILLING ……..	Targeting Induced Local Lesions in Genomes
tRNA……..	transfer RNA
UV…….	ultraviolet
V/cm …….	volt per centimeter
v/v…….	volume/volume
VAD……..	Vitamin A Deficiency
VDE……..	violaxanthin de-epoxidase
w/v…….	weight/volume
wf……...	*white flower*
WT…….	wild type
ZDS……	ζ-carotene desaturase
ZEP……	zeaxanthin epoxidase
ZISO……..	ζ-carotene isomerase
µl……	microliter

Chapter 1

Introduction

Carotenoids comprise a large group of pigments that are ubiquitous in nature. In plants carotenoids play indispensable roles in light harvesting and as precursors of plant hormone abscisic acid and strigolactone. They are produced via the general isoprenoid biosynthetic pathway in chloroplasts of photosynthetic tissues and in the chromoplasts of fruits and flowers. Most of the enzymatic steps in the carotenoid biosynthetic pathway have been characterized and remaining are investigated. All of these enzymes are nuclear-encoded and genes encoding for many of them have been cloned in recent years. The information on gene expression and their regulation on different intermediates of the carotenoid pathway give opportunities for directed manipulation of the types and quantities of these carotenoids.

Tomato is considered as the model system for Solanaceous plants because of the comparatively small genome size and availability of genomic data resources. Tomato has economic value as one of the most consumed vegetables and is a principal dietary source of carotenoids and flavonoids. Both carotenoids and flavonoids are highly beneficial for human health. The compounds such as lycopene reduce the risk of prostate cancer and *β*-carotene is the precursor for vitamin A, an essential nutrient (Bendich, 1993).Tomato is also a favorite model system for studying the regulation of carotenoid biosynthesis because of the dramatic color changes that occur during fruit development, leading to lycopene and *β*-carotene accumulation.

The carotenoid biosynthesis pathway in ripening tomato fruit is essentially similar to that operate in photosynthetic tissue such as leaf, with few fruit specific modification during chloroplast to chromoplast transition. In most of the plant species a chloroplast specific lycopene *β*-cyclase enzyme (LCYB) mediates the conversion of lycopene to *β*-

carotene. In tomato, two different *LCYB* genes are reported (*LCYB1* and *LCYB2*). The expression of these two genes decreases in fruit ripening during the transition of chloroplast to chromoplast. This function is replaced by a paralog of *LCYB* which codes for a different lycopene beta cyclase, but its expression is specific to the chromoplast, hence, called 'chromoplast specific *lycopene betacyclase'* gene (*CYCB*). All the three lycopene cyclases are encoded by nuclear genes. *CYCB* gene has an ORF of 499 codons and is not interrupted by introns. The amino acid sequence of the polypeptide encoded by *CYCB* has only ca.55% similarity to LCYB1 and LCYB2 (Ronen *et al*., 2000).

In most tomato varieties the lycopene content in fruits is at least ten times higher than *β*-carotene content. The mutation in *CYCB* gene can lead to enhanced lycopene in a loss of function mutant and high *β*-carotene in case of enhanced expression in tomato fruits. It is reported that in tomato fruits, *CYCB* is expressed at low levels during the breaker stage of ripening, whereas in '*Beta'* mutant its transcription is dramatically increased. The *B* allele (*CYCB* allele present in *Beta* mutant), originated in wild tomato species and was introduced into the cultivated tomato by crossings. Null mutations in the *CYCB* gene are responsible for the phenotype of deep red fruits and tawny orange flowers in the mutants *old-gold(og)* and *old-gold-crimson(*og^c*)*(Ronen *et al*., 2000).

Functional characterization of *CYCB* gene can also be carried out by screening tomato germplasm collections to identify natural DNA sequence variations in the gene and characterizing the effect of those variations on biosynthesis of *β*-carotene. Tomato has few mutants that have altered nutrition profile, but most of these are spontaneous mutations selected by breeders. Since the forward genetic methods rely on the phenotype screening, they preclude selection of mutants targeted in a specific biosynthetic pathway.

On the other hand, reverse genetics methods allow one to directly target mutations in the gene of interest. TILLING (Targeting Induced Local Lesions in Genomes) is a nontransgenic reverse genetic strategy for rapid and low-cost discovery of induced point mutations in a target gene. Single nucleotide polymorphisms (SNPs) can be detected in a collection of natural accessions very effectively by applying the technique of TILLING, a method known as EcoTILLING (Comai *et al*., 2004).

In the present study, the genetic variations in the form of SNPs, Insertions and Deletions in *CYCB* gene and its promoter in different tomato accessions are analyzed along with their probable effect on respective functions. *In-silico* analysis of promoter region of this gene was carried out to identify the probable response elements. The relative efficiencies of three promoter sequences which are representative of the observed variable promoters were compared by transient expression studies.

Major objectives of the study were:

I. Diversity studies in *CYCB* gene between different plant genera and between different accessions and wild relatives of tomato.

II. Characterization of *CYCB* gene promoter in different wild relatives and accessions of tomato and effect of promoter variations on *CYCB* gene expression.

Chapter 2

Review of Literature

2.1. Tomato as a model system

Tomato (*Solanum lycopersicum* L., formerly *Lycopersicon esculentum* Miller) is a major vegetable crop cultivated and consumed in almost every country. The cultivated tomato belongs to the order Scrophulariales, suborder Solanineae, family Solanaceae, tribe Solaneae, genus Lycopersicon, subgenus Eulycopersicon, species *Lycopersicon esculentum* (lycopersicon = wolf peach, esculentum = edible) (Inbaraj and Chen, 2008).

Tomatoes are consumed as fresh salad as well as in processed forms like ketch up, puree and paste (Nielsen, 1994) which makes it a versatile commodity. According to FAO STAT 2012, India ranks second in world tomato production after China (http://faostat.fao.org/). Other major producers are the United States and Turkey. In addition to its economic importance tomato is an excellent model system for functional genomics studies in plants. The short generation time, routine transformation technology and availability of rich genomic data resources (Solanaceae Genome Network (SGN); http://www.sgn.cornell.edu) made it the most intensively investigated Solanaceous species. It has a diploid genome with 12 chromosome pairs and a genome size of 900 Mb predicted to contain 34,727 protein-coding genes that are largely sequestered in contiguous euchromatic regions (The Tomato Genome Consortium, 2012).

2.2. Wild relatives of tomato

Peru has been hypothesized as the center of origin and domestication of tomato (de Candolle, 1882). However, the name 'tomato' was originated after its domestication and cultivation by Indian tribes that inhabited Mexico (Giordano and Silva, 2000). Spooner *et al*. (1993) changed the genus name '*Lycopersicon'* to '*Solanum* sect. *Lycopersicon'* consisting of 14 closely related species or subspecies including the

domesticated tomato. All members of the above clade are to some degree intercrossable (Rick, 1979) and share a high degree of genomic synteny (Chetelat and Ji, 2007).

The habitat of tomato and its relatives extend from Ecuador to Chile. Eastern and western Andean slopes, and coastal region of northwest South America were identified as richest areas of their occurrence. The striking environmental gradients in this region produced by the dynamic geological and geographical structure generated a broad natural ecological diversity of this group (Young *et al.,* 2002). In addition to that there are certain species endemic to places like Mexico and Galapagos islands. The high variability in this group makes *Lycopersicon* exceptionally suited to ecological and evolutionary analysis at both the population and species level and research on biodiversity, domestication and nutrition perspectives (Peralta and Spooner, 2007).

Most of the wild relative species are smaller in size and green in color with few exceptions. *S. cheesmanii* and two accession of *S. lycopersicum* var. *cerasiforme* from the Galapagos islands are with yellow and orange fruits. The only wild relative species with red fruits is *S. pimpinellifolium*. Compared to other wild relatives, their fruits are very small in size (weighing only few grams). Due to reduced apical dominance and prostrate growth habit it has a large shrub nature and inflorescence carrying many flowers and fruits (Paran and van der Knaap, 2007). The divergence and domestication of tomato is believed to have happened from *S. pimpinellifolium* (de Candolle, 1882), but due to drastic reduction of its natural habitats, *S. pimpinellifolium* has undergone bottleneck and is now an endangered species (Biodiversity-International, 2006).

Crop improvement in tomato targets many aspects like increase in yield and nutritional factors resistance to biotic and abiotic conditions etc. Tomato Crop

Germplasm Committee (Tomato CGC) had conducted a survey on this aspect and suggested an approach to genetically resolve many of these problems using the genes from germplasm banks (Tomato CGC 2003). Identification of several genes of interest from the accessions conserved in germplasm banks, to meet the various requirements in crop improvement have been going on including those from wild relatives (Table 1).

Table 1: Characteristic features of conserved wild type tomato species in germplasm banks around the world (da Silva *et al*., 2008).

Source	**Character of interest**
S. peruvianum	Resistance to several pests; rich source of vitamin C
S. pennellii	Resistance to drought; increased contents of vitamins A and C and sugars
S. pimpinellifolium	Resistance to diseases, lower acidity, intense color, greater content of vitamins and soluble solids
S. esculentum var. *cerasiforme*	Resistance to high temperatures and humidity and to fungi that attack leaves and roots
S. chmielewskii	Intensity of fruit color and sugar content
S. chilense	Resistance to drought
S. cheesmanii	Tolerance to sea water and pedicles without articulations
S. habrochaites	Resistance to insects, spider mites, viruses and other diseases; tolerance to cold
S. parviflorum	Intensity of fruit color and high contents of soluble solids

Attempts to minimize the biotic stress problem have been one of the most focused areas in crop improvement. Selection for disease resistance started in 20th century in the United States and wild species were first used as source of adaptation to biotic stress. In 1912, the first *Fusarium* wilt resistant cultivar, "Tennessee red" was released. Later, identification of dominant resistance genes against many pests and diseases lead to several successful selections for disease resistance. Till now, 12 different disease resistance genes have been derived from wild species (da Silva *et al*., 2008).

Several useful alleles created by natural genetic variations have been identified and were used to make better cultivars. A natural mutation conferring a determinate growth to the plant, named *sp* (for self-pruning) was discovered in early 20th century. This mutation giving a compact fruit set was introduced in a variety "Roma". Another variety developed in 1960s by transferring the same allele is VF145 which was the first mechanically harvested variety and has been the major cultivar for tomato ketchup industry for more than a decade in California (Szymkowiak and Irish, 2005; Quinet *et al*., 2011). Another allele *j2* (*jointless*) which was introgressed from *S. cheesmanii* has been transferred to many processing varieties in combination with *sp*, allowing a large scale mechanical harvest of tomato fruits (Bauchet and Causse, 2012). Specific mutations like *hp* (Lieberman *et al*., 2004) and suitable alleles from wild relatives can be used to improve the potential health beneficial components like lycopene, *β*-carotene etc.

The clade *Solanum* section *Lycopersicon* is ideal for integrating genomic tools and approaches into ecological and evolutionary research (Moyle, 2008). Trait introgression from wild tomato relatives to cultivated tomatoes was pioneered by Prof. Charles Rick. These introgressions could increase the genetic diversity of cultivated tomatoes. Till

now, introgression lines generated in the cultivated tomato background for seven wild tomato species are available. Introgression lines can be more efficiently utilized along with the complete genome sequence information and knowledge about SNPs in the wild relatives compared to cultivated tomatoes and may lead to an intensive bio-diversity based breeding (The Tomato Genome Consortium, 2012). One disadvantage with introgression lines is inherent linkage between "favorable" and "unfavorable" effects of introgressed fragments. Nevertheless, identification of genetic control of complex traits in detail using the advanced molecular techniques will allow pinpointing the key alleles in wild relatives to be transferred to the cultivated tomatoes (Bauchet and Causse, 2012).

Today, the collection of tomato accessions in the seed banks worldwide ranks Ist among vegetable species collected and consists of more than 83,000 lines (FAO 2010). The major seed banks in the world are: the Tomato Genetic Resources Center (TGRC) in California, USA (www.tgrc.ucdavis.edu), the United States Department of Agriculture (USDA), Geneva (www.ars.usda.gov), the World Vegetable Center in Taiwan (www.avrdc.org) and several Europeans collections. More recently, a collection of more than 6,000 domesticated tomato accessions was established and phenotyped in the context of a European Solanaceae project (EU-SOL, www.eu-sol.wur.nl), accompanied by an *ad-hoc* database (Finkers *et al*., 2011).

2.3. Fruit growth and ripening in tomato

The process that occurs during growth and ripening of tomato fruit can be divided into three stages. In the initial stage, during the first 1 to 2 weeks after fertilization, there is a spurt of cell divisions reaching the maximum number of cells by the end of this stage. During the next 3 to 5 weeks which forms the second stage of fruit growth, rapid cell

expansion occurs and the fruit reaches its maximum size. The last stage involves ripening which involves a range of physiological, physical and chemical changes that are triggered by the plant hormone ethylene. Ripening process changes the color, texture, sugar content and flavor of the fruit making it attractive for consumption. Being a climacteric fruit, ripening occurs during development on the plant and continues even after harvest (Srivastava and Handa, 2005). The color change from green to red that occur during ripening of tomato is the result of transformation of chloroplasts in the cell to chromoplast which happens along with the accumulation of lycopene in the chromoplast. The increase in lycopene content during ripening is very dramatic with about 500 fold increase compared to the unripe fruit and becomes the predominant carotenoid in the fruit (80-90%) followed by *β*-carotene (5-10%) (Fraser *et al*., 1994; Kaur *et al*., 2006; Lenucci *et al*., 2006).

2.4. Carotenoids in tomato

Carotenoids are pigments present in plants and animals. In plants they are localized in plastids where they help in harvesting light in the blue-green wave length and pass the energy to photosynthesis reaction centers. This property of carotenoids is due to the presence of extensively conjugated double bond structures. The color of tomato is due to chlorophylls (green color), carotenoids mainly lycopene (red color) and *β*-carotene (orange color), and xanthophylls (oxygenated carotenoids with yellow color). During maturation the green color of tomato decreases and gets replaced by yellowish orange color mainly due to *β*-carotene. But during ripening this color gets replaced by red color due to lycopene accumulation. Tomatoes have been widely considered as the most important source of lycopene and second most important source of *β*-carotene and other

carotenoids such as *γ*-carotene, *ζ*-carotene, *α*-carotene and *δ*-carotene (Garcia-Closas *et al*., 2004). Tomato also contains colorless carotenoid precursors such as phytoene and phytofluene (15-30%), xanthophylls (free and esterified, 6%) and other carotenes (Gross, 1991).

2.4.1. Chemical composition and biosynthesis of carotenoids

Carotenoids are C40 tetraterpenoids consisting of C5 isoprenoid units derived from the acyclic C40H56 polyene lycopene. Generally the carotenoids with hydrocarbons are called carotenes and their oxygenated forms are termed xanthophylls (Britton, 1995). Most of the bacteria, fungi and plants possess the genes for carotenoid biosynthesis and produces vast number of carotenoids, while animals can absorb these carotenoids by consuming these organisms and accumulate in their body. Till now, only one animal from kingdom animalia i.e. pea aphid (*Acyrthosiphon pisum*) is identified to have carotenoid biosynthesis genes like several carotenoid synthases, cyclases, and desaturases. A mutation in one allele of these genes prevents the production of the red carotenoid, torulene, resulting in the green aphid color morph instead of normal red color morph. The green aphid morph has large quantities of greenish-yellow carotenoids: *alpha*-, *beta*-, and *gamma*-carotene. The red aphid has decreased amounts of those carotenoids, and instead has gobs of the red carotenoid, torulene. Indeed, no other known animal genome, including several other insect genomes, contains homologues to these genes. The high similarity of carotenoid biosynthesis genes of the aphid to that of several fungus species suggests the possibility of a horizontal gene transfer at some point in aphid evolution, from a fungus directly into the aphid genome. Possibility of occurrence of similar event from bacteria to plants as well is reported (Moran and Jarvik, 2010).

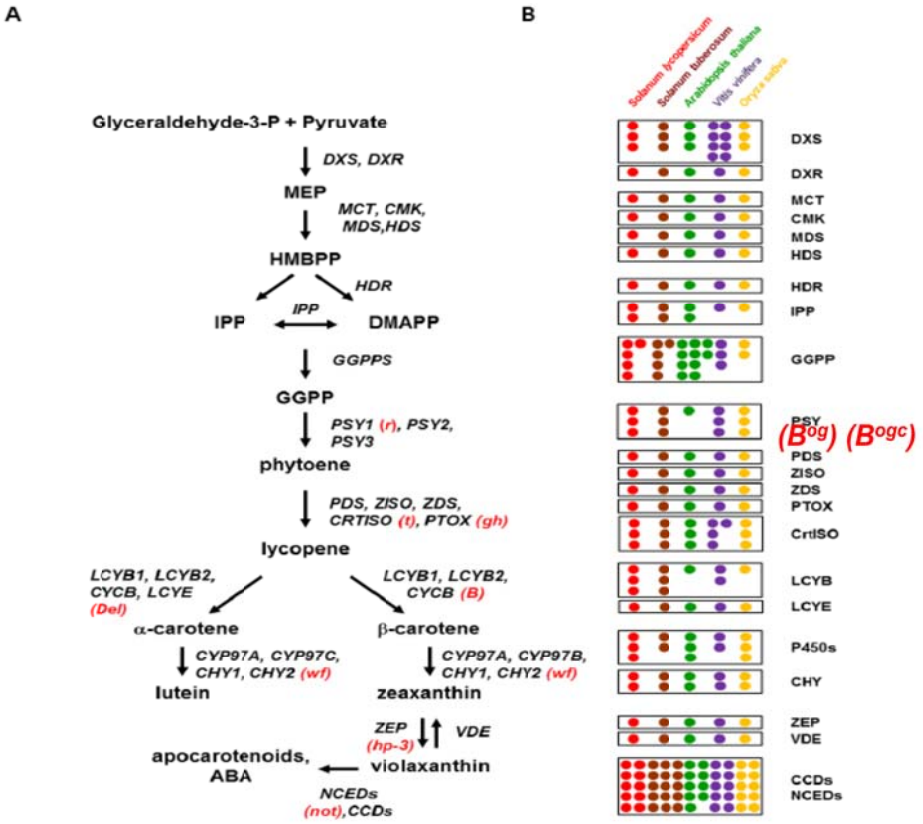
Carotenoids are biosynthesized in plant plastids via MEP (2-C-methyl-D-erythritol 4-phosphate) pathway from the precursor isopentenyl pyrophosphate (IPP), which is formed from mevalonic acid after successive phosphorylation by kinase enzymes and adenosine triphosphate followed by decarboxylation (Kopsell and Kopsell, 2006). IPP on isomerization yields dimethylallyl diphosphate (DMAPP), which in turn condenses with successive addition of IPP forming a C20 geranyl geranyl pyrophosphate (GGPP) mediated by the enzyme GGPP synthase. Phytoene is the first C40 carotenoid and is derived by the condensation of two molecules of GGPP by the phytoene synthase (PSY) enzyme. The enzymes phytoene desaturase (PDS), ζ-carotene isomerase (ZISO), ζ-carotene desaturase (ZDS) and carotenoid isomerase (CRTISO) catalyses sequential introduction of double bonds at alternate sides of phytoene leading to the synthesis of the carotenoids ζ-carotene, lycopene etc. Lycopene consists of 13 double bonds and is a symmetrical molecule (Bartley *et al*., 1999; Park *et al*., 2002).

Carotenoid biosynthesis branches at the cyclization of lycopene by the enzymes LCYE (lycopene epsilon cyclase) and LCYB (lycopene beta cyclase) leading to the synthesis of *α*-carotene with one *β*-ring and one *ε*-ring and *β*-carotene with two *β*-rings. By the addition of oxygen moieties, *α*-carotene and *β*-carotene are converted to many xanthophyll molecules like lutein (from *α*-carotene) and zeaxanthin, violaxanthin and neoxanthin (from *β*-carotene) (Inbaraj and Chen, 2008). Zeaxanthin epoxidase (ZEP) catalyses the conversion of zeaxanthin to violaxanthin by an epoxidation reaction (Marin *et al*., 1996), whereas, violaxanthin de-epoxidase (VDE) catalyses the reverse reaction (Pfundel *et al*., 1994). Conversion of violaxanthin to neoxanthin is catalysed by the enzyme neoxanthin synthase (NXS) (Al-Babili *et al*., 2000). Astaxanthin, capsanthin,

capsorubin and canthaxanthin are other xanthophyll molecules synthesized in various plants (Giuliano *et al*., 2008; Jackson *et al*., 2008). Several enzymes participate in the degradation of carotenoid molecules. CCDs (carotenoid cleavage dioxygenases) are involved in the biosynthesis of strigolactone, whereas, NCEDs (9-cis-epoxycarotenoid dioxygenases) are involved in the ABA biosynthesis (Walter *et al*., 2010; Bouvier *et al*., 2005). Schematic representation of carotenoid biosynthesis pathway in plants is given in Figure 1.

2.4.2. Carotenoid mutants in tomato

Several mutants have been identified with altered levels of carotenoids which include those with mutation in the genes directly involved in carotenoid biosynthesis pathway and those which are indirectly involved in regulation of carotenoid accumulation. A recessive mutation in the gene *PSY1* which codes for phytoene synthase required in chromoplast tissue, eliminates fruit carotenoids and leads to a fruit phenotype with yellow color. Thus the mutant was named as *yellow-flesh* (*r*) mutant (Kachanovsky *et al*., 2012). Another recessive mutation is identified in the gene *CRTISO* which codes for carotenoid isomerase and is expressed in all types of tissues. The defect in *CRTISO* lead to an orange colored mutant fruit called *tangerine* (*t*) due to the accumulation of tetra*cis*–lycopene (orange) instead of all*trans*–lycopene (red). Out of the two mutant alleles identified, *tangerine*^{mic} is a loss of function allele due to deletion in *CRTISO*, whereas, *tangerine*^{3183} has impaired expression of *CRTISO* (Isaacson *et al*., 2002). An epistatic interaction of *tangerine* to *yellow-flesh* was also described (Kachanovsky *et al*., 2012). *PTOX* encodes plastid terminal oxidase which has a role of cofactor for phytoene desaturase. A mutant deficient in this enzyme has reduced ability to synthesize



Supplementary Figure 54. Genes for the MEP/carotenoid pathway.
A. Simplified pathway. Mutant names are shown in red (see **Supplementary Table 14**).
B. Orthologous proteins involved in carotenoid biosynthesis from in *Solanum lycopersicum* (red), *Solanum tuberosum* (brown), *Arabidopsis thaliana* (green), *Vitis vinifera* (purple) and *Oryza sativa* (yellow) identified using the orthoMCL clusters (http://solgenomics.net/tools/genefamily/search.pl) and further manual curation. Each circle represents one gene.

Figure 1: **Schematic representation of a simplified carotenoid biosynthesis pathway in plants (The Tomato Genome Consortium, 2012)**.
GGPP-geranylgeranyl diphosphate, PSY-phytoene synthase, PDS-phytoene desaturase, ZISO-zeta-carotene isomerase, ZDS- zeta-carotene desaturase, CRTISO-carotenoid isomerase, PTOX-plastid terminal oxidase , LCYB-lycopene beta cyclase (chloroplastic), CYCB-lycopene beta cyclase (chromoplastic), LCYE-lycopene epsilon cyclase, CYP-cytochrome P450, CHY-carotene hydroxylase, ZEP-zeaxanthin epoxidase, VDE-violaxanthin de-epoxidase, NCED-9-*cis*-epoxy-carotenoid dioxygenase, CCD-carotenoid cleavage dioxygenase, ABA- abscisic acid. Names of tomato mutants for each gene is given in red letters and brackets. *r:yellow flesh, t:tangerin, gh:ghost , B:Beta, B^{og}:Beta old-gold, B^{ogc}:Beta old-gold-crimson, Del:Delta, wf:white flower, hp-3:high pigment-3, not: notabilis.*

carotenoids leading to variegated leaves due to photobleaching which is termed as *ghost (gh)* phenotype (Shahbazi *et al*., 2007). Three different mutants have been identified for the gene *CYCB* encoding lycopene beta cyclase for chromoplastic tissues. *Beta old-gold* (B^{og}) and *Beta old-gold-crimson* (B^{ogc}) are two loss of function mutants with deep red fruit phenotypes and tawny orange flowers, whereas, *Beta* (*B*) is a partially dominant mutant due to enhanced expression of *CYCB* characterized by orange colored fruits with very high amount of *β*-carotene (45-50% of total carotenoids compared to 5-10% of that in normal tomato fruits) (Ronen *et al*., 2000).

LCYE encodes lycopene epsilon cyclase which converts lycopene to *δ*-carotene. A dominant allele *Del* in *Delta* mutant has orange colored fruit due to the accumulation of *δ*-carotene at the expense of lycopene (Ronen *et al*., 1999). A recessive mutation in the *CrtR-b2* gene which encodes *β*-ring carotene hydroxylase in flowers resulted in *white flower* (*wf*) phenotype due to reduction in the carotenoid concentration (Galpaz *et al*., 2006). An abscisic acid (ABA) deficient mutant *notabilis* (*not*) is impaired in the oxidative cleavage of a 9-cis xanthophyll precursor to form xanthoxin and has a characteristic wilty phenotype (Burbidge *et al*., 1999). Mutations in genes coding for DNA Damage-Binding Protein1 (DDB1) (Lieberman *et al*., 2004; Liu *et al*., 2004), Arabidopsis (*Arabidopsis thaliana*) homolog of De-etiolated1 (DET1) (Mustilli *et al*., 1999) and zeaxanthin epoxidase (Galpaz *et al*., 2008) resulted in the high-pigment mutants *hp1*, *hp2* and *hp3* respectively. These mutants have high lycopene content due to increased number of chromoplasts (Azari *et al*., 2010).

2.4.3. Availability of carotenoids in different food products and their nutritional role

Carotenoids are present in various colored fruits and vegetables. α-carotene and β-carotene are present in apricots, cantaloupe, carrots, pumpkin, sweet potato etc, whereas, water melon, pink grape fruit and tomatoes are rich in lycopene, ζ-carotene, β-carotene, phytofluene and phytoene. Generally fruits with yellow or orange color like mango, papaya, peaches, prunes, squash, oranges etc are good sources of lutein, zeaxanthin, β-cryptoxanthin, α-, β- and ζ-carotene, phytofluene and phytoene, whereas, green fruits and vegetables are sources of lutein, zeaxanthin, α- and β-carotene (Lessin *et al*., 1997).

Even though over 600 carotenoids are available in nature, human consumption is limited to only about 40 out of which 14 carotenoids are found in human serum (Gerster, 1997). Tomatoes are rich in various carotenoids and can contribute to 9 out of those 14 carotenoids (Khachik *et al*., 1995). The red colored pigment lycopene and orange colored β-carotene contributes the major portion of carotenoids in tomato. Lycopene has a characteristic occurrence of very high concentrations in a restricted number of vegetables. The concentration of lycopene in tomato varies with time of harvest, geographic location and plant genotype. Some of the wild cultivars of tomato possess the highest concentration of lycopene as high as double the concentration of that of commercial cultivars (Dorais *et al*., 2001).

Carotenoids are more bioavailable in their processed form than from raw tomatoes (Gärtner *et al*., 1997). Bioavailability of β-carotene from tomato puree is more compared to lycopene availability even though lycopene content is more in tomato. This might be due to the higher solubility of β-carotene in the micelles compared to lycopene (Reboul *et al*., 2005). The bio-availability also depends on the isomeric form. For example, *cis*-

isomer of lycopene is more bioavailable than its *trans*-isomer (Boileau *et al*., 2002). Human blood plasma contains very high concentration of lycopene indicating its essentiality as antioxidant and anti-mutagenic agent. In the human organism the lycopene is present in high concentrations in the blood plasma, indicating its essentiality in natural defense pathway and as an antioxidant.

Several studies have suggested that dietary lycopene may reduce the risk of cardiovascular diseases (Sesso *et al*., 2004, 2005). It is also reported that it may reduce many other health problems like hypertension, hypercholesterolemia and may protect against prostate cancer, breast cancer etc (Uppala *et al*., 2012; Teodoro *et al*., 2012). Neutraceutical containing lactolycopene which is a highly bio-available form of lycopene have been released commercially and the manufacturers argue that one tablet is sufficient to provide bioavailable lycopene than a liter of tomato juice or 100g of tomato paste (http://www.ateronon.com/).

Vitamin A deficiency (VAD) is one of the major health problems in the world mostly affecting the developing world. Its deficiency among the pre-school children and pregnant women are a major concern worldwide affecting millions of people (West, 2003) (Figure 2). Retinoic acid is the active metabolic product of vitamin A. The *β*-ionone ring present in the carotenoids is essential for the pro vitamin A activity of these compounds. Since lycopene lacks a *β*-ionone ring, it does not have this activity while carotenoids like *β*-carotene, *α*-carotene and *γ*-carotene which have one or two *β*-ionone rings, exhibit pro vitamin A activity (Gross, 1991). Vitamin A plays a vital role in organogenesis during development and homeostasis (Clagett-Dame and DeLuca, 2002). Its role in lung development is more critical as retinoic acid deficiency is correlated with

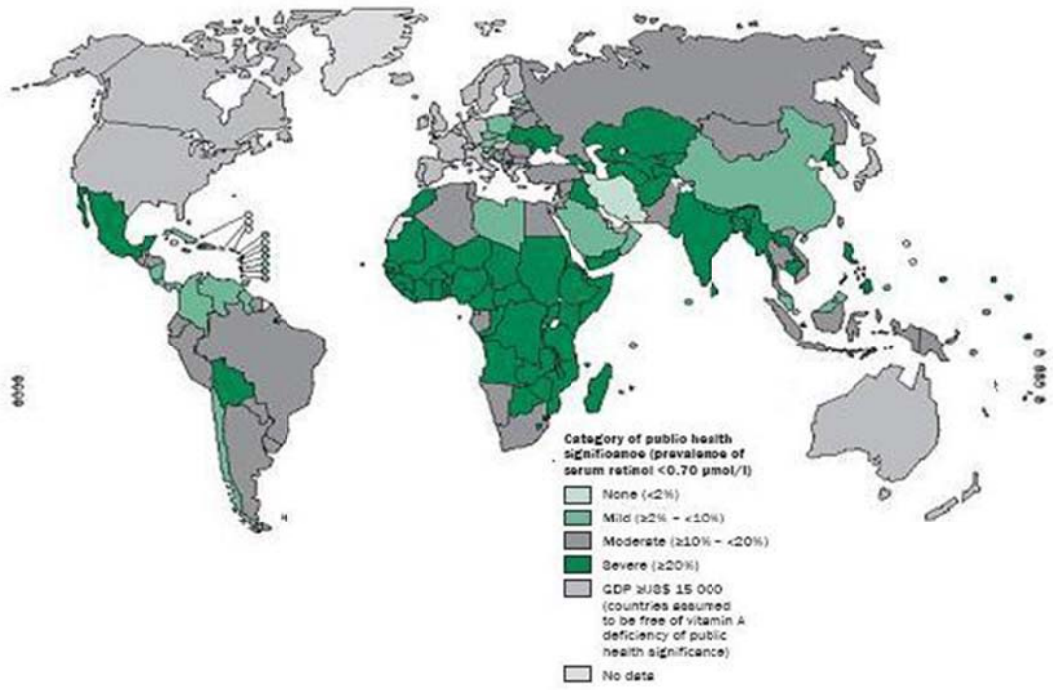


Figure 2: Global prevalence of vitamin A deficiency. Countries and areas with survey data and regression-based estimates: Preschool-age children (Global Prevalence of Vitamin A Deficiency in Populations at Risk, 2009, World Health Organization (WHO).

pulmonary disorders like lung agenesis, tracheoesophageal fistula, and bronchopulmonary dysplasia (Babu and Sharmila, 2010; Kling and Schnitzer, 2007). Very recently it was demonstrated in vivo that *β*-carotene synthesizing intestine-adapted bacteria in mouse is able to deliver vitamin A to the host tissues, indicating a novel way to treat VAD efficiently (Wassef *et al*., 2014).

Carotenoids also show antioxidant properties (Beecher, 1998; Leonardi *et al*., 2000; Stahl and Sies, 1998). Carotenoids derive their antioxidant ability predominantly from their conjugated double bonds but cyclic carotenoids gets this ability partly from their 5, 6 and 5′, 6′ double bonds in their cyclic end groups (Stahl and Sies, 1996; Mortensen *et al*., 2001). The antioxidant activity of carotenoids is based on their ability to quench singlet oxygen and ability to trap peroxyl radicals (Stahl and Sies, 1998). Carotenoids prevent lipid peroxydation mainly by quenching singlet oxygen (Stahl and Sies, 1998). Quenching of singlet oxygen by a carotenoid renders it in an excited state and this additional energy will be later dissipated by the molecule via a succession of rotational and vibrational interactions with the solvent, finally restoring back to the un-excited state making it available for additional cycles of singlet oxygen quenching. Carotenoids act as scavengers of peroxyl radicals especially at low oxygen tension by binding with them forming resonance stabilized adduct radicals which undergo decay to generate non radical products (Burton and Ingold, 1984; Rice-Evans *et al*., 1997; Woodall *et al*., 1997).

Carotenoid intake is associated with a decrease in incidence of chronic diseases. There are different postulated mechanisms for this activity like their pro-vitamin A property, anti oxidant property, modulation of enzymatic activity of lipoxygenases and

regulating expression of connexin protein which is an important component in the gap junctions that are necessary for cell to cell communication (Bendich, 1993). An increase in cell to cell communication probably due to increased connexin might be playing a role in the anti-cancerous property of carotenoids (Sies and Stahl, 1997).

2.4.4. Factors influencing carotenoid accumulation

The accumulation of carotenoids in tomato during ripening process is influenced by many factors.

Temperature: The most conducive temperature for color development is between 12 and 21°C. Lycopene accumulation in fruits gets affected at temperatures below 10°C and above 30°C while *β*-carotene biosynthesis is less susceptible to temperature variations. Thus fruits that ripen in non-optimal conditions tend to be of orange color (Stevens and Rick, 1986).

Light: Light has an important role in ripening and color development of tomato. The spectrum of light has a prominent role with red light inducing maximum accumulation of carotenoids compared to white or green light. In an experiment it was found that harvested fruits when exposed to red light show about 2.3 fold increases in lycopene accumulation compared to fruits that were kept in dark. Interestingly far red light has an inhibitory role in carotenoid synthesis as fruits that are exposed to red light when subsequently treated with far red light tend to show a decrease in accumulation of lycopene. The effect of far red light as an inhibitor possibly indicate role of phytochromes in this process (Alba *et al*., 2000). In tomato fruits phytochromes regulate carotenoid

levels as well as the time required for phase transitions during ripening (Gupta *et al*., 2014).

Soil moisture and nutrients: A decrease in soil moisture content increased the total carotenoid and lycopene content (Matsuzoe *et al*., 1998). Similarly an increase in salinity up to 0.25% w/v NaCl also increased lycopene accumulation (De Pascale *et al*., 2001). Supplementation of nutrients like nitrogen, potassium and phosphorous increased lycopene content (Dumas *et al*., 2003; Fanasca *et al*., 2006) while high levels of calcium, chloride and sulfur levels in fertilizers has a negative impact on lycopene accumulation (Toor *et al*., 2006).

Growth regulators: Several studies had shown that different growth regulators like 2-(4-chlorophenylthio) triethylamine hydrochloride (CPTA), gibberellic acid and cycocel (2-chloroethyl trimethyl ammonium 3-chloride), chlormequat, alar (succinic acid 2,2-dimethylhydrazide), brassinosteroids, ethephon and auxins can increase carotenoid accumulation in tomato (Dumas *et al*., 2003; Vardhini and Rao 2002; Cohen 1996). In addition to this, osmotic solutions like sucrose also increase carotenoid accumulation in tomato (Télef *et al*., 2006).

2.4.5. Regulation of genes in carotenoid biosynthesis pathway

2.4.5.1. Level of gene expression

Different abiotic and biotic factors like quality and duration of light, circadian rhythm, osmotic and oxidative stress, hormones, post-transcriptional feedback regulation and developmental stage are known to affect carotenoid biosynthesis by directly or indirectly regulating the level of expression of genes involved in carotenoid biosynthesis.

Several mutants and transgenic plants with altered expression of different genes in carotenoid pathway had shown either increase on decreased accumulation of different carotenoids. Constitutive expression of the bacterial phytoene desaturase gene *crtI* tripled the concentration of ζ-carotene in tomato fruit (Römer *et al*., 2000). Also, transgenic overexpression of bacterial phytoene desaturase *crtI,* tomato lcyB, tomato cycB, or lycopene *β*-cyclase from the plant *Narcissus pseudonarcissus* increased accumulation of *β*-carotene (Ronen *et al.*, 2000; Wiebke and Ralph, 2009).

2.4.5.2. Developmental stage

Developmental stage is one of the key factors that influence carotenoid biosynthesis in tomato. The expression levels of two genes that are involved in biosynthesis of lycopene *Psy* and *Pds*, were up-regulated 10–20-fold at the breaker stage of ripening (Bramley, 1997) along with a concomitant decrease in expression of lycopene cyclases, *Lcy-b* (*CrtL-b*) and *Lcy-e* (*CrtL-e*) (Ronen *et al.*, 1999). Up-regulation of genes involved in carotenoid biosynthesis and fruit development has also been found in bell pepper (Kuntz *et al.*, 1992), melon (Aggelis *et al*., 1997) and satsuma mandarin (*Citrus unshiu* Marc) (Ikoma *et al*., 2001).

2.4.5.3. Photoregulation

Expression of several genes in the carotenoid biosynthesis pathway is regulated by light. *Phytoene synthase* (*PSY*) is a strongly light induced gene in carotenoid biosynthesis pathway. In Arabidopsis, during de-etiolation, light could up-regulate the expression of *PSY* gene which is the rate limiting step in phytoene synthesis by triggering the degradation of Phytochrome-interacting factor 1(PIF1), an important suppressor of *PSY* gene expression (Welsch *et al*., 2000). Along with the enzymes phytoene desaturase

(PDS), ζ-carotene isomerase (ZISO), ζ-carotene desaturase (ZDS) and carotenoid isomerase (CRTISO), the conversion of 15-*cis*-phytoene to *trans*-lycopene involves a light mediated photoisomerisation (Chen *et al*., 2010). *PDS* gene was over expressed through phytochrome mediated photomorphogenesis (Qin *et al*., 2007). Absence of light for a long period regulates *ZISO* gene expression (Chen *et al*., 2010). There are reports on the *cis–trans* photoisomerisation of lycopene in green tissues in the absence of CRTISO. But the exact molecular mechanism behind the photoisomerisation phenomenon and its regulatory role is yet to be elucidated (Yu *et al*., 2011).

Zeaxanthin is a major carotenoid whose synthesis induced upon light and it acts to limit the light induced damage to photosynthetic membranes. The expression of *β-hydroxylase* gene involved in zeaxanthin synthesis from *β*-carotene is light inducible (Havaux and Niyogi, 1999). Post-translational regulation of violaxanthin de-epoxidase (VDE) an enzyme involved in synthesis of zeaxanthin from violaxanthin also plays a critical role in regulating the zeaxanthin level in response to excessive light (Rockholm and Yamamoto, 1996). This reverse conversion of violaxanthin to zeaxanthin is known as xanthophyll cycle and is believed to be involved in photoprotection by thermal dissipation of excess light energy (Havaux and Niyogi, 1999).

2.4.5.4. Epigenetic control

Carotenoid biosynthetic pathway genes are upregulated by epigenetic mechanisms. The SET DOMAIN GROUP (SDG) family of histone lysine methyltransferases regulates the expression of *CRTISO* in Arabidopsis, by methylating the lysine residues on the tail of histone proteins. This allows other regulatory factors to play their role and induces the transcription of the gene by enhancing the accessibility of

RNA polymerase II complex to this region. The chromatin marks formed by SDG is maintained through mitotic divisions and are reset during meiotic divisions at the time of gamete formation. Regulatory mechanism of SDG enzyme is proved to be restricted to CRTISO as described by the microarray analysis of the SDG mutant (Cazzonelli *et al*., 2009 a,b,c; Ding *et al*., 2011; Saze, 2008).

2.4.5.5. Feedback regulation

Negative and positive feedback regulatory process are playing major role in the regulation of many genes involved in carotenoid biosynthetic pathway through a carotenoid or their degradation products. Feedback regulations are generally operated for genes involved in the synthesis of isoprenoid precursors, for the phytoene synthase gene which works at the rate limiting step of carotenoid biosynthesis or at the bifurcation step converting lycopene to α- and β-carotenes (Cuttriss *et al*., 2007; Qin *et al*., 2007; Bai *et al*., 2009).

In the etiolated Arabidopsis seedlings, the DXS protein induces a positive feedback regulation of *PSY* by increasing the supply of MEP substrates. Arabidopsis seedlings treated with paclobutrazol which is a potent inhibitor of gibberellin biosynthesis enhanced carotenoid levels and PSY activity in the absence of light probably by donating precursor to carotenoid pathway. This enhanced *PSY* expression was correlated with DXS and DXR protein abundance and hence concluded to be due to the positive feedback regulation. Nevertheless, carotenoid accumulation was not enhanced in dark-grown seedlings over-expressed with *DXS* mRNA (Rodríguez-Villalón *et al*., 2009). Despite the low transcript level of *DXS*, an enhanced accumulation of active DXS enzymes is

predicted to increase the *PSY* transcript levels in tomato fruit chromoplast tissues (Fraser *et al*., 2007; Rodríguez-Villalón *et al*., 2009).

The expression of *PSY* is regulated by source and sink metabolites. *OsPSY3* and *OsNCED* gene expression is induced by the external application of ABA. Moreover, an ABA responsive element and coupling element are present in the *OsPSY3* promoter (Welsch *et al*., 2008). An interesting feedback regulation was reported with the plant root arbuscular mycorrhizal (AM) fungi established in the rhizosphere that it can enhance the expression of genes like *DXS*, *DXR* (MEP pathway), *PDS*, *ZDS* (carotenoid biosynthesis pathway) and carotenoid-cleaving dioxygenases in the roots (Strack and Fester, 2006). Nevertheless, the molecular mechanism behind this positive regulation is yet to be elucidated.

Eventhough an increase in *β*-carotene levels was reported in *eLCY* silenced potato (*Solanum tuberosum*) by Diretto *et al*. (2006), the synthesis of *α*- and *β*-carotenes from all *trans*-lycopene is described to be under negative metabolite feedback regulation (Cuttriss *et al*., 2007). An increase in total carotenoids to 200% and 40% in the b*LCY* mutants was observed in *B. napus* and maize respectively and the major carotenoids accumulated were lycopene and δ-carotene (Bai *et al*., 2009).

2.4.5.6. Role of transcription factors in fruit ripening

Various single gene mutations affecting the fruit ripening process have been identified in tomato. Out of them few mutations in transcription factors are controlling the switch to the ripening phase. Map based cloning and chromosomal walking could identify the genes responsible for the mutations rin (ripening-inhibitor), nor (non-ripening) and Cnr (Colourless non-ripening) as *MADS-RIN* (*MCM1/AGAMOUS/DEFICIENS/SRF-*

RIPENING INHIBITOR), *NAC-NOR (NO APICAL MERISTEM /ATAF1/2/CUC2)* and *SPB* (*SQUAMOSA PROMOTER BINDING protein*) respectively (Giovannoni, 2007; Manning *et al*., 2006; Vrebalov *et al*., 2002). Many researchers suggest a network of these genes operating to control many other genes involved in fruit ripening. It is been established that RIN requires CNR as an essential factor to function (Martel *et al*., 2011). Till now, a number of such transcription factors controlling fruit ripening in tomato are identified.

Orthologues of many of these genes have also been identified in many other species controlling their fruit development (Seymour *et al*., 2008). Valve function and development of the dehiscence zone in the Arabidopsis are regulated by genes operating in a network. They include a SQUAMOSA class MADS-box gene, *FRUITFULL (FUL)* and AGAMOUS subgroup MADS-box genes *SHATTERPROOF (SHP)* (Dinneny *et al*., 2005). *FUL* is an orthologue of tomato *TDR4* (*TOMATO DEFICIENS-RELATED4*) gene whose role in ripening is unknown even though it has a ripening-related expression pattern, whereas, *SHP* is an orthologue of tomato *TAGL1* (*TOMATO AGAMOUS-LIKE 1*) which plays a regulatory role in tomato fruit ripening and fruit fleshiness (Itkin *et al*., 2009; Vrebalov *et al*., 2009). Tomato fruit ripening and ethylene production are inhibited when *LeHB1* (*Lycopersicum esculentumhomeobox1*) was silenced. But, an overexpression of this gene caused the modification of sepals to carpel-like structures. LeHB1 interacts with *ACO1* (*ACC oxidase1*) promoter (Lin *et al*., 2008).

Eventhough many mutations affecting the tomato fruit ripening has been reported, rin mutation is most studied at physiological and molecular level. Expression of *RIN* correlates with the fruit ripening process and also affect the lycopene accumulation, tissue

softening etc. RIN binds to the specific DNA sequences known as (CArG) box which was identified as typical binding sequence for MADS-box proteins. It was proved by in-vitro assays that RIN binds to (C/T)(A/T)6(A/G)G, but 'CCA(A/T)(A/t)(A/T)ATAG' is the preferred sequence (Ito *et al*., 2008). Microarray assays showed the suppression of several ripening associated genes in *rin* mutant, indicating the potential control of this gene on other ripening-related genes RIN can regulate the expression of large number of ripening-associated genes by three possible ways. 1) Direct binding of RIN to the promoter of target genes 2) Enhancing the expression of ethylene induced genes through the induction of ethylene during ripening and 3) Induction of ripening-specific gene expression by enhancing the expression of other transcription factors (Fujisawa *et al*., 2011). Eventhough there are many evidences of operation of the regulatory networks of RIN, NOR, CNR, TAGL1 and HB1 by forming heterodimers or multimers, the knowledge on the precise molecular mechanism involved in these networks is lacking (Giovannoni, 2007). Many ERFs (Ethylene Responsive Factors) are also known to have their role in fruit set and early stages of fruit ripening process (Pirrello *et al*., 2012). An AP2/EREBP family transcription factor RAP2.2 binds on *PSY* promoter altering carotenoid pigment synthesis in Arabidopsis (Welsch *et al*., 2007).

In addition to the positive regulators of fruit ripening, certain transcription factors were identified with negative regulatory roles in fruit ripening Karlova *et al*., 2011). They include AP2 (APETALA2), proteins encoded by tomato orthologues of Arabidopsis *DDB-1*(*UV-DAMAGED DNA-BINDING PROTEIN-1*) and *DET-1*(*DE-ETIOLATED-1*) etc. Mutations in *DDB-1*and *DET-1* orthologues resulted in *hp-1* and *hp-2* phenotypes

respectively in tomato characterized by high fruit carotenoid levels in addition to other phenotypic changes (Mustilli *et al*., 1999; Enfissi *et al*., 2010).

2.5. Lycopene cyclases and its homologues

2.5.1. Lycopene beta cyclase

A single gene coding for lycopene beta cyclase is present in case of plant species such as *Arabidopsis thaliana* (Lange and Ghassemian 2003), *Oryza sativa* (Fang *et al*., 2008), and *Zea mays* (Bai *et al*., 2009) where, there is low level of accumulation of carotenoids. But in case of plants which accumulate high levels of carotenoids in non-photosynthetic organs, two different genes coding for lycopene beta cyclases, LCYB and CYCB are present. They differ in cellular localization with LCYB having chloroplast specific function and CYCB having chromoplast specific function. *LCYB* expresses preferentially in photosynthetic tissues while *CYCB* expresses in non-photosynthetic tissues like ripening fruits and flowers (Pecker *et al*., 1996; Bouvier *et al*., 1994). Such differentially expressing genes have been described in tomato (Ronen *et al*., 2000) *Capsicum annuum* (Hugueney *et al*., 1995), *Carica papaya* (Blas *et al*., 2010), *Croccus sativus* (Ahrazem *et al*., 2010) *Citrullus lanatus* (Bang *et al*., 2007) etc.

Tomato (*S. lycopersicum*) fruits are predominantly red because of lycopene accumulation in chromoplast during fruit ripening. In '*Beta*' mutant of tomato a change in the promoter sequence had resulted in enhanced expression of *CYCB* gene in ripening fruits causing conversion of lycopene into *β*-carotene resulting in orange colored fruits. In '*old gold* (*og*)' mutants of tomato frame shift mutations had resulted in loss of function of *CYCB* gene resulting in dark red fruits due to more lycopene accumulation. A transgenic tomato plant expressing antisense *CYCB* had also resulted in deep red colored

fruits while another transgenic which over express *CYCB* had given fruits with increased *β*-carotene accumulation in columellar and placentar region of fruits (Ronen *et al*., 2000).

As in the case of tomato two functionally characterized lycopene cyclases CitLCYb1 and CitLCYb2 are present in the citrus species *Citrus unshiu*, *Citrus sinensis*, *Citrus limon* etc. *CitLCYb1*expresses in chloroplast while *CitLCYb2*expresses in chromoplast. The expression of *LCYb2* is highly induced in the fruit peel and pulp of *Citrus sinensis* during ripening indicating the role of this gene in accumulation of *β,β*-xanthophylls in fruits. Conversely, *LCYb1* shows constitutive but low level of expression (Alque´ zar *et al*., 2009). Several lycopene accumulating mutants are known in grape fruit (*Citrus paradisi*) and pummel (*Citrus grandis*) and three in orange (*Citrus sinensis*): Shara, Cara Cara, and Hong Anliu, where, the phenotype is suspected to be because of mutations in *LCYb* gene (Monselise and Halevy, 1961; Lee, 2001; Liu *et al*., 2007).

In *Carica papaya* two lycopene cyclase genes *LCYB1* and *LCYB2* are known. The expression of *LCYB2* increased in fruits upon ripening while both the genes showed low level of constitutive expression in leaves (Devitt *et al*., 2010). In *Dacus carota,* out of the two lycopene cyclase genes *DcLCYB1* gene is expressed in both leaves and storage roots. Increased expression of *DcLCYB1* gene correlating the enhanced accumulation of carotenoids in mature carrot roots indicated the involvement of this gene in carotenoid accumulation (Fuentes *et al*., 2012). An *LCYB* gene (*Lyc-β Fc)*was cloned from *Ficus carica* and heterologous expression of this gene in an *E. coli* strain that accumulate lycopene resulted in 90% conversion of lycopene to *β*-carotene (Araya-Garay *et al*., 2011). Similarly, a *Vitis vinefera LCYB* gene (*VvLBCY2*) was also cloned and it was reported to be able to convert 72% percent of the accumulated lycopene into *β*-carotene

(Young *et al*., 2012). A lycopene beta cyclase gene was isolated from water melon (*Citrullus lanatus*). This gene shows sequence polymorphisms which co-segregate with yellow and red flesh colour of fruit (Bang *et al*., 2007). The orange fruited pepper varieties harbor an *LYCB* gene allele that is different from that of the red fruited pepper varieties (Hugueney *et al*., 1995).

All plant lycopene cyclases have a conserved pyridine nucleotide binding site and an N-terminal signal peptide that directs the protein into plastids (Pecker *et al*., 1996; Krubasik and Sandmann, 2000; Sandmann *et al*., 2006; Hugueney *et al*., 1995). In addition to a conserved FAD/NAD(P) binding sequence motif/ dinucleotide binding motif LYCBs have cyclase motifs and plant LCYB specific motif which might have roles in membrane attachment and substrate interaction (Hugueney *et al*., 1995). Typically the signal peptide in lycopene beta cyclase is the first 100 to 150 amino acids (Sandmann *et al*., 2006; Hugueney *et al*., 1995).

2.5.2. Lycopene epsilon cyclase

LCYE could act on one end of lycopene molecule to form a ε-ring producing δ-carotene which when acted upon by LCYB produces α-carotene (Bang *et al*., 2007). *LCYE* isolated from lettuce (*Lactuca sativa*) (*CRTL-E*) is an exception as it codes for a bi-cyclase which could convert lycopene to α-carotene (Cunningham and Gantt 2001). In a high amount of sequence similarity with about 30% amino acid identity is observed between LCYE and LCYB in plants like Arabidopsis and tomato. Like LCYB, LCYE also contain a characteristic FAD/NAD(P) binding sequence motif (Pecker *et al*., 1996). CitLCYe when over expressed in lycopene accumulating strain of *E. coli* along with CitLCYb1 and CitLCYb2, resulted in production of α-carotene (Zhang *et al*., 2012).

Likewise VvLECY1, a putative lycopene ε-cyclase in *Vitis vinifera* L. converted 42.1% of the available lycopene to δ-carotene in heterologous expression studies (Young *et al*., 2012).

2.5.3. Capsanthin capsorubin synthase

Oxygenated derivatives of carotenoids are called xanthophylls. β-carotene acts as the precursor of xanthophylls like antherxanthin and violaxanthin. CCS converts antheraxanthin to capsanthin and violaxanthin to capsorubin. Capsanthin and capsorubin are red colored xanthophylls that give pepper the usual red color. *CCS* gene shows very much homology to *CYCB* gene with pepper *CCS* gene showing 86.1% identity to tomato *CYCB* gene (Bouvier *et al*., 1994). Also the genomic location of *CCS* in pepper is identical to the genomic location of *CYCB* in tomato (Thorup *et al*., 2000). In pepper a mutation in *CCS* gene is known to give rise to yellow fruits (Guzman *et al*., 2010). Also a lot of cultivar specific variations in the promoter of *CCS* gene are known in red, orange and yellow fruited pepper varieties (Ha *et al*., 2007). In tiger lilly (*Lolium lancifolium*) the orange colour of flower petals is due to accumulation of capsanthin and capsorubin. The expression of *Llccs* is restricted to flowers and flower buds but absent in vegetative tissues (Jeknic *et al*., 2012).

2.5.4. Neoxanthin synthase

Neoxanthin is considered as the final product of carotenoid synthesis and is a precursor to the hormone abscisic acid (Parry *et al*., 1990). In plants neoxanthin is synthesized from violaxanthin (Galpaz *et al*., 2008). The enzyme which is involved in this conversion is neoxanthin synthase (NXS) and the gene coding for this belongs to the lycopene cyclase gene family (Al Babili *et al*., 2000; Bouvier *et al*., 2000). Expression in

E.coli produced a 56-kDa protein for tomato NXS and was found to be catalyzing the conversion of violaxanthin to neoxanthin. Transient expression of the same in tobacco leaves also proved the functionality of the enzyme (Bouvier *et al*., 2000). The very high similarity (99%) of tomato CYCB and NXS suggest the possibility of a single enzyme performing two different functions in tomato. However, the unaltered neoxanthin synthesis in tomato *CYCB* null mutant *ogc* (Ronen *et al*., 2000) explains the existence of an independent *NXS* gene in tomato. Pepper is known to produce neoxanthin in green unripe fruits but no *NXS* gene is cloned from it yet (Guzman *et al*., 2010). The Arabidopsis *aba4* gene (North *et al*., 2007) and another putative gene from tomato (Neuman *et al*., 2014) which are involved in the conversion of violaxanthin to neoxanthin suggest the existence of *NXS* which does not belong to lycopene cyclase gene family. However, the identity of these enzymes is not confirmed by their in-vitro activity.

2.6. Functional genomics approaches in tomato

The present picture on the global requirement and studies on nutritional value pinpoints the importance of increasing the yield as well as enhancing the nutrient levels in tomato. Further, because of the simple diploid genetics, small genome size, fast propagation, short life cycle and nominal fruit phenotypes at different stages, it is accepted as a model system for functional genomics studies in dicotyledonous plants (Barone *et al*., 2008). Protocols for several functional genomics strategies have been developed and many bioinformatics resources are also available in tomato. Tomato Expression Database (TED) (Fei *et al*., 2006) got expanded to Tomato Functional Genomics Database (TFGD) with other 'omics' data in addition to the transcriptomics data (Fei *et al*., 2011). Resources of mutants, wild species, TILLING populations, gene

silenced tomato lines, VIGS libraries (for transient silencing), mapping population, BAC libraries etc are also available for genomic research in tomato. Transcript and protein analysis, screening of posttranslational modifications and protein-protein interactions, metabolite analysis to generate metabolic networks etc are also being carried out for a better understanding of structural and functional aspects of its genome. Tomato genomic resources have been developed in the last two decades in the form of linkage maps with various markers, ESTs, full length cDNA sequences, gene expression profiles, and genome sequences with annotations. Currently, most information from these resources has been released to databases in the public domain such as the National Center for Biotechnology Information (NCBI), the DNA Data Bank of Japan (DDBJ), the Solanaceae Genome Project Network (SGN), and the J. Craig Venter Institute (JCVI, formed through the merger of several organizations including The Institute for Genomic Research, TIGR) (Yano *et al*., 2007).

The International Solanaceae Genome Project (SOL) and the Solanaceae Genomics Network website (SGN; http://www .sgn.cornell.edu) are playing major roles in the advance of tomato genomic researches worldwide. While SOL coordinates the Solanaceae research of different groups and develops knowledge network, SGN facilitate distribution of genomic information for tomato in particular and for Solanaceous species in general in a comparative genomic context (Mueller *et al*., 2005 a,b). The full sequence information of tomato genome along with "-omics" and next generation sequencing techniques will lead to high throughput functional genomics for tomato diversity studies and expression studies to correlate transcriptome, proteome and metabolome levels (The tomato genome consortium, 2012).

2.6.1. Forward genetics

Approaches like mutagenesis, genetic transformation, and transcriptome analysis have been exploited in understanding the gene functions, identifying their role in metabolic pathways and studying the regulatory mechanism of genes at various stages. Forward genetic approaches are based on searching for the gene responsible for a specific phenotype produced. Tomato Genetics Resource Center (TGRC) at the University of California, Davis has a collection of 1,017 monogenic mutants at 622 loci affecting many aspects of plant development (http://tgrc.uctavis.edu). The mutant collection contains spontaneous and induced mutations brought to different backgrounds. In addition TGRC has a large collection of wild relatives and around 1,500 miscellaneous genetic stocks including landraces, cultivars, prebred lines, introgression lines, backcross recombinant inbred lines, stress-tolerant stock, and cytogenetic stock containing trisomics, tetraploids, and translocations. The seeds stored in the TGRC collection are available for research purpose on request. "Genes That Make Tomatoes" (http://Zamir.sgn.cornell.edu/mutants/) is a site in SGN, where a catalogue of 3,400 isogenic mutations in the genetic background of the inbred variety 'M82'is available (Menda *et al*., 2004). All these mutant resources are playing a catalytic role in boosting up the gene function identification through forward genetics. A major drawback with forward genetics techniques is that they require longer time to correlate phenotype with genotype compared to reverse genetics techniques.

2.6.2. Reverse genetics

The functional genomics approach developed to identify the gene function by analyzing and correlating the genetic variations in the gene of interest to the phenotypic variations is known as reverse genetics. It is based on silencing or overexpression of the

targeted gene and is complementary to forward genetic approaches. Several reverse genetics approaches are being developed in tomato and most of them are based on transformation. Development of a well standardized and efficient transformation protocol was a good support for all these techniques (Sun *et al*., 2006). Insertional mutagenesis is a reverse genetic strategy where exogenous transposon systems have been exploited (Meissner *et al*., 2000). Gene silencing is one of the powerful tool helps in the identification of gene functions. Sense and antisense silencing had been successfully used in tomato fruit ripening (Smith *et al*., 1990; Gray *et al.,* 1992). The first transgenic plant released to the fresh market was *Flavr-savr*® tomato which is a delayed ripening tomato due to reduced expression of polygalacturonase in cell wall. It was obtained by antisense silencing and could improve the fruit shelf life and storage quality (Kramer and Redenbaugh, 1994; Sanders and Hiatt, 2005). But due to consumers concern about GMO, it become a commercial failure (Bauchet and Causse, 2012). RNA interference (RNAi) and virus induced gene silencing (VIGS) are other gene silencing tools proved successful in tomato functional genomics recently. Availability of extensive EST collection in tomato supported the transcriptional profiling (Mueller *et al*., 2005). A new strategy in reverse genetics which is based on the identification of induced mutations in gene of interest was described by McCallum *et al*., (2000) called Targeting Induced Local Lesions IN Genomes (TILLING). Metabolomics has an important role to play in characterization of natural diversity in tomato (Menda *et al*., 2004; Fernie *et al*., 2011). As well, it can boost the biochemical understanding of fruit content and be an enhancer for quality breeding (Fernie and Schauer, 2009; de Vos *et al*., 2011).

2.6.2.1. TILLING and EcoTILLING

By taking full advantage of the nearly completed genome sequence information, reverse genetic tools can be used for identifying mutants of gene of interest. Unlike approaches like RNAi, gene knockout, site-directed mutagenesis and transposon tagging which rely on the creation of transgenic material, TILLING is a non-transgenic reverse genetics approach to identify EMS-induced point mutations (McCallum *et al*., 2000a, b). In contrast to transgenic methods, mutagenesis is random, cost effective and is not submitted to GMO regulation. The biggest advantage with TILLING is that variants can be generated and transfer rapidly into any cultivated genetic background (Nicolaï *et al*., 2010; Gady *et al*., 2009).

TILLING has been applied successfully in many organisms belonging to different kingdoms. Successful examples are available in drosophila (Winkler *et al*., 2005), zebra fish (Sood *et al*., 2006), Arabidopsis (Till *et al*., 2006), wheat (Slade *et al*., 2005), maize (Till *et al*., 2004), soybeans (Cooper *et al*., 2008), lotus (Perry *et al*., 2003) etc. Several countries (USA, France, Italy, Japan and India) are currently involved in the development and high throughput screening of their populations in tomato through TILLING (http://tilling.ucdavis.edu/index.php/Tomato_Tilling; Minoia *et al*., 2007; Watanabe *et al*., 2007; Srinivasan *et. al*., 2006). Various mutagenesis strategies have been adopted by different groups to make the populations. They include chemical (e.g., EMS), physical (e.g., X-ray or fast-neutron irradiation), and insertional (e.g., transposable elements or T-DNA) mutagenesis (Hildering and Verkerk, 1965; Thomas *et al*., 1994; Meissner *et al*., 1997; Li *et al*., 2001; Emmanuel and Levy, 2002; Menda *et al*., 2004). Generally, physical mutagens are less preferred for induced mutagenesis because they induce

chromosomal breakages leading to larges deletions generating truncated genes and lethal alleles (Naito *et al*., 2005). The mutagenesis rate of gamma-rays is estimated at the rate of one per 6.19 Mb by Sato *et al*. (2006).

Among chemical mutagens ethylmethanesulfonate (EMS) is most effective and so is widely used for mutagenesis. EMS generally induces the transition A (Adenine) to G (Guanine) which may result in stop codon, synonymous or nonsynonymous changes. Based on the codon usage in a species, various sublethal and substerile alleles may be generated. It causes very low (<5%) truncations compared to other mutagens (Greene *et al*., 2003). Repositories of EMS mutagenised seeds and DNA is maintained by several groups around the world (Menda *et al*., 2004; Minoia *et al*., 2010, Okabe *et al*., 2011).

One drawback with TILLING is the time required to develop a mutagenised population after balancing lethality with adequate mutation frequency. This can be avoided by using a collection of natural accessions with high variability and is termed as EcoTILLING (Comai *et al*., 2004). Single Nucleotide Polymorphisms (SNPs) and small In-Dels are reported to be the most common form of nucleotide variation in natural population (Perkel, 2008). Eventhough sequencing is the most accurate approach; the relatively high expense when applied in large number of individuals limits its use in detection of nucleotide variations (Till *et al*., 2004). The first successful example reported for detection of genetic variations associated with target traits for crop improvement was that of Arabidopsis population (Comai *et al*., 2004). 19 elite inbred lines with SNPs and indels in *LHCP* region in sunflower was identified by Fusari *et al.* (2008). Allelic variations in *mlo* and *Mla* resistance genes (Mejlhede *et al*., 2006) and drought resistance genes (Cseri *et al*., 2011) were identified in barley. A low erucic acid (LEA) resource is

discovered in *B. rapa* (Wang *et al.*, 2010). Five SNPs in *Lhcb1* gene of barley are potentially identified as marker for crop improvement (Xia *et al*., 2012). New allelic variants of eIF4E, that control virus susceptibility was reported in *Cucumis sp*. (Nieto *et al*., 2007). EcoTILLING was successfully used in the study of drought tolerance in rice (Yu *et al*., 2012). It can also be used to study the natural variations in case of out breeding species also as reported in case of *P. trichocarpa* (Gilchrist *et al.*, 2006).

Generation and identification of promising alleles suitable for improving nutritional quality traits in tomato could be done using TILLING and EcoTILLING methods. Tomato is one of the richest sources of lycopene and other carotenoids like *β*-carotene. Quantity and relative proportion of these nutrients in tomato could be further improved using TILLING. The rich genetic diversity in tomato and the free availability of tomato germplasm collection to researchers all over the world through various sources makes EcoTILLING a good strategy for identifying candidate alleles for improving carotenoid biosynthesis in commercial cultivars.

Chapter 3

Materials and Methods

3.1. Materials

3.1.1. Plant material

An EcoTILLING population consisting of 543 accessions was used for screening of naturally occurring SNPs. The accessions were obtained from NBPGR (New Delhi, India), IIVR (Varanasi, India), TGRC (California, USA) and Bejo Sheetal Seeds (Jalna, India). *S. lycopersicum cv.* Arka Vikas (originally obtained from Indian Institute of Horticulture Research, Bangalore) was used as the reference variety. The list of these accessions with accession numbers and other details is given as Annexure I.

3.1.2. Chemicals and consumables

All chemicals for DNA isolation and PCR were from Sigma-Aldrich, MO, USA and the organic solvents used in the procedure were from Qualigens Limited, India. Stock solutions for DNA isolation and EcoTILLING were prepared in sterile MQ water (Millipore water purification system, Millipore, USA). IRDye labeled oligonucleotides used for EcoTILLING were procured from Biomer, Germany and all other oligonucleotides from BioServe Biotechnologies (India) Pvt Ltd.100 µM solution of each primer was prepared in TE buffer (10 mM Tris-HCl, pH 7.5; 1 mM EDTA, pH 8.0) and aliquots were stored at ¯80°Cto avoid repeated freeze–thaw cycles that could reduce fluorescent activity. dATP, dCTP, dGTP and dTTP required for PCR were obtained from Merck, Germany.

3.2. Methods

3.2.1. Extraction and estimation of total carotenoids

Extraction of total carotenoids was carried out using the protocol described by Sadler *et al*. (2006) with few modifications. About 0.5 g of red ripe fruit tissue (pericarp and mesocarp) was homogenized with mortar and pestle. The homogenate was filtered using a sieve to remove the debris. 50 µl of homogenized extract was transferred to a 2.0 ml microcentrifuge tube and 1.25 ml of organic mixture (Hexane: Acetone: Absolute Alcohol in 2:1:1, v/v/v ratio) was added. The mixture was thoroughly vortexed and kept on ice. To this 195 µl distilled water was added, mixed by inverting the tubes and centrifuged at 10,000 rpm for 10 min at 4°C. 1 ml of the extract was separated and used for estimation.

A method based on mean absorption wavelength (451 nm) and mean absorption coefficients (Biehler *et al*., 2010) was used for estimation. $A_{451nm}$of the supernatant was measured using UV spectrophotometer (UVIKON Spectrophotometer 922, Kontron Instruments) in a 1 cm path length quartz cuvette. All the steps in the extraction and estimation were carried out under dark condition. The amount of total carotenoid content in µg/100g fresh weight (FW) tissue was calculated using the formula described by Rodriguez-Amaya and Kimura, 2004.

Total carotenoid content (µg/100g) =

$$\{A_{451} \text{ x Volume (ml) x } 10^4 \text{ x100}\} \,/\, \{A^{1\%}_{1cm} \text{ x sample weight (g)}\}$$

Where, A_{451}= absorbance at 451 nm; Volume = total volume of extract (1.25 ml) $A^{1\%}_{1cm}$=absorption coefficient of a carotenoid (absorbance at a given wavelength of a 1%

solution in 1 cm light-path spectrophotometer cuvette). The mean absorption coefficient for various carotenoids (Rodriguez-Amaya and Kimura, 2004) was calculated as 2701. Sample weight=0.05 g (almost equivalent to 50 µl of homogenized extract).

3.2.1.1. Statistical analysis

Mean, standard deviation (SD) and coefficient of variation (CV) were calculated for total carotenoid contents of the fruits from 183 accessions. The accessions with values > (mean + SD) and > (mean + 2SD) were identified as desirable and highly desirable ones, whereas, those with values < (mean – SD) and < (mean – 2SD) were identified as undesirable and highly undesirable ones. All others were considered as with average performance (Shakhatreh *et al*., 2010).

3.2.2. EcoTILLING

3.2.2.1. Seed germination and cotyledon harvesting

The seeds were surface sterilized with 2% (v/v) sodium hypochlorite solution for 15-20 min at room temperature and were washed thoroughly under running tap water. Seedlings were grown in wells (perforated at bottom) of 96 deep well plates (2 ml) filled with Soilrite mixture (Keltech Energies Limited, Karnataka, India). The plates were kept in large plastic trays filled with 1 cm water. Seedlings were grown under white fluorescent light (100 µmol/m^2/sec) in growth room at 25±2°C. Cotyledons from ten days old seedlings were harvested and placed in the wells of sterile 96 deep well plates (2 ml) (Axygen Limited, India) for DNA extraction.

3.2.2.2. DNA extraction

Large scale DNA extraction was done by a protocol developed in lab (Sreelakshmi *et al*., 2010). Cotyledon tissues (80-100 mg) were ground with three steel balls of ~2 mm in diameter in Mini Bead Beater (BioSpec Products Inc.) for 2 min in the presence of 750 µl preheated (65°C) extraction buffer (0.1 M Tris-HCl, pH 7.5; 0.05 M EDTA, pH 8.0; 1.25% (w/v) SDS) containing 0.2 M *β*-mercaptoethanol and 20 mg of insoluble polyvinylpolypyrrolidone (PVPP). The plate was then incubated at 65°C for 30 min. Contaminating RNA was removed by adding 4 µl RNase (10 mg/ml) and subsequent incubation for 30 min at 37°C. The plates were kept at room temperature for 10 min to reduce the sample temperature. 400 µl of cold 6 M ammonium acetate was added to the samples, mixed well and incubated for 15 min at 4°C. The precipitated protein and other debris were separated by centrifuging the plates at 4700 rpm in a swing out plate rotor (SH-3000) in Sorvall RC Evolution centrifuge for 30 min. 650 µl of the clear supernatant was transferred to fresh 2 ml 96-well plate. For DNA precipitation, equal volume of ice cold isopropanol was added to the supernatant, mixed well and incubated at ¯20°C for 1 hr. The plates were centrifuged at 4700 rpm for 30 min at 4°C. Two 70% (v/v) ethanol washes were given to remove traces of salts from the samples. The plates were kept at 65°C for 10-15 min to dry the DNA pellet. 200 µl TE (10 mM Tris, pH 7.5, 1 mM EDTA pH 8.0) supplemented with 3.2 µg/ml RNase, was added to each of the wells and the plates were kept at 4°C overnight for dissolution. To pellet any undissolved material, the plates were centrifuged at 4700 rpm at 4°C for 30 min. 180 µl of the supernatant was transferred to fresh 1 ml plate using 96 well pipettor (PP550 DS, ApricotDesigns).

3.2.2.3. Quantification and equalization of DNA

The DNA samples were incubated at 65°C for 15 min and then transferred to ice. The DNA samples were quantified by comparative visual estimation of fluorescence based on a known standard (λ DNA) on 1% (w/v) agarose (SeaKem LE, Rockland, ME, USA) gel containing 0.5 µg/ml ethidium bromide. Electrophoresis was carried out in 1X TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA (pH 8.0)), at room temperature using 100-120 V current. The ethidium bromide stained DNA bands were visualized in a gel documentation system (Alpha Imager™ 2200, Alpha Innotech). The samples were equalized to 5 ng/µl concentration. Samples were further diluted to 1 ng/µl concentration and aliquoted to 96-well PCR plates (5 ng each) using 96-well pipettor. The aliquoted plates were vacuum dried and stored at -20°C.

3.2.2.4. SNP detection

Screening for SNPs was done by a protocol described by Till *et al*. (2006) and Colbert *et al*., (2001) with few modifications. The overall strategy of SNP detection is diagrammatically shown in Figure 3. The steps are explained in detail below.

3.2.2.4.1. Nested PCR strategy

In order to increase the concentration of PCR products and to increase the specificity, nested PCR approach was used. The first step PCR was carried out using unlabelled primers covering an extended gene region. The PCR products of this reaction was quantified on agarose gel using known standard and diluted to use as the template for the second step PCR. Second step reaction was carried out with a cocktail of labeled and unlabelled primers for promoter and exon separately. Figure 4 shows the strategy of nested PCR.

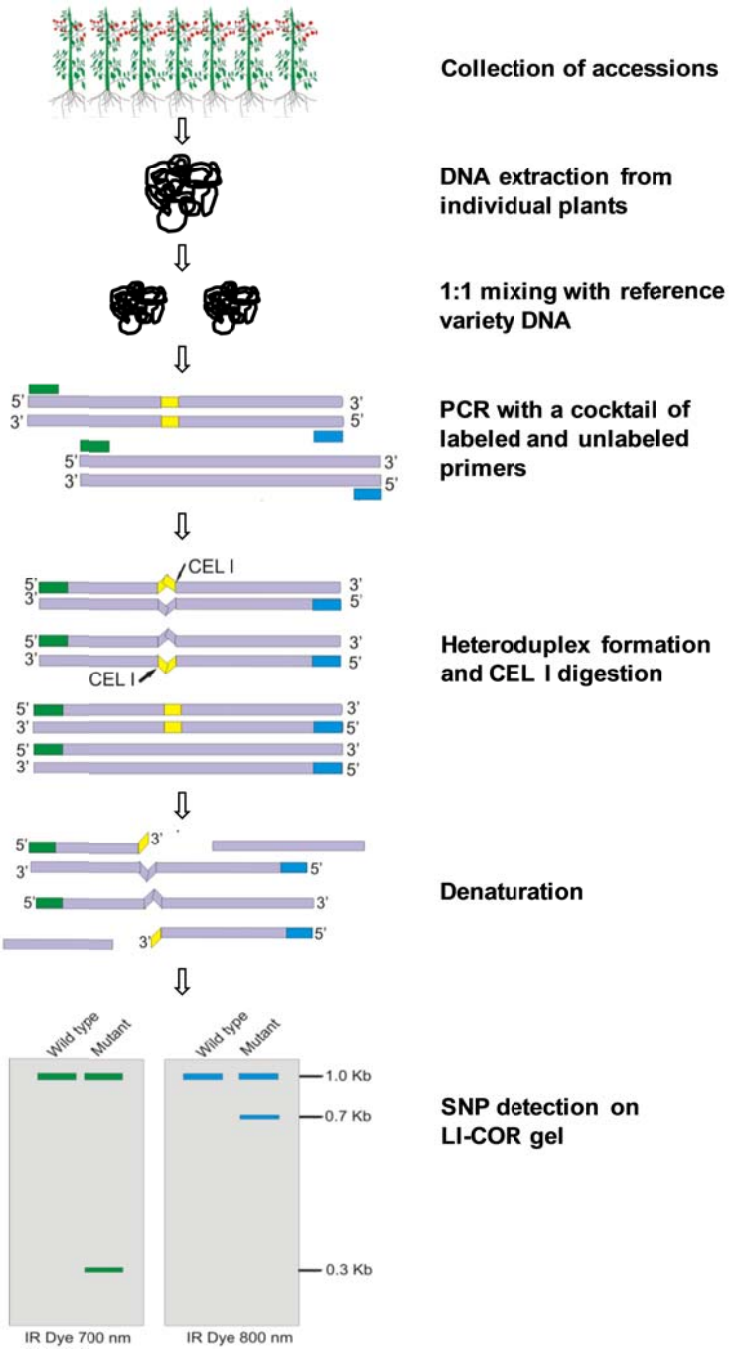


Figure 3: Schematic representation of SNP detection by Eco TILLING. Arka Vikas was used as the reference variety. Green and blue colors indicate $IRDye_{700}$ and $IRDye_{800}$ labeling respectively. Yellow color represents the position of an SNP. SNP detection was done by separating the CEL I digested fragments on 6.5 % polyacrylamide gel and detecting the complementary fragments (here 0.3 Kb and 0.7 Kb) on $IRDye_{700}$ and $IRDye_{800}$ channels. Undigested PCR products (here 1.0 Kb) are visible on both channels.

3.2.2.4.2. Primer designing

The sequence of *Beta* gene (Solyc06g074240.1) and around 1 Kb upstream region was obtained from SGN database (www.solgenomics.net/). The CODDLE (Codons Optimized to Discover Deleterious Lesions) was used as a primer design tool (http://www.proweb.org/coddle/) which emphasizes on the portions of the gene having a high probability for inducing point mutations which can bring about drastic phenotypic changes, by EMS (Ethyl Methane Sulphonate). One set of primers were designed to cover both the coding region and ~1 Kb upstream sequence of the coding region with an expected amplicon size of 2684 bp. The primers for the second step reaction were designed to amplify around 991 bp and 1231 bp fragments of the upstream sequence and the coding sequence respectively, from the first step product. In order to reduce the cost of IRDye labeled primers, common M_{13} forward and reverse tails were attached to the 5′ ends of the forward and reverse primers respectively. The details of the primers are given in Table 2 and the primer binding sites on the sequence is diagrammatically represented in Figure 4.

3.2.2.4.3. PCR conditions

All amplification reactions were carried out in a Tetrad 2 DNA Engine (M J Research Bio Rad). PCR were set up in 96-well microtiter plates and the reaction volume was 20 µl per well.

Step I: Pre-amplification mixing of the DNA samples of each accession was done with that of Arka Vikas in a 1:1 ratio and 5 ng of template was added to each reaction. Other components include 1X PCR buffer (10 mM Tris, 50 mM KCl, 1.5 mM $MgCl_2$, 0.1% (w/v) gelatin, 0.005% (v/v) Tween-20, 0.005% (v/v) NP-40, pH 8.8), 0.2 mM each

Table 2: Details of the primers designed for nested PCR strategy. (Universal M13 sequences are indicated in bold letters).

Primer Name	Sequence 5′- 3′	Length (bp)	Amplicon length (bp)
NF	CTGATTGTGCTCTTCCTTTACTTG	24	2684
NR	GAAAAGACACAAGCTGAGTAAACC	24	
PNF	**TGTAAAACGACGGCCAGT**TCCACAGTGAGCATTTCGATC TAC	42	991
PNR	**AGGAAACAGCTATGACCAT**TGGGTGCTAAATCAAGAAA GCTAC	43	
ENF	**TGTAAAACGACGGCCAGT**CTCTTCTCAAGCCTTTTCCAT CTC	42	1231
ENR	**AGGAAACAGCTATGACCAT**CTAACACATCTTCTATCCAA AGGC	43	

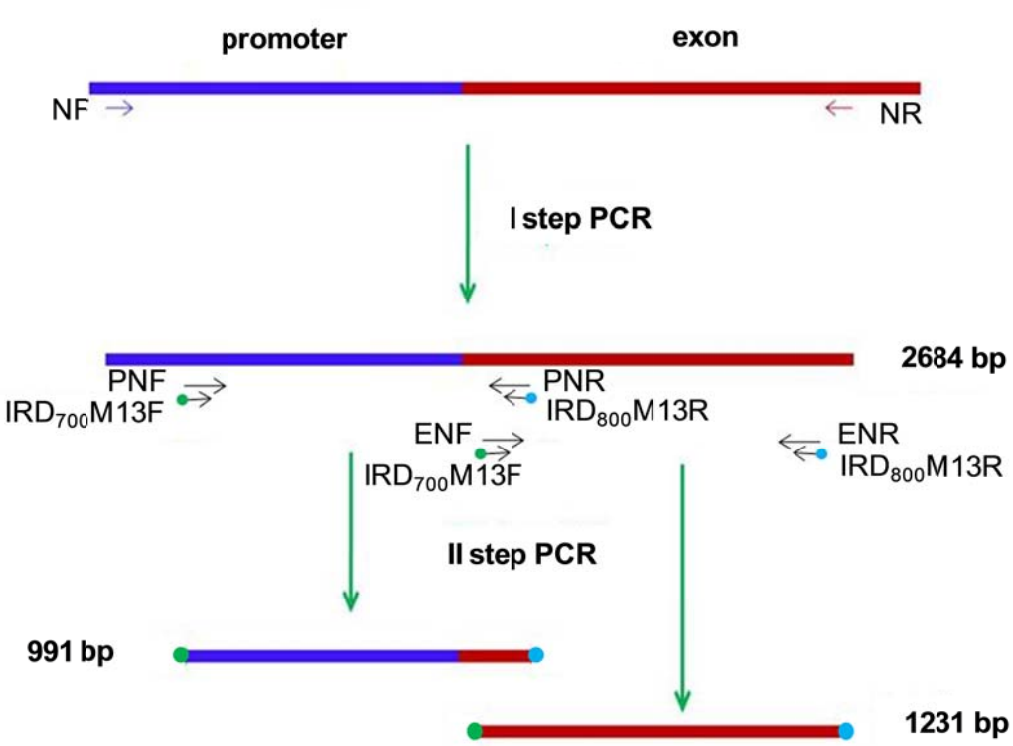


Figure 4: Nested PCR strategy with M13 tailed primers for *CYCB* gene and promoter. First step reaction was carried out with sequence specific unlabeled primers (NF and NR). Second step reactions were carried out with a mixture of sequence specific primers with M13 tail (unlabeled) and universal M13 primers (IRDye labeled). PNF and PNR represents set of forward and reverse primers respectively to amplify promoter. ENF and ENR represents set of forward and reverse primers respectively to amplify exon. IRD_{700}M13F: Universal M13 forward primer with IRD_{700} labeling at 5′ end. IRD_{800}M13R: Universal M13 reverse primer with IRD_{800} labeling at 5′ end.

dNTPs, 0.18 µl Taq polymerase (in-house isolated) and 0.15 pmoles each of forward and reverse primers. The thermocycling conditions for amplification were 94°C-4 min; 10 cycles of 94°C-20 sec, 67°C-45 sec, 72°C-2 min 30 sec; step down: 25 cycles of 94°C-20 sec, 60°C-45 sec, 72°C-2 min 30 sec; 72°C-10 min; held at 4°C.

Step II: 1 µl each of the first step product was used as template in the second step reaction. The other components include 1X PCR buffer, 0.2 mM each dNTPs, 0.10 µl Taq polymerase (in-house isolated) and 0.3 pmoles of primer cocktail in the ratio 2:3:1:4 (Forward unlabelled:M13Forward labeled: Reverse unlabelled:M13Reverse labeled). The thermocycling conditions for amplification were 94°C-4 min; 10 cycles of 94°C-20 sec, 67°C-45 sec, 72°C-1 min 20 sec; step down: 25 cycles of 94°C-20 sec, 60°C-45 sec, 72°C-1 min 20 sec; 72°C-10 min; heteroduplex formation: 99°C-10 min, 80°C-20 sec, 70 cycles of 80°C-7 sec with a decrement of 0.3°C per cycle and held at 4°C.

3.2.2.4.4. Mis-match cleavage

The mismatch cleavage reaction was performed in a total volume of 45 µl containing 20 µl PCR product, 1X CEL I digestion buffer (10 mM HEPES buffer pH 7.0, 10 mM KCl, 10 mM $MgCl_2$, 0.002% (v/v) Triton X-100 and 10 µg/ml BSA) and CELI enzyme (isolated previously from celery as described by Till *et al*. (2006) and Oleykowski *et al*. (1998)) at 1: 300 dilution (1 µl/300 µl CELI digestion buffer). The reaction was incubated at 45°C for 15 min and then stopped by adding 10 µl stop solution (2.5 M NaCl, 75 mM EDTA, pH 8.0 and 0.5 mg/ml blue dextran). Precipitation was done by addition of 125 µl of cold absolute ethanol and a brief incubation in -80°Cfollowed by centrifugation at 4500 rpm in a SH-3000 rotor for 30 min. The pellet was washed with a 70% (v/v) ethanol, dried in a dry bath at 80°C and stored in -80°C till loading.

3.2.2.4.5. Scanning for SNPs

The pellets were suspended in 8 µl formamide loading buffer (37% (v/v) deionized formamide, 1 mM EDTA and 0.02% (w/v) bromophenol blue). The products were denatured by heating to 94°C for 2 min and then placed on ice. About 0.5 µl of the sample was loaded on a 6.5% (w/v) denaturing polyacrylamide gel and was electrophoresed in TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 8.3) at 1500 V, 40 mA and 40 V setting on LI-COR 4300 DNA Analyzer (Li-COR Biosciences). The two TIFF images of IRDye700 and IRDye800 channels were analyzed in Adobe Photoshop software (Adobe Systems Inc.) and the gel was visually assessed for mutations based on the presence of complementary fragments in IRDye700 and IRDye800 channels. The SNPs were confirmed by repeating the protocol using fresh DNA samples from the selected accessions. Sequencing of the amplified fragments from the confirmed accessions was done and the exact nucleotide changes were identified.

3.2.3. Detection of promoter size variations

The promoter region of *CYCB* gene was amplified from various accessions and wild relatives using the forward and reverse primers 5′-caatccagtggattctcgttctggcacct-3′ and 5′-tattgatgcttttgggtacctactttggc-3′ respectively designed to amplify 1007 bp fragment from the reference cultivar. The reaction mixture consisted of 5 ng of DNA template, 1X PCR buffer, 0.2 mM each dNTPs, 0.18 µl Taq polymerase (in-house isolated) and 0.15 pmoles each of forward and reverse primers. The thermocycling conditions for amplification were 94°C-4 min; 35 cycles of 94°C-20 sec, 65°C-45 sec, 72°C-1 min; 72°C-10 min; held at 4°C. The PCR product lengths were compared by electrophoresis on

2% (w/v) agarose gel in 1X TAE buffer at room temperature using 50-70 V current. The fragments which showed size variations were re-amplified and sequenced.

3.2.4. Sequencing and analysis of sequences

All sequencing reactions were carried out by Bioserve Biotechnologies (India) Pvt. Ltd using dideoxynucleotide chain termination method. Sequences were assembled using Chromas software (http://www.technelysium.com.au/chromaslite.html). The sequence variations (substitutions and In-Dels) were detected by aligning the sequences with wild type sequence using 'multialign' software (http://multalin.toulouse.inra.fr/multalin/). The zygosity of the SNPs was determined based on the presence of single (homozygous) or double (heterozygous) peaks in the chromatogram. The amino acid changes corresponding to each nucleotide changes and their effect on gene functions were obtained using PARSENP (Project Aligned Related Sequences and Evaluate SNPs) analysis (http://www.proweb.org/parsesnp/). Phylogram based on the amino acid sequences was constructed using MEGA6 (Tamura *et al*., 2013). The evolutionary history was inferred using the UPGMA method (Sneath and Sokal, 1973). The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated. Phylograms based on the nucleotide sequences were constructed using ClustalW2 software (http://www.ebi.ac.uk/Tools/msa/clustalw2/). The promoter sequences were analysed using 'Softberry N site'(http://linux1.softberry.com/berry.phtml?topic=nsitep&group=programs&subgroup=

promoter) and PLACE (PLAnt Cis-acting regulatory DNA Elements) software (http://www.dna.affrc.go.jp/PLACE/) to identify the regulatory elements.

3.2.5. Transcript quantification of selected accessions

3.2.5.1. Total RNA extraction and quantification

The tomato accessions used for Real-Time PCR were *S. lycopersicum* cv. Arka Vikas (WTAV), Pearson (*LA0012*), B^{og} mutant (*LA0348*), B^{og} mutant (*LA4026*), *Beta* mutant (*LA2374*), *S. chilense* (*LA1969*) and *S. habrochaites* (*LA1362*). Fruit tissues were collected at 30 DPA stage from all these accessions. In addition, for *S. hirsutum,* fruit tissue at 60 DPA stage was also collected. Fruit pericarp tissues were frozen in liquid N_2 and homogenized using a pre-cooled pestle and mortar to fine powder using a mortar and pestle. About 100 mg of homogenized powder was transferred to a 1.5ml microcentrifuge tube and 1 ml of TRI Reagent (Sigma, MO, USA) was added. The samples were thoroughly mixed and incubated for 5 min at room temperature. To this 0.2 ml chloroform was added and the tubes were subjected to vigorous vortexing. Thereafter tubes were incubated at room temperature for 3 min and centrifuged at 12,000 rpm for 15 min at 4°C. After phase separation, the upper aqueous phase was carefully transferred into another micro centrifuge tube and 0.5 ml of isopropanol was added to it. The tubes were incubated at room temperature for 10 min and centrifuged at 12,000 rpm for 10 min at 4°C. The supernatant was discarded slowly and the pellets were washed with 75% (v/v) ice-cold ethanol followed by centrifuging at 10,000 rpm for 10 min at 4°C. Finally, the pellet was air–dried and dissolved in 50 µl RNAse free water. The samples were stored at -80°C until further use.

Concentration of RNA was determined by measuring UV absorbance (A_{260} and A_{280} nm) in NanodropND-1000 spectrophotometer. The integrity of RNA was checked on 1.2% (w/v) formaldehyde agarose gel containing 1X formaldehyde agarose gel buffer (20 mM 3-[N-morpholino]propanesulfonic acid (MOPS), 5 mM sodium acetate, 1 mM EDTA, pH 7.0) 1.8 ml 37% (12.3 M) formaldehyde/100 ml and 0.1 µg/ml ethidium bromide. Prior to running, the gel was equilibrated in 1X Formaldehyde Agarose gel running buffer (1X formaldehyde agarose gel buffer, 20 ml 37% (12.3 M) formaldehyde/1000 ml) for 30 min. RNA samples for loading were prepared by adding 1 volume of 5X loading buffer (16 µl saturated aqueous bromophenol blue solution, 80 µl 500 mM EDTA, pH 8.0, 720 µl 37% (12.3 M) formaldehyde, 2 ml 100% glycerol, 3084 µl formamide, 4 ml 10X formaldehyde agarose gel buffer, RNase-free water to 10 ml) per 4 volumes of RNA sample and mixing. The samples were incubated for 3–5 min at 65°C, chilled on ice, and loaded onto the equilibrated Formaldehyde Agarose gel. Electrophoresis was carried out at 5–7 V/cm in 1X formaldehyde agarose gel running buffer.

3.2.5.2. cDNA synthesis

To remove the contaminating DNA, all the RNA samples were treated with DNase, using RQ1 RNase-Free DNase kit (Promega, Madison, USA) according to manufactures protocol. Reverse transcription was performed with 2 µg of total RNA in a total volume of 20 µL using SuperScript™ III first strand synthesis system for RT-PCR (Invitrogen, Rockville, USA) following manufactures protocol.

3.2.5.3. Real-Time PCR

Real-Time PCR was performed using cDNAs corresponding to 5 ng of total RNA in 10 µl reaction volumes with 0.2 pmoles each of forward and reverse primers using the SYBR Green PCR Master Mix (Takara, Otsu, Japan) on a 7300 Fast Real Time PCR system (ABI). The primer sequences used for Real-Time PCR were designed using primer3 software (Table 3). Actin was used as the internal control. The relative expression was calculated using the formula $2^{(-\Delta\Delta CT)}$, where ΔCt=(Ct value of target gene) - (Ct value of actin) and ΔΔCT=ΔCt of accession-ΔCt of reference.

Table 3: List of primers used for Real-Time PCR

Primer name	**Sequence 5´- 3´**
CycB F	TCTTCTCAAGCCTTTTCCATC
CycB R	TGGTGGGACTTAGAAAAGAAGG
*FLuc*RTF	GAGGCGAACTGTGTGTGAGA
*FLuc*RTR	GTGTTCGTCTTCGTCCCAGT
Actin F	GAAATAGCATAAGATGGC
Actin R	ATACCCACCATCACACCA

3.2.6. Transient expression

3.2.6.1. Preparation of constructs

The double luciferase expression vector, pGreenII 0800LUC (Figure 5) was kindly provided by Dr. Roger Hellens, Plant and food research, New Zealand. The plant materials selected for the study were *S. lycopersicum* cv.Arka Vikas, *Beta* mutant (LA2374) and B^{og} mutant (LA0348). The primer sequences which were designed to amplify 886 bp sequence (WT) upstream to the start codon of *CYCB* gene are given in Table 4. Since the GC content of the reverse primer was very low, modified PCR conditions were used (Figure 6). The reaction mixture consisted of 5 ng of template, 1X PCR buffer, 0.2 mM each dNTPs, 0.18 µl Taq polymerase (in-house isolated) and 0.15 pmoles each of forward and reverse primers. The thermocycling conditions for amplification were 94°C-4 min; 5 cycles of 94°C-20 sec, 55°C-45 sec, 72°C-1 min (carried out with reverse primer only); 5 cycles of 94°C-20 sec, 55°C-45 sec, 72°C-1 min (carried out after the addition of forward primer); 25 cycles of 94°C-20 sec, 60°C-45 sec, 72°C-1 min; 72°C-10 min; held at 4°C.

The PCR product sizes were compared by electrophoresis on 2% (w/v) agarose gel in 1X TAE buffer at room temperature using 50-70 V current. The promoter fragments from all the accessions were gel eluted using QIAquick Gel Extraction Kit (Qiagen, Germany). They were then ligated to the RcoRV (blunt end) site of the cloning vector pMosBlue and transformed to *E.coli* (DH5-α) ultra competent cells by heat-shock method. The colonies with required orientation of the promoter were selected by specific PCR amplification check using combinations of vector specific (T7 and M13; universal) and gene specific primers (Figure 7). They were then confirmed by sequencing.

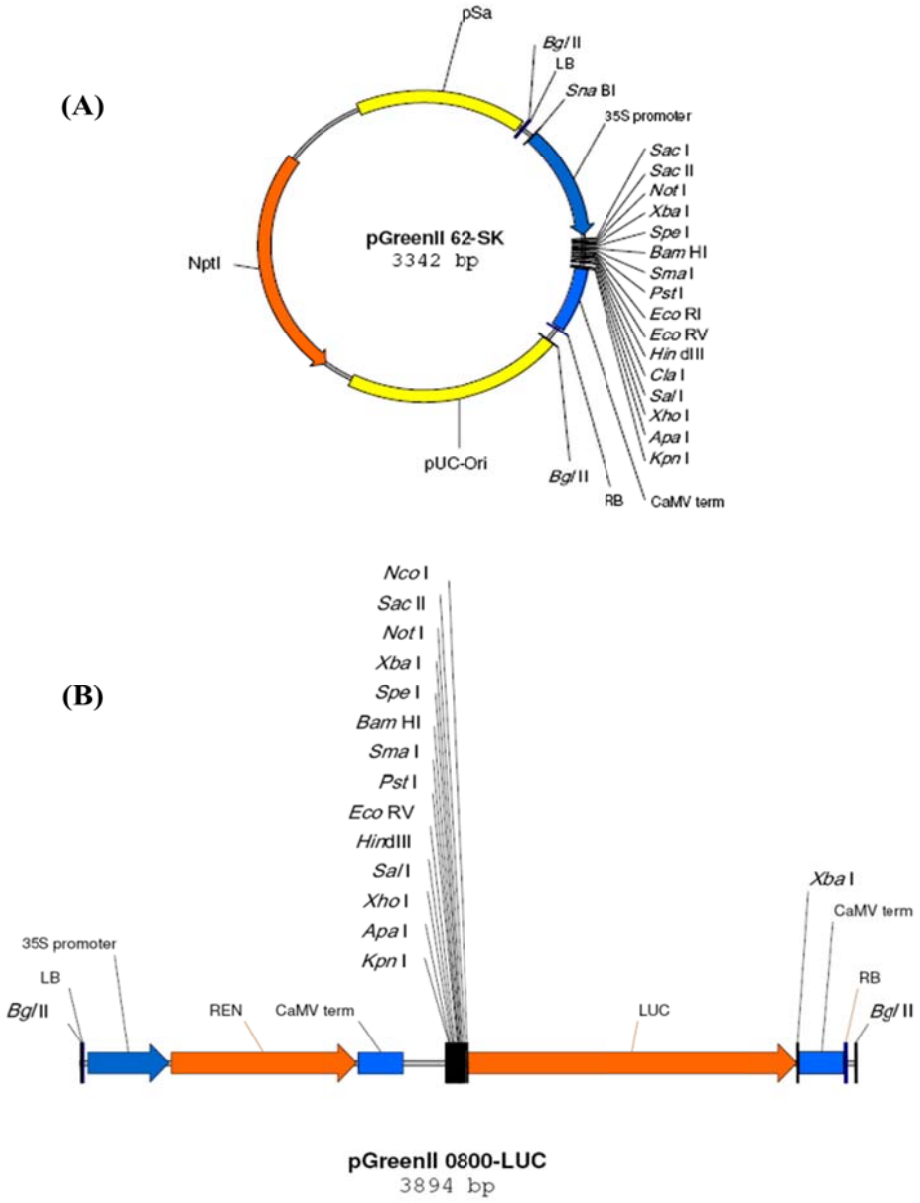


Figure 5: Map of the transient expression vector pGreenII 62-SK (A). This vector was used as a backbone for making the expression vector pGreenII 0800-LUC by Hellens *et al*. (2005). The size of the vector is 3342 bp.
T-DNA region of the transient expression vector pGreenII 0800-LUC (B). REN: *Renilla* luciferase (RLUC) gene under cauliflower mosaic virus (CaMV) 35S promoter (constitutively expressing internal control), LUC: Firefly luciferase (FLUC) gene (will be expressed depending on the promoter inserted in the Multiple Cloning Site (MCS) upstream to it). The size of the T-DNA region is 3894 bp (Hellens *et al*., 2005). The total size of the expression vector pGreenII 0800-LUC is around 7.2 Kb.

Table 4: Primers designed to amplify *CYCB* promoter region for transient expression

Primer Name	Sequence 5´- 3´	Tm °C	GC%	Amplicon length
TFP	ACACCAGGGTTGTCAAAAATGTCTC	61.3	44	886 bp
TRP	TATAGAGAATGTATAAGATTGATAATGGT	56.8	24.1	

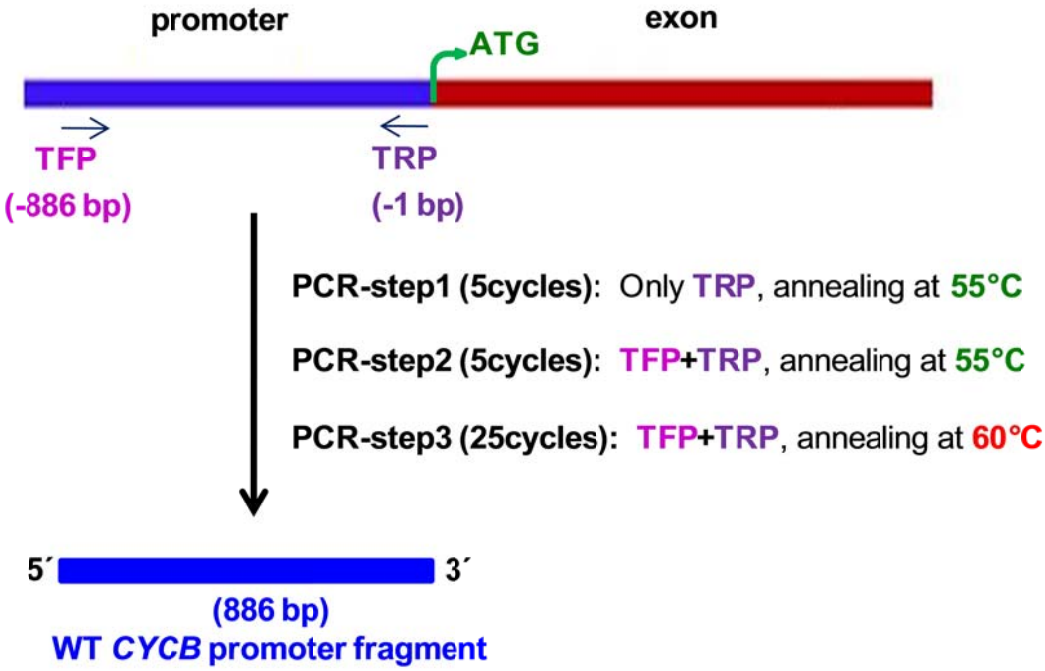


Figure 6: Diagrammatic representation of PCR amplification strategy of *CYCB* promoter fragment for transient expression. Binding sites of forward (TFP) and reverse (TRP) primers are –886 bp and –1 bp respectively on a 'WT' (Arka vikas) template. PCR was carried out in three steps. The conditions used for different steps are indicated. The same strategy was used to amplify respective fragments from '*Beta* type' and '*Chilense* type 'templates, where, the binding sites of TFP are –933 bp and –1192 bp respectively.

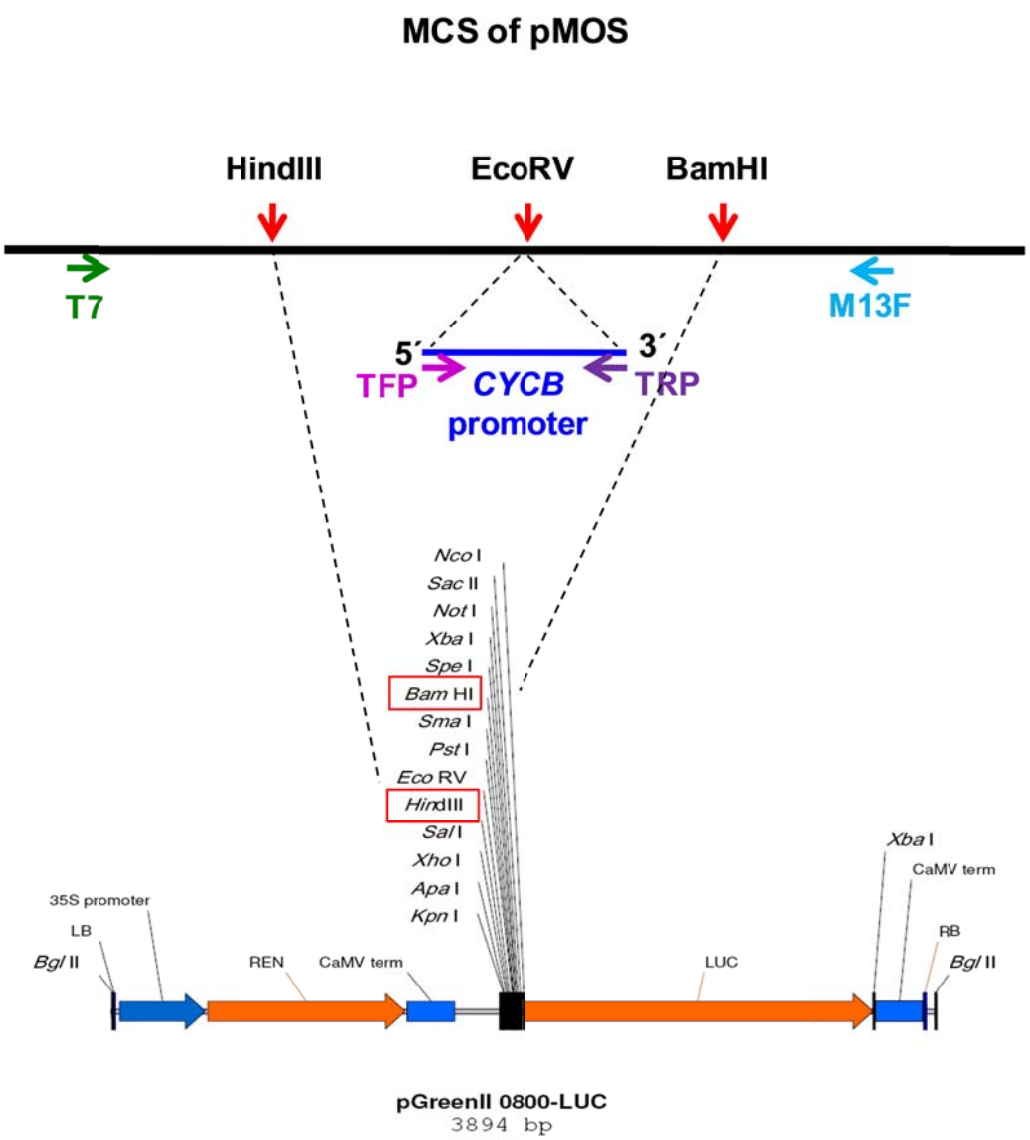


Figure 7: Diagrammatic representation of construct preparation with *CYCB* promoter fragment using pGreenII 0800-LUC for transient expression. The promoter fragments were inserted into the EcoRV blunt end site of the cloning vector pMosBlue. The MCS of pMosBlue with various restriction sites and binding sites of universal T7 and M13F primers are diagrammatically shown. Required orientation of the *CYCB* promoter in pMosBlue is also shown and were selected based on specific PCR amplification using universal T7 and insert specific TRP primer pair and universal M13F and insert specific TFP primer pair. Absence of PCR amplification with the opposite combination of primers confirmed the selection. The selected promoter fragments in the required orientation were cloned into the double luciferase vector pGreenII 0800-LUC upstream to the *LUC* (Firefly luciferase) gene using Bam HI and Hind III sites.

The promoter fragments were then derived by double digestion using BamHI and Hind III (Fermentas, Thermo Fisher Scientific, India) and gel elution following manufactures protocols. These fragments were then ligated into the Bam HI and Hind III sites of pGreenII 0800LUC vector (Figure 7) using T4 DNA ligase (Fermentas, Thermo Fisher Scientific, India) according to manufactures instructions and transformed to DH5-α cells. The positive colonies were selected by colony PCR using the primers specific to the backbone. The primers used for this were FP: 5′-TCGTTCGTTGAGCGAGTTCT-3′ and RP: 5′-CGTAAGTGATGTCCACCTCGAT-3′and was confirmed by double digestion using Bam HI and Hind III. Plasmid DNA samples for each construct were extracted from the selected DH5-α cells for further use. Preparation of ultra competent DH5-α cells and transformation of the plasmid constructs were done by the protocol described by Inoue *et al* (1990).

3.2.6.1.1. Plasmid DNA extraction from DH5-α cells

Plasmid DNA was extracted from DH5-α cells using alkali lysis method. 2 ml overnight grown culture was taken in a 2.0 ml microcentrifuge tube. The cells were pelleted by centrifugation at 13,000 rpm for 1 min at 4°C. The supernatant was discarded and 100 µl of ice-cold solution I (50 mM Glucose; 25 mM Tris-HCl, pH 8.0; 10 mM EDTA, pH 8.0) and 8 µl of RNAse (10 mg/ml) was added to the pellet. The cells were resuspended and kept on ice for 5 min. To this 100 µl of freshly prepared solution II (0.2 N NaOH; 1%(w/v) SDS) was added and mixed by inverting the tubes. Within 3 min of addition of solution II, 150 µl ice-cold solution III (3 M potassium acetate, pH 5.5 adjusted with glacial acetic acid) was added, mixed by inversion again and kept on ice for 5 min. The tubes were then centrifuged at 13,000 rpm for 2 min at 4°C to remove the

debris. The supernatant was transferred to a fresh microcentrifuge tube and repeated the centrifugation step to remove any impurities remaining. To the fresh supernatant 0.7 volume of chilled isopropanol was added, mixed gently and centrifuged at 13,000 rpm for 15 min at 4°C to pellet the plasmid DNA. The supernatant was removed and the pellets were washed twice with 500 µl 70% (v/v) ethanol by centrifugation at 13,000 rpm for 5 min at 4°C. The pellets were air dried and dissolved in 50 µl of sterile MQ water and stored at -20°C until further use.

3.2.6.2. Transformation of fruit tissues by particle bombardment

3.2.6.2.1. Preparation of fruit tissues

Fruit tissues of *S. lycopersicon* cv. Arka Vikas at 30 DPA stage was used for transient expressions experiments. Before bombardment, osmotic treatment was given to the fruit tissues. The fruits were washed and surface sterilized with 75% (v/v) ethanol. Thin slices of mesocarp fruit tissues (3-5 mm X 3-5 mm) were arranged over a sterile filter paper (3 cm diameter) at the center of petriplates (9 cm diameter) with an osmotic medium (0.4 M mannitol, 0.4 M sorbitol, 0.8% agar, 0.1 M sucrose). The plates were sealed with parafilm and incubated for 4 hrs at 26°C.

3.2.6.2.2. Preparation of tungsten particles

In a 1.5 ml microcentrifuge tube, 50 mg (1.1 µ size) Tungsten particles (Bio-rad Laboratories) were suspended in 1 ml of 95% (v/v) ethanol. Particle suspension was sonicated for 30 sec using a sterile tip to destroy particle aggregates and the tube was centrifuged for 1 min and the supernatant was removed. Then the particles were re-suspended in 1 ml of ethanol, centrifuged and supernatant was removed. The particles

were then washed three times in 1.0 ml sterile distilled water (SDW) by centrifugation. Finally, the particles were suspended in 1.0 ml of SDW, made into 50 µl aliquots and stored at -20°C.

3.2.6.2.3. Preparation of tungsten-DNA mixture

1 µg of the plasmid DNA sample of each construct was added to different microcentrifuge tubes with 50 µl aliquots of tungsten particle. To this 50 µl of 2.5 M $CaCl_2$ and 20 µl of 100 mM spermidine were added. The tubes were vortexed briefly and placed on ice for 5 min. When the DNA coated tungsten particles settled down, around 50 µl of the clear supernatant was discarded and the rest of the aliquot was used for microprojectile bombardment.

3.2.6.2.4. Particle bombardment

The inside chamber of the PIG (Finer *et al*., 1992) was sprayed with 70% (v/v) ethanol. The aperture inlet valve was opened at 360°. Following this, the helium cylinder valve was opened completely and the high-pressure gas regulator valve was opened to the required pressure of 12 kg/cm^2. The tungsten-plasmid DNA mixture was resuspended well and 25 µl of the mixture was loaded into the syringe filter unit of the bombardment device. The filter holder was screwed finger tight into place. The petri-plates containing the fruit tissues were placed one after the other on the adjustable shelf at 10 cm distance from the tip of the filter unit. The door was closed and the handles were latched. Immediately, the vacuum valve was opened and the pump was switched on till the vacuum reached 600 mm Hg. The fruit tissues were bombarded under a helium gas pressure of 12 kg/cm^2 under partial vacuum by flicking and releasing the timer relay

switch during which the tungsten was propelled into the tissue and displaced by force. The vacuum was then released and the displaced explants pieces were collected. The above steps were repeated for bombarding samples in all the petri-plates. Each construct was transformed in three replicates. The bombarded tissues were then left on the osmotic medium (petriplates sealed with parafilm) and incubated at 26ºC and 45% relative humidity and with 16 h light / 8 h dark photoperiod for 48 hrs for post-bombardment treatment.

3.2.6.3. RNA extraction, cDNA preparation and Real-time PCR

RNA extraction was carried out with 100 mg of the transformed fruit tissues in each case using the protocol outlined in section 3.2.5.1. The protocol for cDNA preparation and Real-time PCR were same as outlined in section 3.2.5.2 and 3.2.5.3 respectively. The primer sequences used for Real-Time PCR were designed using primer3software (Table 3). Relative *FLUC* expression was calculated with Actin as the internal control.

3.2.6.4. Dual Luciferase assay

Quantity of Firefly luciferase and *Renilla* luciferase were assayed using the dual luciferase assay reagents (Promega, Madison,USA) following manufactures instructions with few modifications for plant tissue. After transformation and a transient incubation of 2 days, ca.100 mg of the fruit tissue discs were harvested and homogenized in liquid N_2 to make a fine powder using a pestle and mortar. To this 1 ml of 1X Passive Lysis Buffer (PLB) was added immediately, mixed well and the contents were transferred to a 1.5 ml micro centrifuge tube. The tubes were incubated at 22°C for 15 min with gentle shaking

in between for the complete lysis of the cells. For removing the debris, the tubes were centrifuged at 12,000 rpm for 30 sec at 4°C and the supernatant was transferred to a new micro centrifuge tube. 20µl of this crude extract was assayed in 50 µl of Luciferase Assay Reagent (LARII), and the chemiluminescence was measured (FLUC). 50 µl of Stop and Glow™ buffer was then added and a second chemiluminescence measurement was made (RLUC). Chemiluminescence measurements were made using GloMax® 96 Luminometer (Promega, Madison,USA) by manual loading method with 10 sec delay and 5 sec reading in both cases. Ratio of FLUC to RLUC was taken as a measure of relative promoter strength. Background controls were run with the empty vector pGreenII 0800LUC (without any promoter for *FLUC*).

Chapter 4

Results

4.1. Diversity in chromoplast specific lycopene cyclase gene (*CYCB*) and its homologues between different plant genera

To study the diversity of *CYCB* gene and its homologues in different plant genera, the amino acid sequences of various homologues were obtained from NCBI database through homology search using tomato *CYCB* gene sequence. The various homologues identified include enzymes like LCYB (Lycopene beta cyclase, chloroplastic), LCYE (Lycopene epsilon cyclase), and CCS (Capsanthin capsorubin synthase). Phylogenetic analysis of CYCB homologues showed three major groups. One group consisted mostly of LCYB; another group consisted mostly of LCYE while the third group consisted of enzymes like CYCB and CCS (Figure 8). It shows that the amino acid sequences of these enzymes are significantly conserved among different plant genera. LCYB from different plant species shows more homology between themselves than to CYCB of the same species. Similar observations could be made for other lycopene cyclase genes like *CYCB* and *LCYE* too. The CYCB lineage appears to have diverged from the LCYB lineage. In some plants CYCB lineage genes appears to have evolved novel functions like CCS in pepper (Kim *et al*., 2014). Phylogenetic analysis shows that the gene duplication event that leads to the evolution of *LCYE* might have occurred earlier than the duplication event that leads to the divergence of LCYB and *CYCB*/*CCS* (Figure 8).

CYCB of tomato shows highest similarity (86%) with CCS from pepper. The highest amino acid similarity was found to be between LCYB1 and LCYB2 of tomato (87%). The similarities of LCYE with other lycopene cyclases were equal to or less than 40%. Even though CYCB and LCYB have the same function in different tissues, they have highly diverged amino acid sequences (Table 5). Major difference between LCYB

and CYCB, two enzymes catalyzing the same reaction, lies in their cellular localization as well as spatial and temporal expression pattern. A comparison of transit peptide between LCYB and CYCB across different species using *'TargetP 1.1'* software (Emanuelsson *et al*., 2007) indicates that *CYCB* genes have a stronger chloroplast localization signal (Table 6). It could also be that chromoplast targeting requires a different transit peptide sequence.

Alignment of amino acid sequences of CYCB homologues indicated maximum sequence divergence in the N-terminal region of the protein sequence partly due to high divergence in the signal peptide region (Figure 9). The transit peptide region shows high divergence between the three different groups and the transit peptide of CYCB group shows higher plastid targeting scores, probably indicating more effective targeting of this protein to the plastids (Table 6).

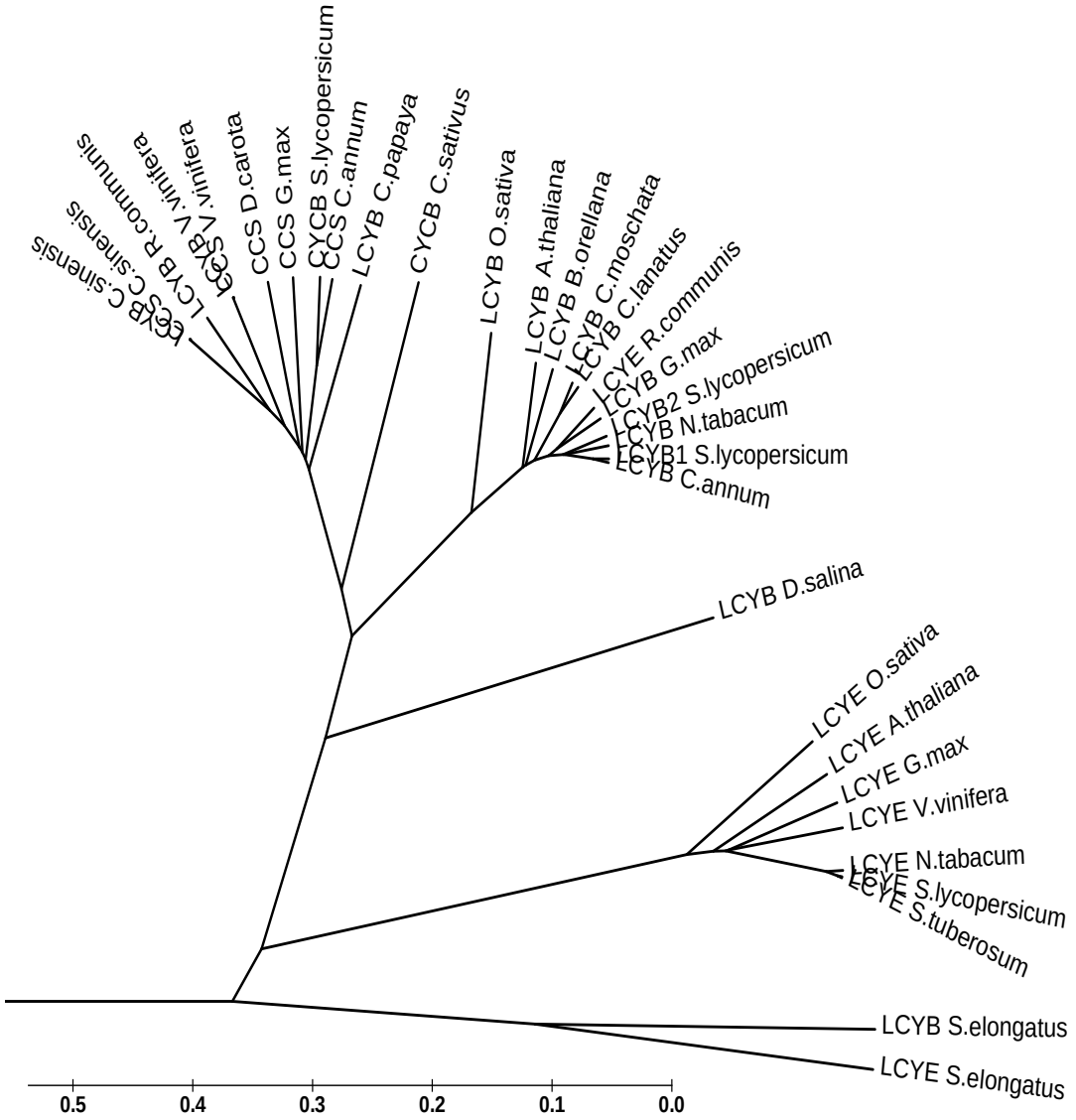


Figure 8: Phylogenic tree showing the similarities between amino acid sequences of CYCB and its homologues. The evolutionary history was inferred using the UPGMA method (Sneath and Sokal, 1973). The optimal tree with the sum of branch length = 4.42204667 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. The analysis involved 32 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 383 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura *et al*., 2013).

Table 5: Amino acid sequence identities (%) between CYCB homologues (NCBI blastp suite-2sequences)

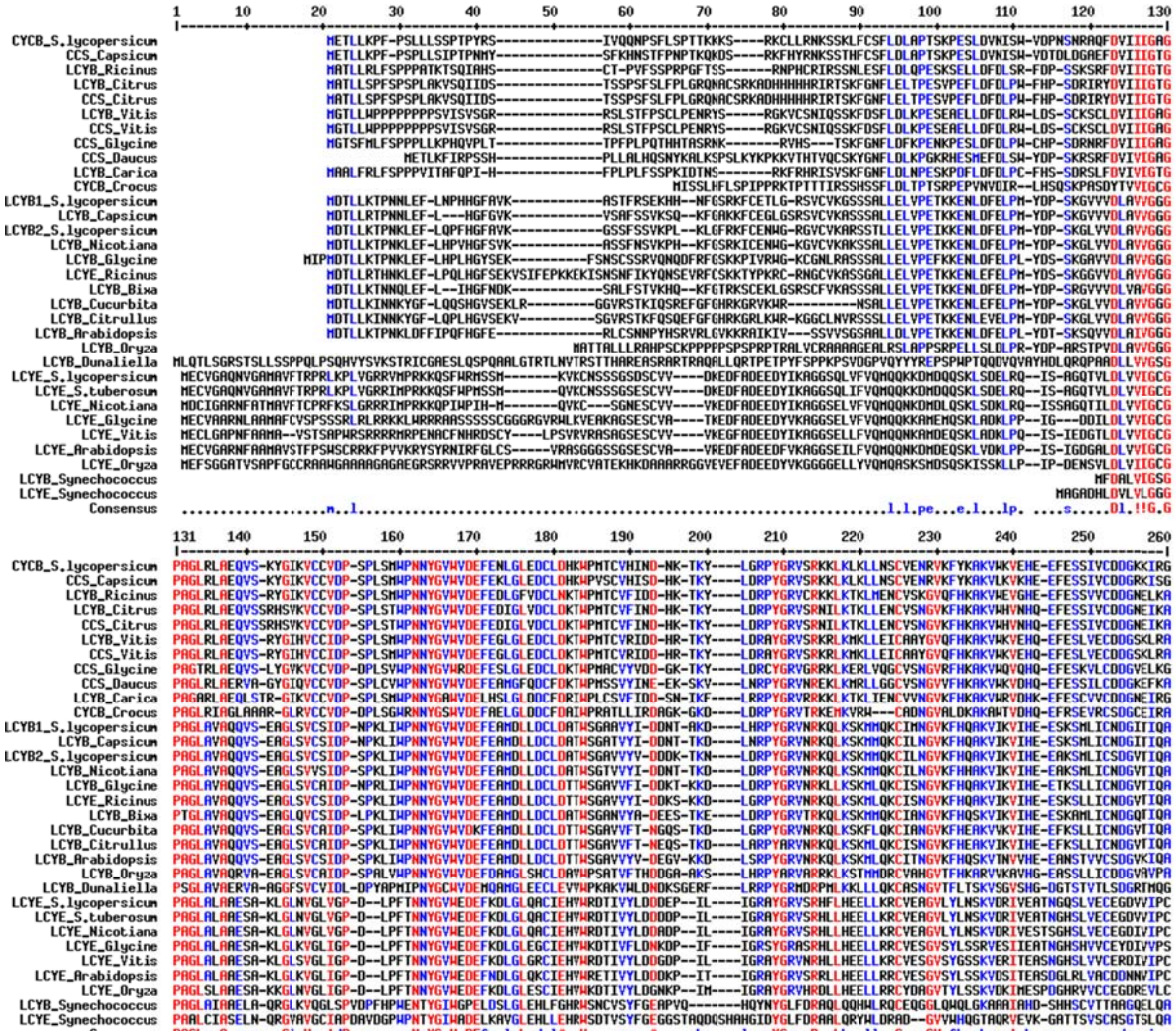
	CYCB_Tomato(AK327886.1)	LCYB1_Tomato (AK319553.1)	LCYB2_Tomato (AK323472.1)	LCYE_Tomato (NP_001234337)	CCS_Pepper (GU122939.1)
CYCB_Tomato(AK 327886.1)	100				
LCYB1_Tomato (AK319553.1)	55	100			
LCYB2_Tomato (AK323472.1)	56	87	100		
LCYE_Tomato (NP_001234337)	40	37	37	100	
CCS_Pepper (GU122939.1)	86	52	58	39	100

Table 6: Subcellular localisation scores of CYCB homologues predicted using *'TargetP 1.1'* (www.cbs.dtu.dk/services/TargetP/)

Name*	Amino acid Length	cTP	mTP	SP	other	Locus	RC	Tp length
CYCB_*Solanum lycopersicum*(AK327886.1)	498	0.895	0.083	0.009	0.048	C	1	42
CYCB_*Crocus sativus*(GQ202141.1)	468	0.851	0.262	0.01	0.049	C	3	22
CCS_*Capsicum annum*(GU122939.1)	498	0.81	0.052	0.03	0.268	C	3	39
CCS_*Daucus carota*(DQ192191.1)	492	0.089	0.524	0.034	0.475	M	5	58
CCS_*Vitis vinifera*(XM_002273826.1)	497	0.937	0.221	0.021	0.032	C	2	42
CCS_*Glycine max*(XM_003541917.1)	493	0.764	0.145	0.041	0.114	C	2	41
CCS_*Citrus sinensis*(AF169241.1)	503	0.933	0.03	0.034	0.091	C	1	12
LCYB1_*Solanum lycopersicum*(AK319553.1)	500	0.26	0.144	0.044	0.553	-	4	-
LCYB2_*Solanum lycopersicum*(AK323472.1)	500	0.205	0.133	0.033	0.609	-	3	-
LCYB_*Capsicum annum*(ADH04277.1)	498	0.188	0.17	0.027	0.542	-	4	-
LCYB_*Nicotiana tabacum*(CAA57386.1)	500	0.188	0.14	0.096	0.638	-	3	-
LCYB_*Bixa orellana*(CAD70565.1)	499	0.259	0.071	0.048	0.534	-	4	-
LCYB_*Ricinus communis*(EEF30885.1)	495	0.884	0.347	0.004	0.03	C	3	45
LCYB_*Citrus sinensis*(FJ516403.1)	503	0.933	0.03	0.034	0.091	C	1	12
LCYB_*Vitis vinifera*(JQ319643.1)	497	0.937	0.221	0.021	0.032	C	2	42
LCYB_*Carica papaya*(FJ839872.1)	494	0.612	0.109	0.064	0.089	C	3	44
LCYB_*Oryza sativa*(BAD16478.1)	489	0.878	0.212	0.007	0.056	C	2	33
LCYB_*Arabidopsis thaliana*(NP_187634.1)	501	0.091	0.05	0.075	0.816	-	2	-
LCYB_*Synechococcus elongates*(YP_401079.1)	411	0.058	0.098	0.676	0.473	S	4	17
LCYB_*Cucurbita moschata*(AEN94903.1)	497	0.139	0.308	0.021	0.577	-	4	-
LCYB_*Citrullus lanatus*(ABM90918.1)	504	0.177	0.129	0.023	0.73	-	3	-
LCYB_*Glycine max*(XP_003554131.1)	507	0.385	0.059	0.024	0.713	-	4	-
LCYB_*Dunaliella salina*(ACA34344.1)	584	0.861	0.214	0.005	0.064	C	2	68
LCYE_*Solanum lycopersicum*(NP_001234337)	527	0.497	0.305	0.029	0.252	C	5	45
LCYE_*Solanum tuberosum*(DAA33890.1)	527	0.494	0.358	0.029	0.225	C	5	45
LCYE_*Arabidopsis thaliana*(NP_200513.1)	524	0.751	0.442	0.016	0.045	C	4	45
LCYE_*Synechococcus elongates*(ZP_01470358.1)	414	0.039	0.081	0.66	0.578	S	5	22
LCYE_*Ricinus communis*(EEF48090.1)	514	0.119	0.138	0.038	0.761	-	2	-
LCYE_*Glycine max*(XP_003546468.1)	531	0.986	0.341	0.014	0.003	C	2	54
LCYE_*Nicotiana tabacum*(ADZ48238.1)	524	0.312	0.412	0.046	0.223	M	5	9
LCYE_*Vitis vinifera*(AFP28798.1)	530	0.961	0.162	0.023	0.02	C	2	48
LCYE_*Oryza sativa*(BAC05562.1)	540	0.587	0.194	0.067	0.08	C	4	52

*Gene bank/NCBI reference sequence accession numbers are given in bracket.

C: Chloroplast, **M**: Mitochondria, **S**:Secretory pathway, **cTP**: plastid transit peptide, **mTP**: mitochondrial targeting peptide, **SP**: secretory pathway signal peptide, _: Any other location, **RC**: Reliability class (1 indicates the strongest prediction), **TP**: target peptide



…cont'd in the next page

…cont'd from previous page

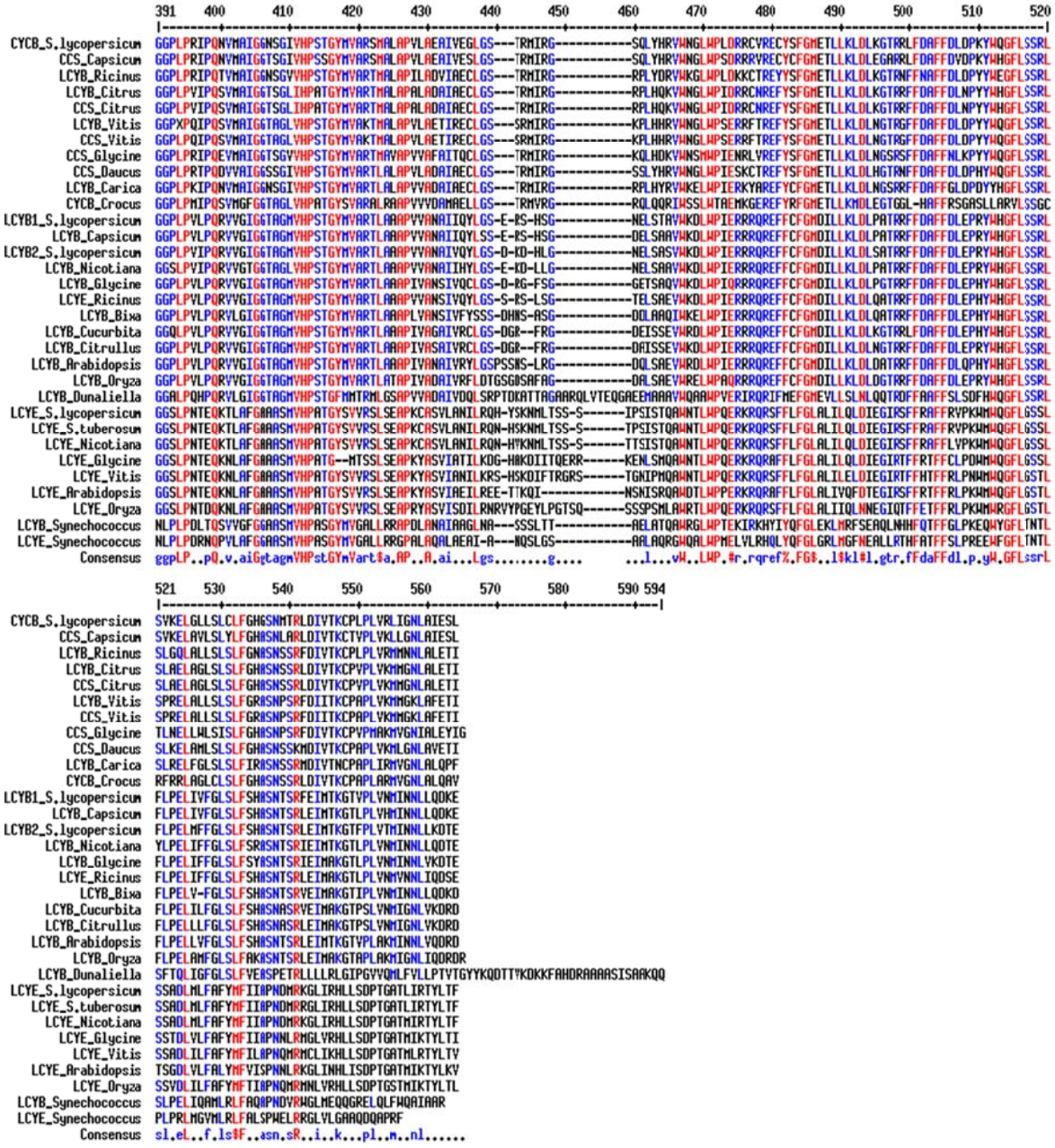


Figure 9: Alignment of amino acid sequences of CYCB and its homologues. Multalin software (http://multalin.toulouse.inra.fr/multalin/multalin.html) (Corpet, 1988) was used for making the alignment. Red color indicates high consensus (>90%) and blue color indicates low consensus (<50%) amino acid residues. The N-terminal regions of all the enzymes are highly diverged.

4.2. Diversity in tomato *CYCB* gene (coding region) between different accessions and wild relatives of tomato

CYCB gene is located on chromosome 6 of tomato genome. It is an intronless gene with a total size of 1.6 Kb (including the UTR region), coding for a protein of 498 amino acids. The upstream 900 bp region of the gene is considered as the promoter region (http://solgenomics.net/). The genetic variations in *CYCB* gene and its promoter region were studied in a collection of wild relatives of tomato and among tomato accessions from various resources. An Indian cultivar of *Solanum lycopersicum*, Arka Vikas (Sel 22), was used as a reference variety.The mutation in *CYCB* can lead to either enhanced lycopene in loss of function situation or more *β*-carotene in gain of the function situation in chromoplast. Depending on the type and nature of variations in *CYCB*, alterations in the major carotenoid levels is expected.

4.2.1. Genetic variations in the wild relatives of tomato

A 1.2 Kb region of the coding sequence of *CYCB* gene was amplified from *S. chilense, S. pennellii, S. parviflorum, S. habrochaites, S. cheesmanii, S. pimpinellifolium and S. lycopersicum* (cv. Arka Vikas). The amplicon sizes were found to be the same for all species. All the PCR products were sequenced to find out the SNPs. The sequence of each wild relative was aligned with that of *S. lycopersicum*. The number of SNPs detected in *S. chilense, S. pennellii, S. parviflorum, S. habrochaites, S. cheesmanii* and *S. pimpinellifolium* compared to *S. lycopersicum* are 13, 17, 13, 14, 2 and 2 respectively. The details of the SNPs are presented in Table 7. The amino acid changes corresponding to each nucleotide changes and their effect on gene functions were obtained by sequence analysis using PARSENP (http://www.proweb.org/parsesnp/). Lists of restriction enzyme

polymorphisms caused by the changes are also shown (Table 8). For each amino acid change, the change in their properties like side chain polarity, side chain charge and hydropathy index were studied (Table 9).

4.2.2. Genetic variations in the collection of accessions

Since SNPs contribute a major part of genetic variations in nature, tomato accessions were screened for SNPs using a modified strategy of TILLING (Targeting Induced Local Lesions IN Genomes) (Till *et al*., 2004) known as EcoTILLING (Comai *et al*., 2004). Since CELI digest only a proportion of the heteroduplexes at a single position, multiple SNPs in an amplicon can also be detected. Figure 10 shows a schematic representation of multiple fragment formation and a LI-COR gel image showing multiple fragments of a CELI digested heteroduplex sample indicating the presence of multiple SNPs. Moreover, a high throughput screening is also possible where all steps of EcoTILLING screening such as PCR, heteroduplex formation, CELI digestion and separation of fragments can be done with 96 samples in a 96-well PCR plate and digested samples can be run on a single LI-COR gel. For examining natural genetic polymorphism, EcoTILLING was found to be an excellent strategy for screening of large number of accessions for multiple SNPs.

Screening of the *CYCB* exon from 543 tomato accessions revealed the presence of several SNPs. Figure 11 shows representative LI-COR gel image showing detection of SNPs in *CYCB* exon. The presence of SNPs in a given accession was also confirmed by repeating the experiment with unlabelled primers and separation of CELI digested PCR product fragments on agarose gel. Final confirmation for the presence of an SNP was done by sequencing of the amplicons which revealed the exact type and location

Table 7: List of SNPs in *CYCB* gene in various haplotypes and wild relatives

Sl. No.	Nucleotide change	Amino acid change*	*S.pimpinellifolium* (LA1589)	*S.cheesmanii* (LA0483)	*S.habrochaites* (LA1362)	*S.parviflorum* (LA2133)	*S.pennellii* (LA0716)	*S.chilense* (LA1969)	HT.2	HT.3	HT.4	HT.5	HT.6	HT.7	HT.8
1	T55C	Y19H					√	√					√		
2	G59A	R20K			√			√						√	
3	G60A	R20=			√									√	
4	G67T	V23F			√	√	√	√				√	√	√	√
5	A77T	N26I													
6	T81C	P27=			√		√	√				√	√	√	
7	C98T	T33I										√			
8	T125C	L42P				√									
9	G131A	R44K													√
10	A232G	N78D				√	√	√				√	√		
11	A233G	N78S					√								√
12	C249T	D83=				√									
13	C261T	I87=				√		√				√	√		√
14	A264G	G88=													√
15	C286T	L96=					√								
16	A317C	K106T													√
17	C345T	L115=			√							√		√	
18	G406A	D136N							√				√		
19	A407G	D136G									√				
20	A443G	H148R									√				
21	A459G	K153=			√	√	√					√	√	√	√
22	T462C	T154=				√									
23	G465A	K155=													√
24	A463ATA	Frameshift											√		
25	G476A	R159K													√
26	A493T	S165C													√

Sl. No.	Nucleotide change	Amino acid change*	*S.pimpinellifolium* (LA1589)	*S.cheesmanii* (LA0483)	*S.habrochaites* (LA1362)	*S.parviflorum* (LA2133)	*S.pennellii* (LA0716)	*S.chilense* (LA1969)	HT.2	HT.3	HT.4	HT.5	HT.6	HT.7	HT.8
27	G497A	R166K													√
28	G510A	K170=										√			√
29	T525C	N175=			√		√							√	
30	A556G	K186E		√				√					√		
31	G570A	W190*													√
32	A600G	S200=													√
33	A614G	D205G									√				
34	A623G	K208R									√				
35	A683G	D228G									√				
36	G686A	R229K			√							√	√	√	√
37	G712C	A238P											√		
38	G795A	L265=			√	√		√				√	√	√	√
39	T798C	G266=				√		√					√		√
40	A867T	R289S			√									√	
41	G868A	D290N	√	√	√	√	√	√	√	√		√	√	√	√
42	T912G	V304=			√	√	√	√				√	√	√	√
43	A915G	L305=													√
44	A930G	V310=												√	
45	A977G	K326R					√								
46	G1003T	V335L			√			√					√	√	√
47	A1068T	S356=						√							
48	C1088T	T363I									√				
49	A1089G	T363=					√								
50	G1147A	E383K				√									

√ symbol indicates presence of the SNP

Table 8: List of SNPs in *CYCB* gene and their probable effects as predicted by PARSESNP. (Refer to corresponding serial numbers in Table 7 for species/haplotype details)

Sl. No.	Nucleotide Change	Amino acid change	Restriction Enyzme Differences from REBASE		PSSM Difference[a]	SIFT Score[b]
			Gained in Variant	Lost from Reference		
1	T55C	Y19H	-	SfeI	-1.7	0.85
2	G59A	R20K	-	SfeI	1.0	1.00
3	G60A	R20=	BglII, DpnI, Hin4I, MboI, XhoII	-	-	-
4	G67T	V23F	-	-	-2.8	0.54
5	A77T	N26I	ApoI, TspEI	-	1.5	0.85
6	T81C	P27=	MnlI	-	-	-
7	C98T	T33I	BccI, HphI	-	-0.2	0.92
8	T125C	L42P	MnlI	BbvII, MboII	-3.5	0.21
9	G131A	R44K	MseI	DdeI	-	-
10	A232G	N78D	BinI, DpnI, Hpy188I, MboI	AsuII, HinfI, TaqI, TfiI	0.1	1.00
11	A233G	N78S	BplI, PleI	AsuII, TfiI	-	-
12	C249T	D83=	-	BtrI, Hpy99I, MaeII	-	-
13	C261T	I87=	-	BsaBI, Hpy188I	-	-
14	A264G	G88=	-	AluI, Hpy188I	-	-
15	C286T	L96=	-	Cac8I, CviJI, MaeI, NheI	-	-
16	A317C	K106T	Tsp4CI	MseI	-	-
17	C345T	L115=	-	-	-	-
18	G406A	D136N	TspEI	MboII	-0.1	0.00

Sl. No.	Nucleotide Change	Amino acid change	Restriction Enyzme Differences from REBASE		PSSM Difference[a]	SIFT Score[b]
			Gained in Variant	Lost from Reference		
19	A407G	D136G	-	MboII	0.3	0.70
20	A443G	H148R	-	CviRI, MslI	3.5	1.00
21	A459G	K153=	-	-	-	-
22	T462C	T154=	BstXI	DdeI	-	-
23	G465A	K155=	SspI	DdeI	-	-
24	A463ATA	Frameshift	-	DdeI	-	-
25	G476A	R159K	-	BbvII, MboII	-	-
26	A493T	S165C	-	-	-	-
27	G497A	R166K	-	-	-	-
28	G510A	K170=	TspEI	Eco57I, Eco57MI	-	-
29	T525C	N175=	Tsp4CI	-	-	-
30	A556G	K186E	TspDTI	PsiI	0.2	1.00
31	G570A	W190*	-	-	-	-
32	A600G	S200=	TaqI	MfeI, TspEI	-	-
33	A614G	D205G	-	-	0.3	0.70
34	A623G	K208R	MnlI	MboII	-1.0	0.90
35	A683G	D228G	Cac8I		0.3	0.70
36	G686A	R229K	-	HaeIII	1.0	1.00
37	G712C	A238P	ApoI, MnlI		-3.1	0.32
38	G795A	L265=	-	-	-	-
39	T798C	G266=	BsrDI	-	-	-
40	A867T	R289S	-	-	2.4	0.92
41	G868A	D290N	ApoI, TspEI	-	-0.1	0.00
42	T912G	V304=	-	-	-	-
43	A915G	L305=	-	-	-	-

Sl. No.	Nucleotide Change	Amino acid change	Restriction Enyzme Differences from REBASE		PSSM Difference[a]	SIFT Score[b]
			Gained in Variant	Lost from Reference		
44	A930G	V310=	-	-	-	-
45	A977G	K326R	-	-	-2.8	1.00
46	G1003T	V335L	-	-	0.6	0.89
47	A1068T	S356=	-	-	-	-
48	C1088T	T363I	-	-	-0.2	0.92
49	A1089G	T363=	-	-	-	-
50	G1147A	E383K	-	Hpy99I, MnlI, TaqI	4.0	0.57

[a]PSSM: Position Specific Scoring Matrix (>10:deleterious)

[b]SIFT: Sorting Intolerant From Tolerant (<0.05:deleterious)

Table 9: List of amino acid changes identified and changes in their properties. (Refer to corresponding serial numbers in Table 7 for species/haplotype details)

Sl. No.	Nucleotide change	Amino acid change	Change in amino acid properties		
			Side-chain Polarity*	Side-chain charge (pH 7.4)*	Hydropathy index@
1	T55C	Y19H	polar:polar	neutral:positive	−1.3:−3.2
2	G59A	R20K	basic polar:polar	positive:positive	−4.5:−3.9
4	G67T	V23F	nonpolar:nonpolar	neutral:neutral	4.2:2.8
5	A77T	N26I	polar:nonpolar	neutral:neutral	−3.5:4.5
7	C98T	T33I	nonpolar:nonpolar	neutral:neutral	4.5:4.5
8	T125C	L42P	nonpolar:nonpolar	neutral:neutral	3.8:−1.6
9	G131A	R44K	basic polar:polar	positive:positive	−4.5:−3.9
10	A232G	N78D	polar:polar	neutral:negative	−3.5:−3.5
11	A233G	N78S	polar:polar	neutral:neutral	−3.5:−0.8
16	A317C	K106T	polar:nonpolar	positive:neutral	−3.9:4.5
18	G406A	D136N	polar:polar	negative:neutral	−3.5:−3.5
19	A407G	D136G	polar:nonpolar	negative:neutral	−3.5:−0.4
20	A443G	H148R	basic polar:basic polar	positive(10%) neutral(90%) :positive	−3.2:−4.5
25	G476A	R159K	basic polar:polar	positive:positive	−4.5:−3.9
26	A493T	S165C	polar:nonpolar	neutral:neutral	−0.8:2.5
27	G497A	R166K	basic polar:polar	positive:positive	−4.5:−3.9
30	A556G	K186E	polar:polar	positive:negative	−3.9:−3.5
33	A614G	D205G	polar:nonpolar	negative:neutral	−3.5: −0.4
34	A623G	K208R	polar:basic polar	positive:positive	−3.9:−4.5
35	A683G	D228G	polar:nonpolar	negative:neutral	−3.5: −0.4
36	G686A	R229K	basic polar:polar	positive:positive	−4.5:−3.9
37	G712C	A238P	nonpolar:nonpolar	neutral:neutral	1.8:−1.6

Sl. No.	Nucleotide change	Amino acid change	Change in amino acid properties		
			Side-chain Polarity*	Side-chain charge (pH 7.4)*	Hydropathy index[@]
40	A867T	R289S	basic polar:polar	positive:neutral	−4.5:−0.8
41	G868A	D290N	polar:polar	negative:neutral	−3.5:−3.5
45	A977G	K326R	polar:basic polar	positive:positive	−3.9/−4.5
46	G1003T	V335L	nonpolar:nonpolar	neutral:neutral	4.2:3.8
48	C1088T	T363I	nonpolar:nonpolar	neutral:neutral	4.5:4.5
50	G1147A	E383K	polar:polar	negative:positive	−3.5/−3.9

* As described by Hausman and Cooper (2004)

[@]As described by Kyte and Doolittle (1982)

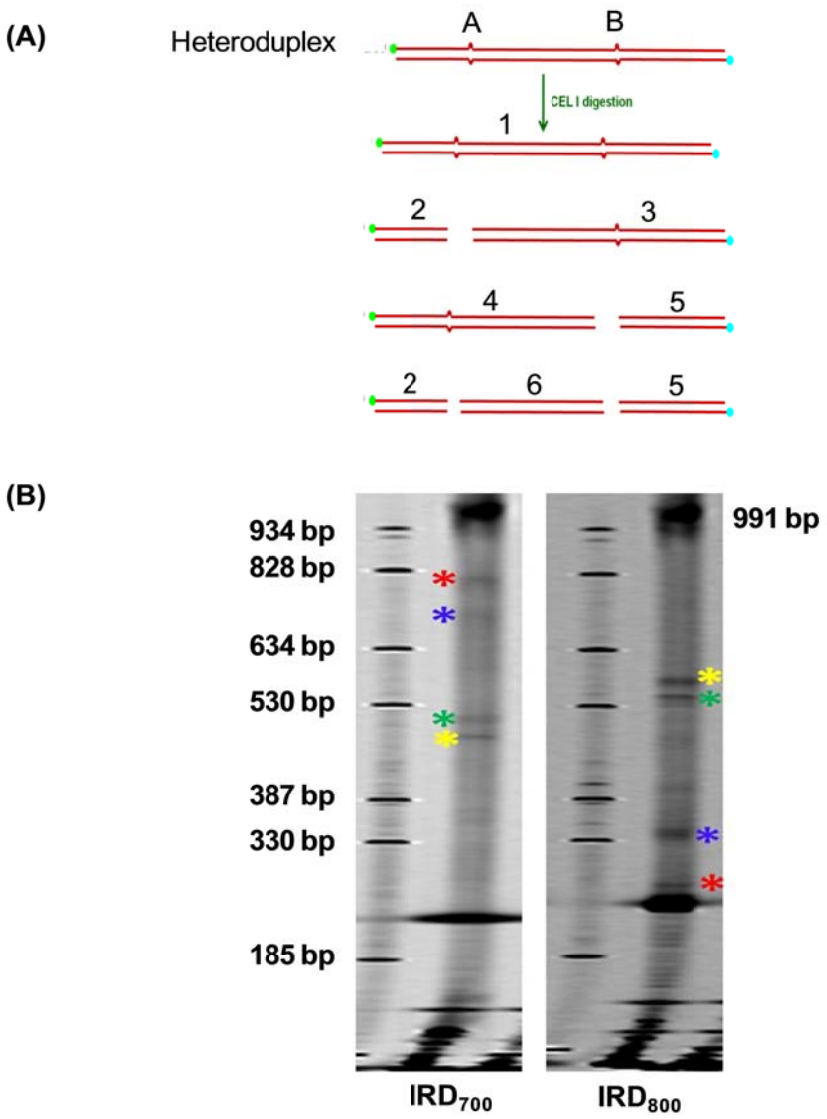
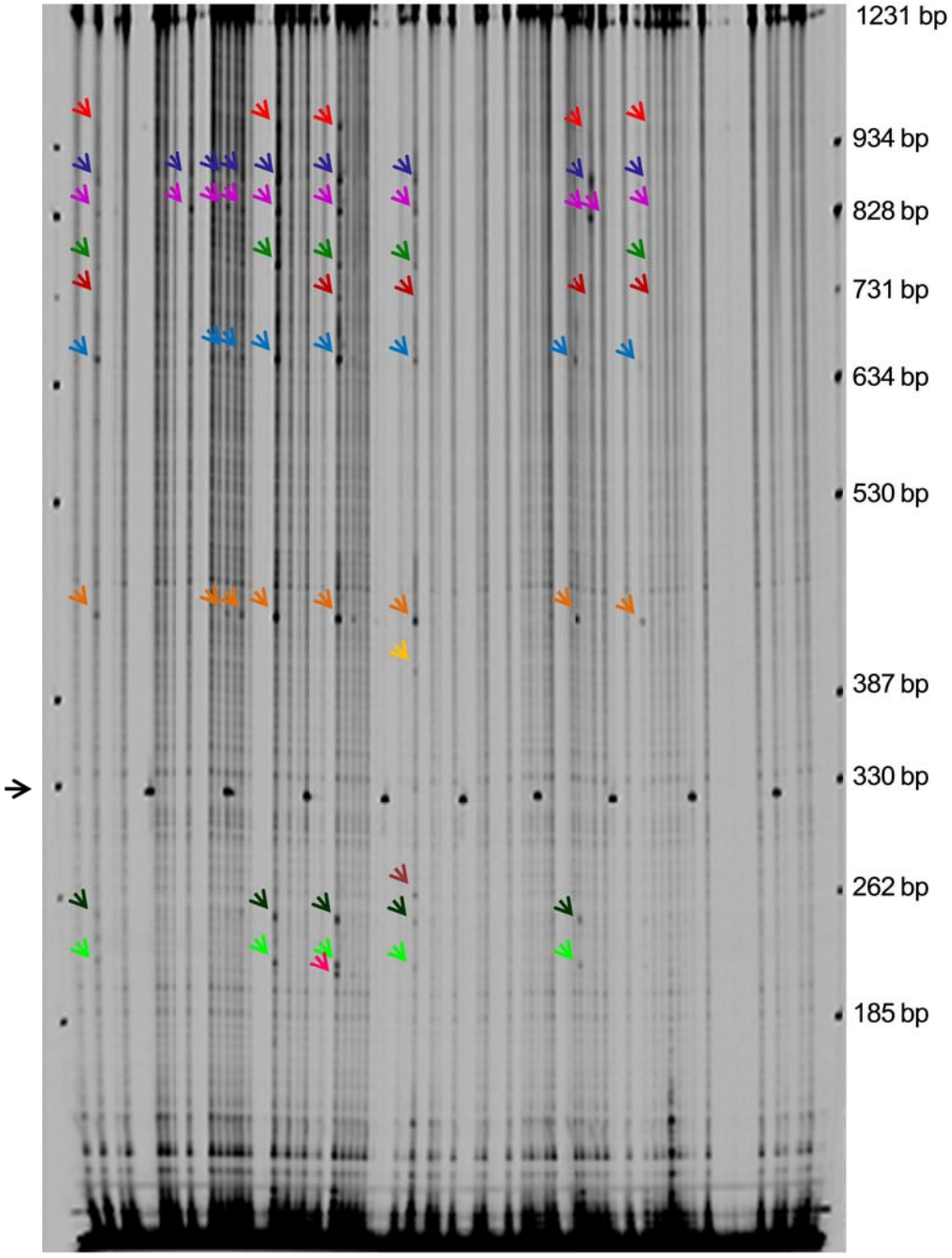


Figure 10: Schematic representation of multiple product formation after CEL I digestion (A). Green and blue colors represents IRD_{700} and IRD_{800} labeling respectively. A heteroduplex with two mis-matches (A and B) can give 8 different products after mis-match cleavage by CEL I digestion. The numbers indicate different types of products formed. Product 1 is completely undigested one having both the IRDye labels and so will be visible on both IRD_{700} and IRD_{800} channels. Product 2 and 3 are formed when only mis-match A is cleaved. Product 4 and 5 are formed when only mis-match B is cleaved. Products 2, 5 and 6 are formed when both the mis-matches are cleaved. Products 2 and 4 will be visible in IRD_{700} channel only, whereas, products 3 and 5 will be visible in IRD_{800} channel only. Product 6 lacks any of the labels and so will not be visible in any channel.
A LI-COR image showing detection of 4 SNPs in a sample (B). Same colour asterisk symbol in both channels indicates the position of complementary cut products. Molecular size marker is shown on the left side of both the channels. Size of the main PCR product is 991 bp.



…cont'd in the next page

…cont'd from previous page

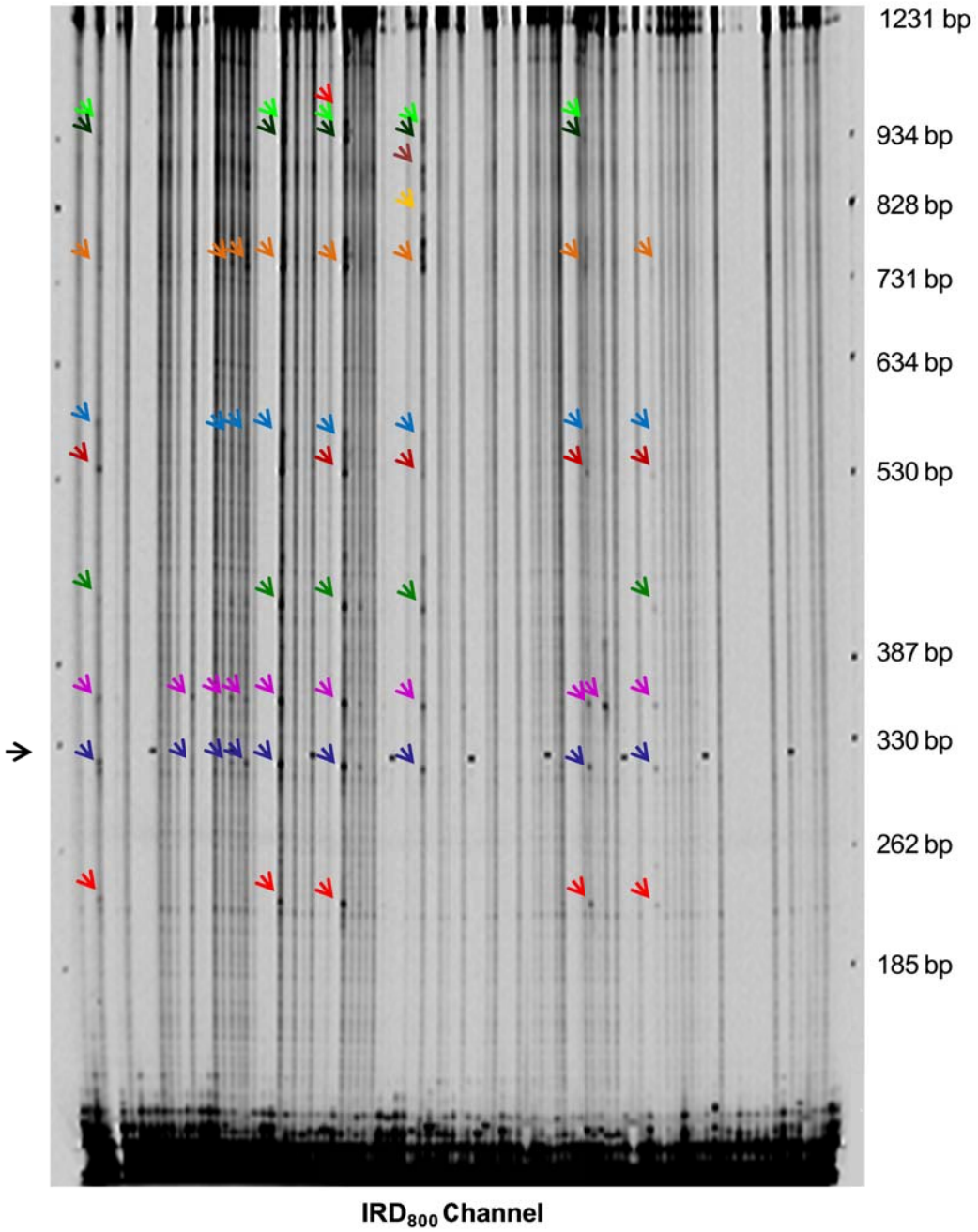


Figure 11: LI-COR image showing the presence of SNPs in *CYCB* gene. 96 samples each representing one accession were loaded in the gel. *S. lycopersicum* cv. Arka Vikas was used as WT for heteroduplex formation. Molecular size markers were loaded on both sides, the sizes of which are given on the right hand side. Size of the main product is 1231 bp. Another marker loaded on every 10th lane is visible at 330 bp position (indicated with horizontal black arrow). SNPs are marked with colored arrows. Each color indicates an SNP in different accessions. Complementary positions in IRD700 and IRD800 channels are indicated with same color of the arrow.

of the SNPs. In all the cases the chromatograms showed single peak at the position of base change indicating that all the SNPs are present in a homozygous condition in the respective accessions. Being homozygous in nature, any of these SNPs can be conveniently used for crossing experiments in future for transferring associated trait.

4.2.2.1. Frequency of SNPs in the population

Out of 543 accessions screened, 13 accessions showed at least one SNP. A total of 94 SNPs were detected from 6, 68, 433 bp screened (1231 bp each from 543 accessions). The SNP frequency in *CYCB* gene in the population was calculated as 1/7.11 Kb. The details of the SNPs in different accessions are given in Table 7. Out of 50 SNPs identified in various wild relatives and other tomato accessions, 18 are common in wild relative/s and other accession/s.

4.2.2.2. Haplotypes

Based on the number and type of SNPs detected in the *CYCB* region, 8 different haplotypes were identified. 530 accessions including the reference cultivar Arka Vikas were included in haplotype 1 (HT.1). Two accessions with a single SNP were classified in haplotype 2 (HT.2). Three accessions with two SNPs each were classified as haplotype 3 (HT.3).Only one accession each was identified for haplotypes 5, 6, 7 and 8 (HT.5, HT.6, HT.7 and HT.8). Maximum number of SNPs was identified in the accession EC20636 (HT.8). The details of the haplotypes are given in Table 10. The schematic representation of the distribution and types of SNPs in the screened region in various haplotypes is shown in Figure 12.

Table 10: List of haplotypes in *CYCB* gene identified from natural populations

Haplotype	No.of accessions	Acc. No.	No. of SNPs
HT.1	530	Arka Vikas and other accessions	-
HT.2	2	EC520046, EC129602	1
HT.3	3	EC520052, EC34477, Cerasiformae	2
HT.4	4	EC8936, EC34480, EC20639, EC25563	6
HT.5	1	EC163598	12
HT.6	1	LA0348	15
HT.7	1	LA3000	14
HT.8	1	EC20636	21

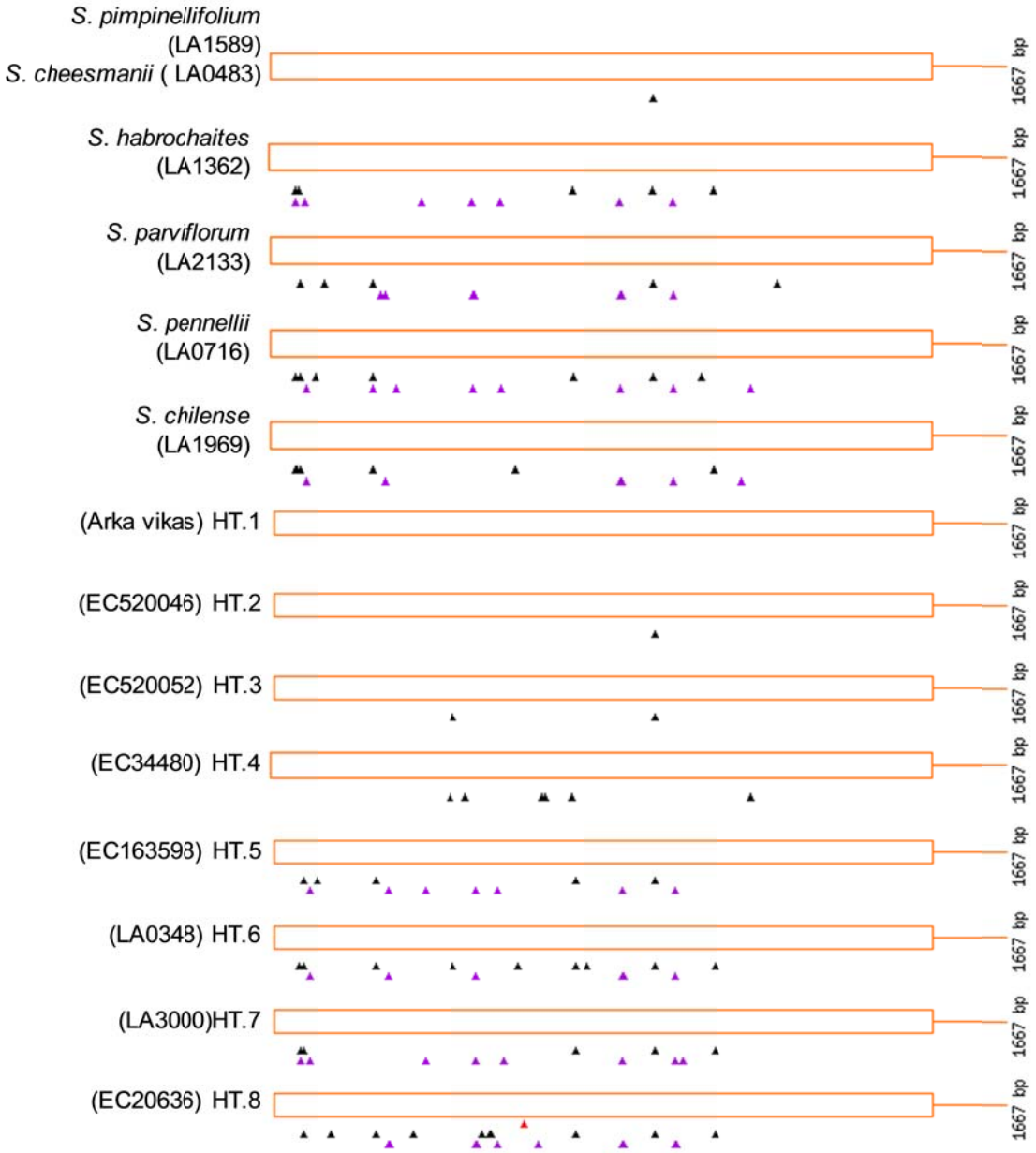


Figure 12: Distribution of SNPs in *CYCB* gene in various wild relatives and haplotypes identified from natural populations. The exon (1-1497 bp) and 3´UTR (1498-1667 bp) of the gene is diagrammatically represented with bar and line respectively. Black, purple and red triangles indicate the positions of nonsynonymous, synonymous and nonsense nucleotide substitutions respectively. TGRC accession numbers of the wild relatives are given in brackets. For each haplotype, accession number of one representative accession is given in brackets. Arka vikas is considered as haplotype 1.

4.2.2.3. Transitions and transversions

The base substitutions identified in the accessions were classified into four different types of transitions and five different transversions. In general transitions occurred with much higher frequency than transversions with a transition/transversion ratio 5.0. The interchange of the purine rings (A↔G) was more predominant with maximum frequency which is more than double that of the pyrimidine rings (C↔T), whereas, very low frequencies were observed for the transversions A→C, T→G and G→C. Graphical representation of the frequencies for various substitution reactions is given in the Figure 13.

4.2.2.4. Synonymous and non-synonymous changes

Most of the base substitutions resulted in synonymous mutations. The synonymous/nonsynonymous ratio of the SNPs in the screened region of *CYCB* is 1.7.

4.2.2.5. Amino acid changes and other possible effects

Online software: PARSESNP (Project Aligned Related Sequences and Evaluate SNPs) (http://www.proweb.org/parsesnp/) was used to analyze the effects of the polymorphisms on the expressed gene product. The PSSM (Position Specific Scoring Matrix) scores for all SNPs were very low and so these changes were predicted as non-deleterious. The SIFT (Sorting Intolerant From Tolerant) scores for the nucleotide changes G406A (HT.2 and HT.6) and G868A (all haplotypes except HT.4) were 0.00 and are predicted as deleterious. Table 8 shows lists of restriction enzyme polymorphisms caused by the changes. For each amino acid change, the change in their properties like side chain polarity, side chain charge and hydropathy index were studied (Table 9). It

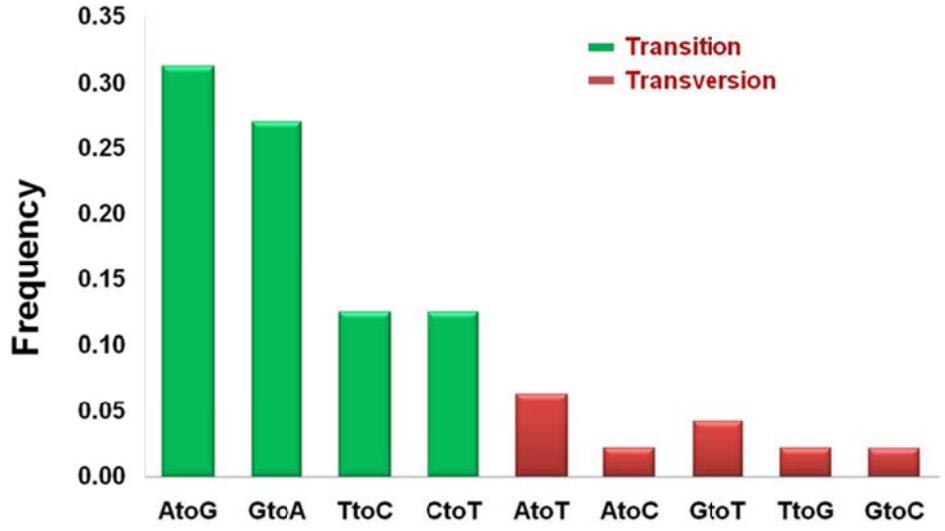


Figure 13: Frequency of various base substitutions in *CYCB* gene.

shows that in many cases the amino acid changes are causing a change in the side chain polarity, side chain charge and hydropathy index.

Of interest was a nucleotide change A317C which caused an amino acid change from Lysine to Threonine at position 106. Lysine has a polar side chain with positive charge and with a hydropathy index −3.9, whereas, threonine has a nonpolar side chain which is neutral in charge and with a hydopathy index 4.5. This amino acid change might have affected the enzyme activity of CYCB protein. This SNP was identified only in one accession EC20636 (HT.8), which is a green fruited one. In view of this it is difficult to ascribe any function to it and relate to any phenotypic effect such as change in lycopene and *β*-carotene composition. Further studies are needed to see the effect of this SNP by crossing it with a red fruited tomato line.

4.2.2.6. Truncation of CYCB enzyme

A base change G570A identified in the accession EC20636 converts the codon TGG (codes for tryptophan) into TGA (stop codon) and thus would cause a truncation in the CYCB protein after 189 amino acids (Figure 14). Considering that the above accession is a green fruited one, where the carotenoid biosynthetic pathway is inactive in fruit ripening, the functional role of this allele on the fruit phenotype is not discernible. Nevertheless, this loss of function allele is a good candidate for developing a high lycopene containing cultivar and hence could act as a donor plant for back cross breeding with established cultivars.

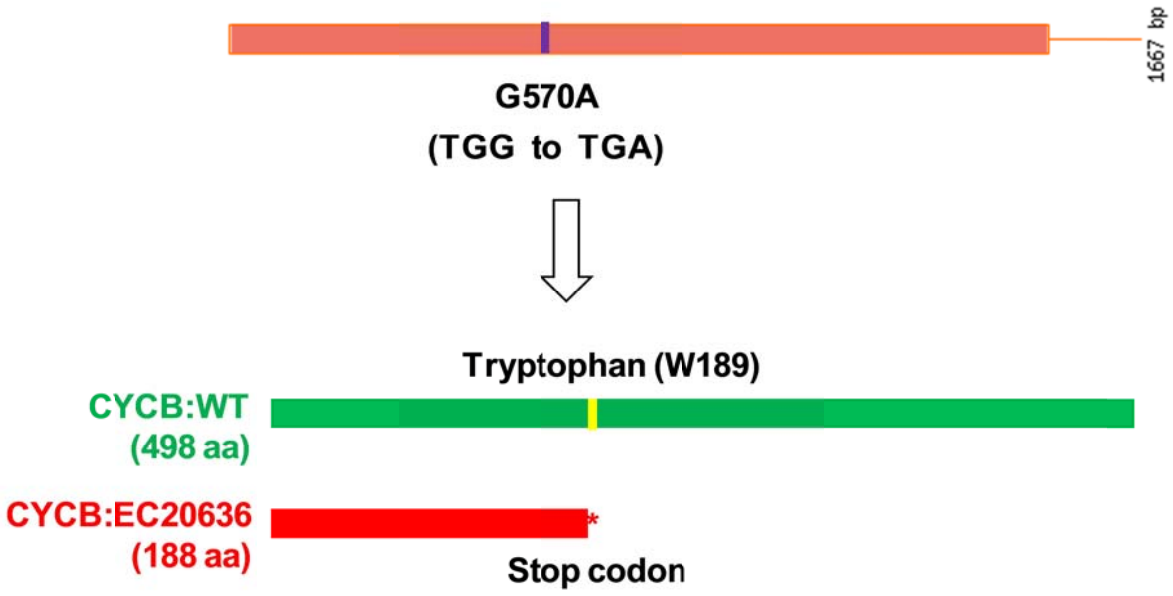


Figure 14: Probable loss of function allele of CYCB in EC20636. A base change G570A converts the codon TGG (codes for tryptophan) to TGA (stop codon). Purple color represents the position of base substitution in the nucleotide sequence. CYCB in wild type and EC20636 are represented by green and red bars respectively. Yellow color represents the position of tryptophan (W189).

4.2.2.7. Change in codon usage efficiency

In case of SNPs which results in silent mutation (synonymous codons), a probable change in the translational efficiency is expected due to a difference in the abundance of various tRNA molecules and the preference of various genes for the type of tRNA molecule. The change in the codon usage efficiency for such SNPs observed, based on the information from codon usage database (http://www.kazusa.or.jp/codon/) is also given in Table 11. For most of the amino acids the change in nucleotide causes change in the codon usage efficiencies. But, in case of Lysine (K) there is not much difference in the codon usage efficiencies. In case of the silent changes P27=, G88=, T154=, N175=, S200=, L265=, G266=, V304= and T363= there is drastic decrease in the codon usage efficiency compared to the reference (*S. lycopersicum*). The SNPs corresponding to these are present mostly in green fruited accessions and wild relatives. It is an indication of increase in the translational efficiency in colored fruited species compared to the green fruited ones. The change in the codon usage efficiencies might be altering the quantity of the CYCB enzyme synthesized and thus may affect the rate of conversion of lycopene to *β*-carotene.

A combined phylogram was prepared based on the nucleotide sequences of *CYCB* from different wild relatives and accessions, to study the evolutionary pattern (Figure 15). All the green fruited wild relatives and accessions belonging to haplotypes 5, 6, 7 and 8 were clustered together, whereas, haplotypes 1,2,3 and 4 were clustered with colored fruited species. The genetic variability of *CYCB* in cultivated tomato accessions is low and the few accessions that showed SNPs may have accumulated them through hybridization with wild relatives. The low allele frequencies observed in this gene in colored fruited accessions indicate that this gene is under stabilizing selection, whereas,

green fruited wild relatives of tomato and other accessions showed high amount of variation in *CYCB* locus. In the green fruited varieties this locus might lack any selection pressure as accumulation of SNPs in this locus has no phenotypic consequence in them.

Table 11: List of synonymous nucleotide changes and the corresponding change in the codon usage efficiencies. (Refer to corresponding serial numbers in Table 7 for species/haplotype details)

Sl. No.	Nucleotide change	Amino acid change	Original codon*	Change in codon*
3	G60A	R20=	AG<u>G</u> (11.9)	AG<u>A</u> (16.4)
6	T81C	P27=	CC<u>T</u> (19.2)	CC<u>C</u> (5.7)
12	C249T	D83=	GA<u>C</u> (15.0)	GA<u>T</u> (39.3)
13	C261T	I87=	AT<u>C</u> (14.0)	AT<u>T</u> (28.2)
14	A264G	G88=	GG<u>A</u> (25.6)	GG<u>G</u> (10.8)
15	C286T	L96=	GG<u>C</u> (9.7)	GG<u>T</u> (23.7)
17	C345T	L115=	CT<u>C</u> (11.2)	CT<u>T</u> (24.9)
21	A459G	K153=	AA<u>A</u> (31.1)	AA<u>G</u> (31.0)
22	T462C	T154=	AC<u>T</u> (19.9)	AC<u>C</u> (8.6)
23	G465A	K155=	AA<u>G</u> (31.0)	AA<u>A</u> (31.1)
28	G510A	K170=	AA<u>G</u> (31.0)	AA<u>A</u> (31.1)
29	T525C	N175=	AA<u>T</u> (30.5)	AA<u>C</u> (17.3)
32	A600G	S200=	TC<u>A</u> (20.7)	TC<u>G</u> (5.6)
38	G795A	L265=	TT<u>G</u> (24.2)	TT<u>A</u> (14.4)
39	T798C	G266=	GG<u>T</u> (23.9)	GG<u>C</u> (9.7)
42	T912G	V304=	GT<u>T</u> (28.0)	GT<u>G</u> (16.0)
43	A915G	L305=	TT<u>A</u> (14.4)	TT<u>G</u> (24.2)
44	A930G	V310=	GT<u>A</u> (11.2)	GT<u>G</u> (16.0)
47	A1068T	S356=	TC<u>A</u> (20.7)	TC<u>T</u> (21.2)
49	A1089G	T363=	AC<u>A</u> (17.9)	AC<u>G</u> (4.6)

*Nucleotide change is underlined; codon usage efficiency is given in brackets.

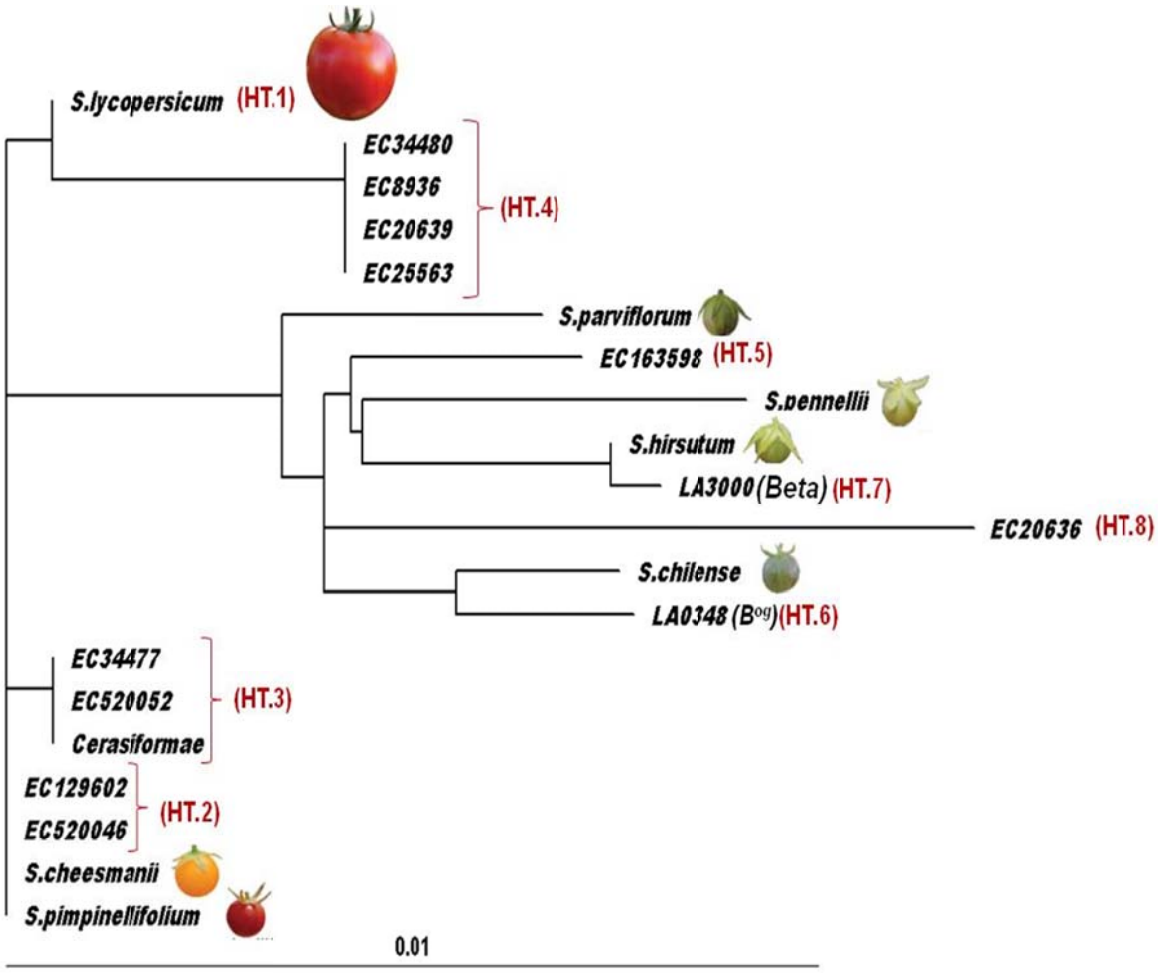


Figure 15: Phylogram based on *CYCB* gene sequences from different tomato accessions and wild relatives. Clustering was done using ClustalW2 software (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Distance scale is given at the bottom. Haplotypes of the accessions are indicated in brackets. Fruit phenotypes of wild relatives are shown.

4.3. Diversity studies of *CYCB* gene promoter in different wild relatives and accessions of tomato

4.3.1. Genetic variations of *CYCB* promoter in wild relatives of tomato

In order to find out natural variations of *CYCB* promoter among different wild relatives of *S. lycopersicum*, approximately 1 Kb 5′upstream sequence of *CYCB* gene was PCR amplified and the PCR product lengths were compared. It was found that the PCR product from *S. pimpinellifolium* and *S. cheesmanii* were of the same size as that from *S. lycopersicum,* whereas the products from *S. pennellii, S. habrochaites* and *S. parviflorum* were larger in size. (Figure 16A). Alignment of the respective sequences of various *Solanum* species showed high variation in the form of SNPs and In-Dels (Figure 16B). A 13 bp insertion (5′GAGTT/CTGGGTTC/-A/T3′ (-758 bp)) was observed in all the wild relatives compared to *S. lycopersicum.* A 27 bp insertion (5′ACTTCACCCTTCTTTCTTGTCTTGGTG3′ (-863 bp)) and two 8 bp insertions (5′CGAAGTAT3′ (-477 bp) and 5′CTAAATAT3′ (-310 bp)) were observed in *S. pennellii, S. habrochaites* and *S. parviflorum.*

4.3.2. Genetic variations of *CYCB* promoter in various tomato accessions

The study on the variations of the promoter region was extended to different tomato accessions. As a preliminary screening, the amplicon sizes of this region from 543 tomato accessions were compared to that of Arka Vikas cultivar. Several accessions have bigger sized amplicons than that of the reference cultivar indicating the presence of several natural variations in this region among the accessions (Figure 17).

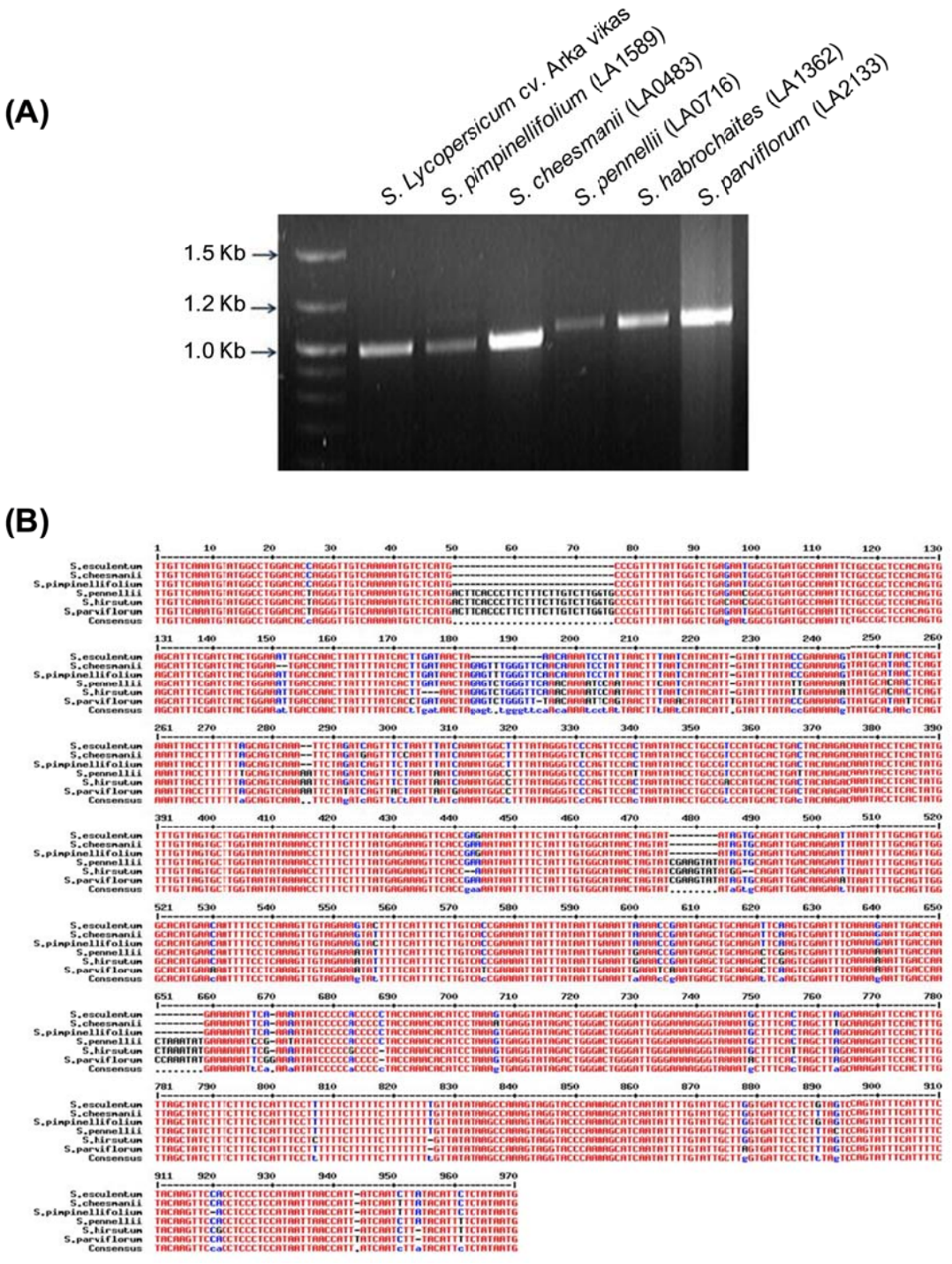


Figure 16: Variations in *CYCB* promoter length in different wild relatives of tomato **(A)**. Alignment of the sequences of *CYCB* promoter in some of the tomato species showing insertions in case of the green fruited species **(B)**.

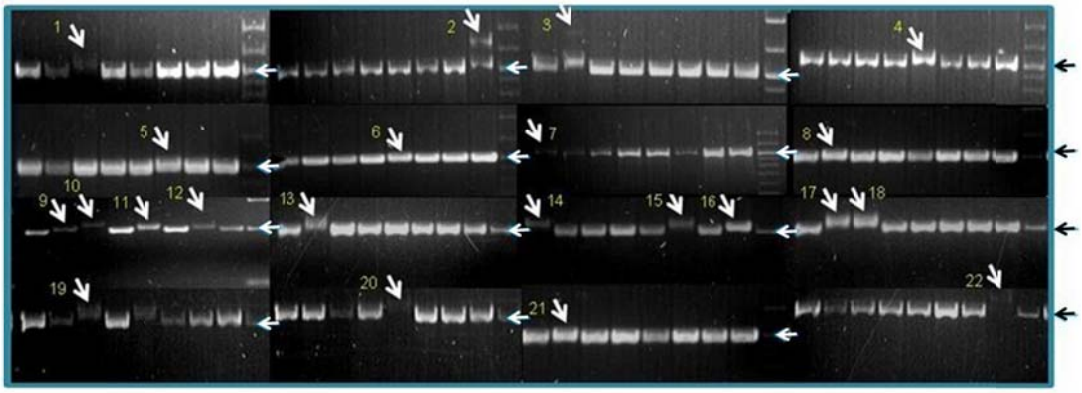


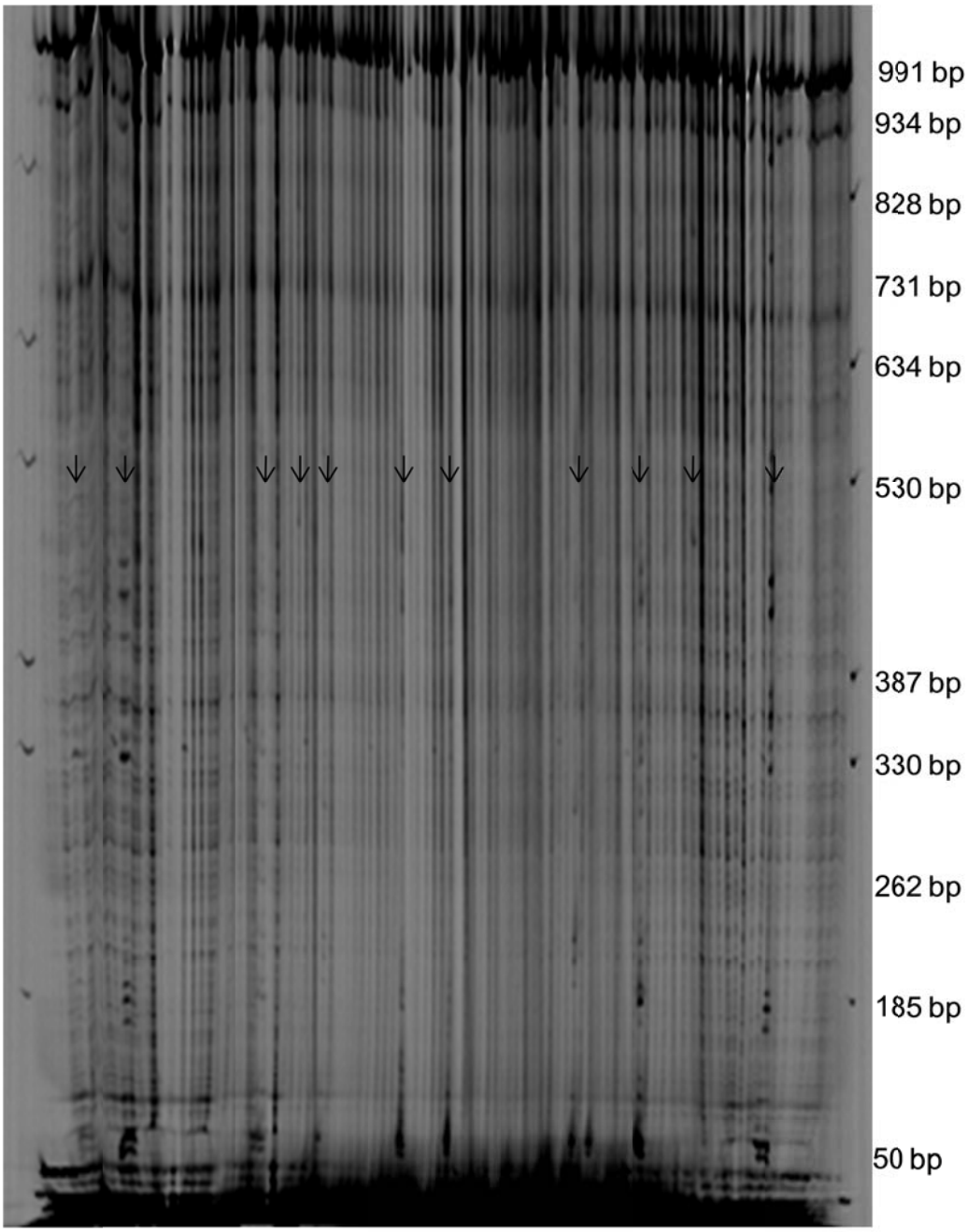
Figure 17: Variation of *CYCB* promoter length in different accessions of tomato. Horizontal arrows indicate 1Kb position of the molecular size marker. Vertical slanting arrows indicate the fragments showing increase in length compared to Arka vikas. The list of accession numbers corresponding to the numbers indicated near the arrows is given below.

1) EC538141	7) EC50055	13) EC8936	19) EC32311
2) WIR3768	8) EC490136	14) EC25563	20) EC89248
3) EC163598	9) EC252	15) EC20639	21) IC469700
4) EC27910	10) EC202	16) EC20636	22) EC2630
5) EC520046	11) EC1129	17) EC34480	
6) IC402307	12) EC492	18) EC34477	

Screening of all these accessions for the presence of SNPs by EcoTILLING showed the presence of large number of nucleotide substitutions. Figure 18 shows an example of LI-COR gels showing the presence of SNPs in various accessions. Sequence analysis revealed the occurrence of several In-Dels and substitutions. A total of 298 SNPs were detected from 4,33,314 bp screened (798 bp from 543 accessions). Out of 543 accessions, 12 accessions have at least one SNP/In-Del. Nine different haplotypes were identified with an SNP frequency of 1/1.45 Kb. Maximum number of SNPs are present in the accession EC20636 (HT.9), whereas the HT.6 has highest number of In-Dels (Table 12).

4.3.2.1 Identification of a Nup (Nuclear localized plastid) DNA fragment

When the promoter lengths were compared for different accessions by PCR, three accessions showed about 250 bp longer PCR products compared to that of Arka vikas. The accessions are LA0348, LA0500 (accessions carrying the *Beta-old-gold* mutation) and LA0458 (an accession of *S. chilense*) (Figure 19A). Nucleotide BLAST in NCBI database showed similarity of the 250 bp additional sequence to a fragment of *rps4 (ribosomal protein small subunit 4)* gene of the tomato chloroplast genome (Figure 19B). Literature search indicated that source of *og* was a plant from the F_2 progenies of a cross between *S. chilense* and *S. lycopersicum. Old-gold* phenotype has been transferred to the latter species by nine successive backcrosses (Rick and Smith, 1953; Rick and Dempsey, 1961). Therefore, it is likely that the Nup DNA fragment might have first got integrated into the nuclear genome in *S. chilense* and later got introgressed into *S.lycopersicum* during crossing.



…cont'd in the next page

…cont'd from previous page

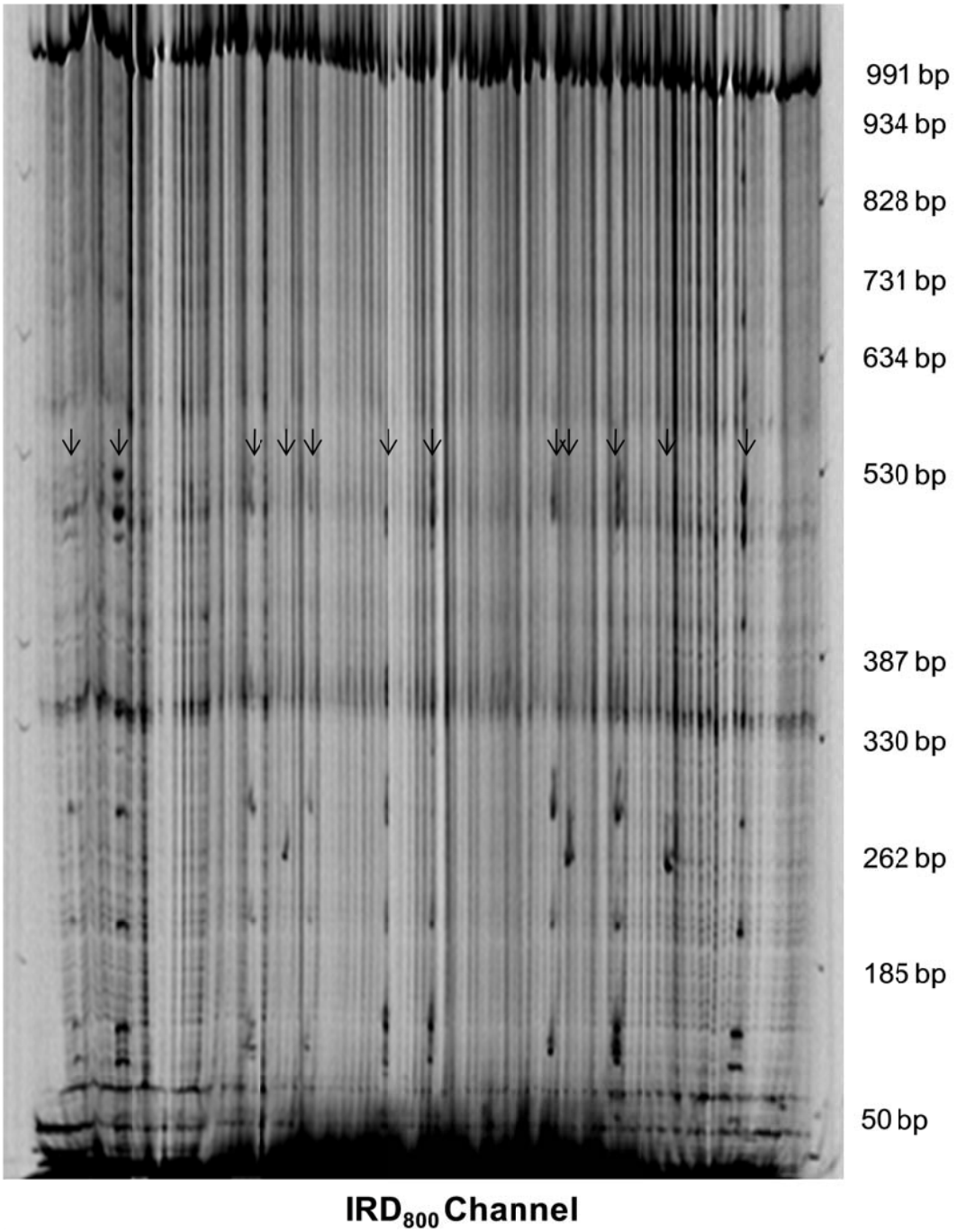


Figure 18: LI-COR image showing the presence of SNPs in *CYCB* gene promoter. 96 samples each representing one accession were loaded in the gel. *S. lycopersicum* cv. Arka Vikas was used as WT for heteroduplex formation. Molecular size markers were loaded on both sides, the sizes of which are given on the right hand side. Size of the main product is 991 bp. Lanes marked with vertical arrows indicate samples with SNPs.

Table 12: List of haplotypes based on *CYCB* gene promoter identified from natural populations

Haplotype	No. of accessions	Acc. No.	No. of SNPs	No. of In-Dels
HT.1	531	Arka Vikas and other accessions	-	-
HT.2	1	WIR3768	-	2
HT.3	2	EC520046, EC129602	2	1
HT.4	2	EC520052, EC34477	3	2
HT.5	1	LA0348	32	6
HT.6	2	EC34480, EC20639	32	11
HT.7	1	LA3000	34	9
HT.8	2	EC8936, EC25563	38	10
HT.9	1	EC20636	82	10

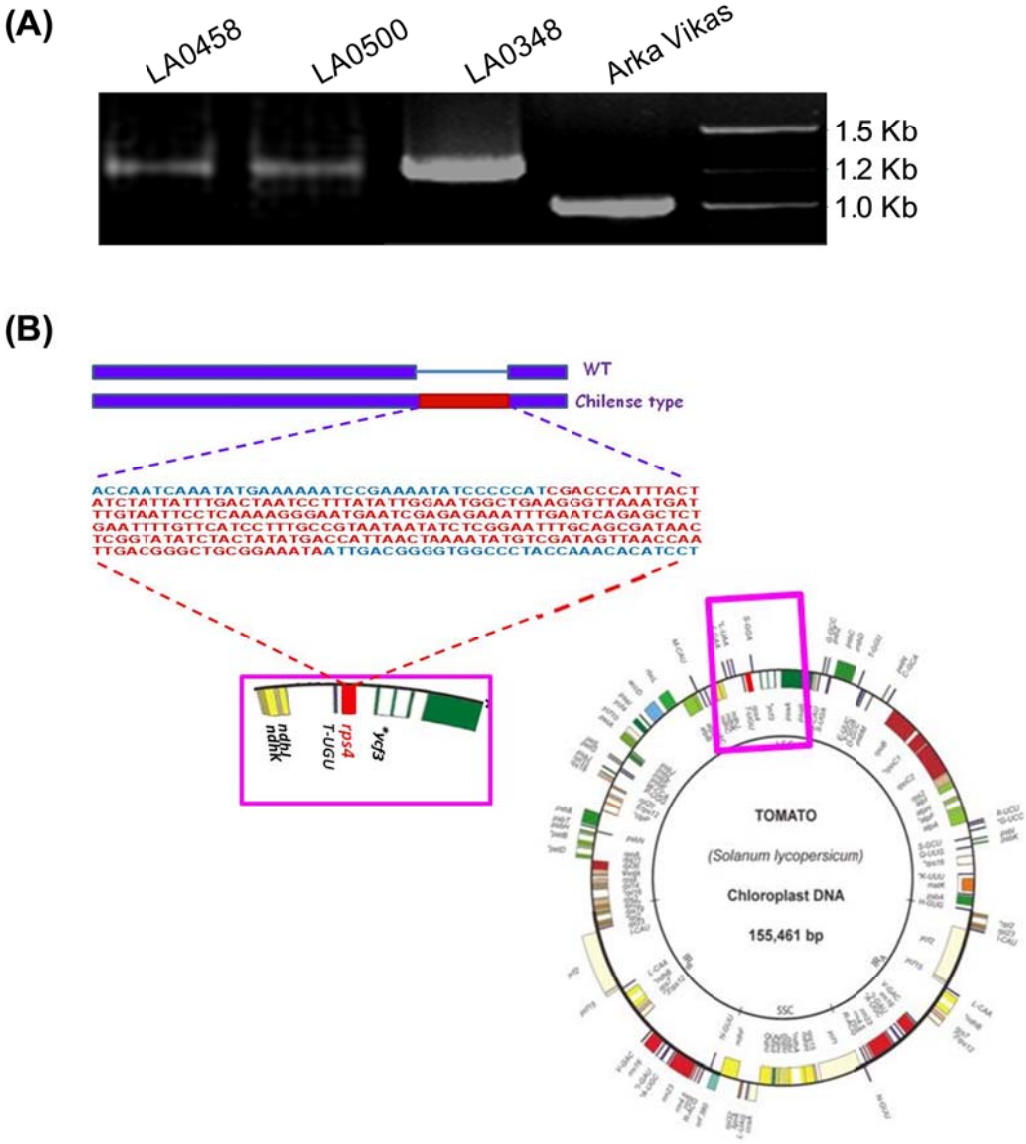


Figure 19: Highly variable *CYCB* promoter sequences identified in few accessions (A). LA0458: *Cmr Lv*, a mutant accession in *S. challenge* background. LA0500 and LA0348: B^{og} mutant accessions in *S. lycopersicum* background. *S. lycopersicum* cv. Arka Vikas was used as WT for comparison. The size of the amplicon from WT is 1007 bp.
Identification of a probable Nup (Nuclear localized plastid) genome fragment (B). WT and *Chilense* type sequences are diagrammatically represented. Red region in the diagram represents the insertion. Red and blue color sequences represent the 240 bp insertion sequence and the flanking sequence. A complete gene map representing the circular chloroplast genome from *S. lycopersicum* and an enlarged fragment including *rps4* (*ribosomal protein subunit 4*) gene is given.

4.3.3. Evolutionary relations between various tomato accessions and wild relatives

A combined phylogram was prepared based on the nucleotide sequences from different haplotypes and wild relatives. It shows the similarity of few haplotypes with the wild relatives indicating that those accessions might have obtained the variations through hybridizations. Green fruited accessions and wild relatives were clustered together. The accession EC20636 (HT.9) was found to be the most distant one with maximum nucleotide variations. Colored fruited accessions were clustered with colored fruited tomato species (Figure 20).

4.3.4. Regulatory elements in *CYCB* gene promoter from various sources

In order to find out natural variations (insertions/deletion/substitutions), in the 1Kb upstream region of *CYCB* gene, a comparison of the sequences from different wild relatives of *S. lycopersicum* was carried out. To identify specific *cis*-elements which are involved in repression or induction of gene expression, the sequences were analyzed using the online software 'Softberry N site' (http://linux1.softberry.com/berry.phtml?topic=nsitep&group=programs&subgroup=promoter). The response elements identified by the site along with the known binding factors for different promoters are given in Table 13. The functions of various response elements are given in Table 14. Several hormone responsive, stress responsive and light responsive elements and binding sites for transcription enhancers were identified. Certain elements like ABRE, AT-rich H, box 1, box d, ERE2, ERE3, poly (dA-dT) element, S2 and VP1RE were identified in almost all the promoter sequences analyzed. But, an ABA response element (ABRE3) was present only in the promoters of green fruited tomatoes, whereas, an ethylene response element (ERE1), an auxin response element (AIV) and a light response

element (LRE) were present only in the colored fruited species. ERE1, AIV and LRE must be playing certain crucial roles in tomato fruit ripening process.

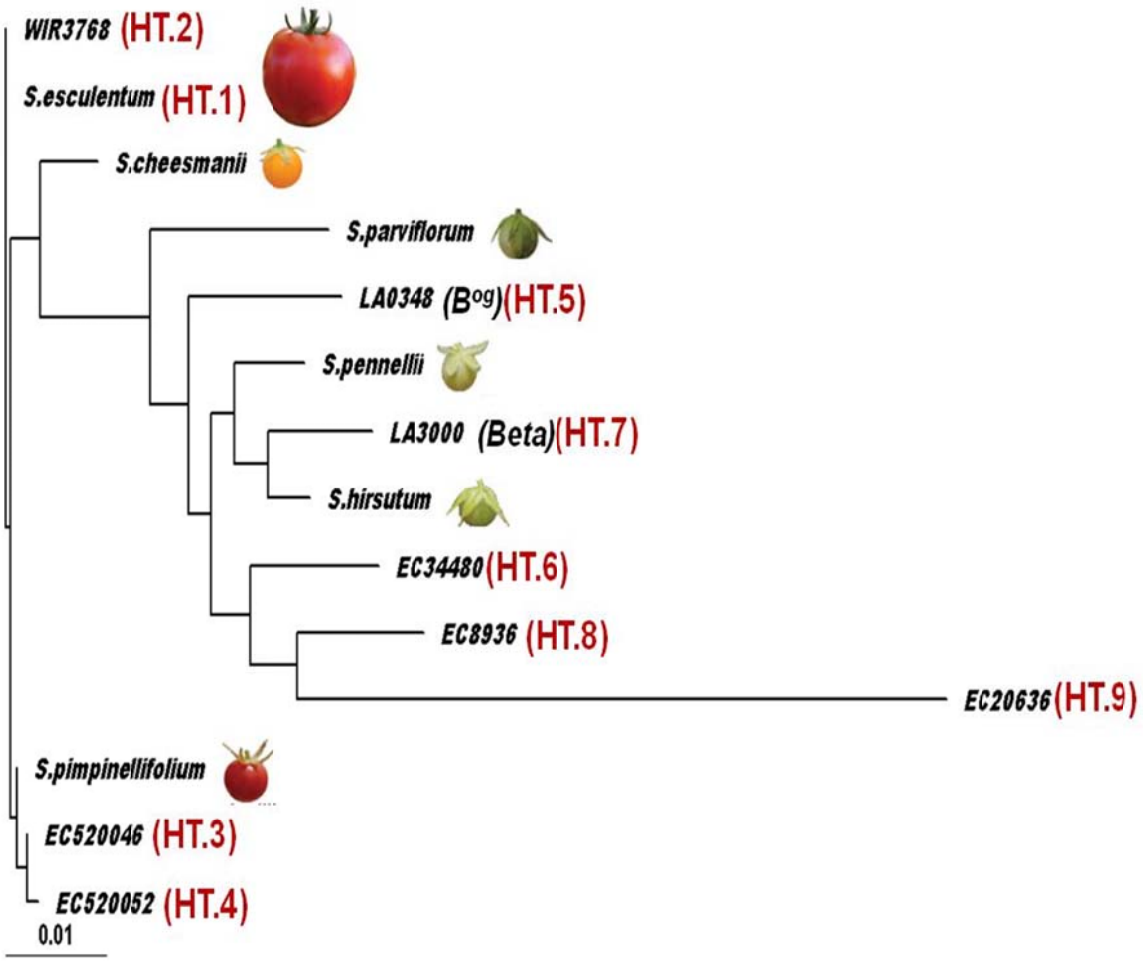


Figure 20: Phylogram based on *CYCB* gene promoter sequences from different tomato accessions and wild relatives. Clustering was done using ClustalW2 software (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Distance scale is given at the bottom. Haplotypes of the accessions are indicated in brackets. Fruit phenotypes of wild relatives are shown.

Table 13: List of response elements identified in *CYCB* promoter of different wild relatives and accessions

Response elements	**Binding Factor**	***S. lycopersicum* (HT.1)**	***S. pimpinellifolium***	***S. cheesmanii***	***S. habrochaites***	***S. parviflorum***	***S. pennellii***	**HT.2**	**HT.3**	**HT.4**	**HT.5**	**HT.6**	**HT.7**	**HT.8**	**HT.9**
ABRE	-	Y(3)	Y(3)	Y(3)	-	Y(3)	Y(3)	Y(3)	Y(3)	Y(3)	Y(3)	-	Y(3)	Y(3)	-
ABRE3	bZIP67 + LEC1/L1L-[NF-YC2]	-	-	-	Y(1)	Y(1)	Y(1)	-	-	-	-	Y(1)	Y(1)	Y(1)	Y(1)
AC-II	-	-	-	-	-	-	-	Y(2)	-	-	-	-	-	-	-
AIV	-	Y(1)	Y(1)	Y(1)	-	-	-	Y(1)	Y(1)	Y(1)	-	-	-	-	-
Alfin1 BS2	Alfin1	-	-	-	-	-	-	-	-	-	-	-	-	-	Y(2)
AT-2b	nuclear proteins from cotyledons	-	-	-	Y(0)	-	Y(0)	-	-	-	-	-	Y(0)	-	-
AT-rich H	-	Y(4)	Y(4)	Y(4)	Y(4)	-	Y(4)	Y(4)	Y(4)	Y(4)	Y(4)	Y(4)	Y(4)	-	-
box 1	GT-1	Y(1)	Y(1)	Y(1)	Y(1)	Y(1)	Y(1)	Y(1)	Y(1)	Y(1)	Y(1)	Y(1)	Y(1)	Y(1)	Y(1)
box d	DOF1	Y(2)	Y(2)	Y(2)	Y(2)	Y(2)	Y(2)	Y(2)	Y(2)	Y(2)	Y(2)	Y(2)	Y(2)	-	Y(2)
CArG (RIN-a)	RIN	-	-	-	-	-	-	-	-	-	-	Y(0)	-	-	-
C1-box	-	-	-	-	-	-	-	-	Y(2)	Y(2)	-	-	-	-	-
CBF-C BS	CBF-C	-	-	-	-	-	-	-	-	-	-	-	-	-	Y(0)
C1BS2	C1	-	-	-	-	-	-	-	-	-	Y(1)	-	-	-	-
ERE1	-	Y(0)	Y(0)	Y(0)		-	-	Y(0)	Y(0)	Y(0)	-	-	-	-	-
ERE2	-	4Y(0, 0,1,1)	4Y(0, 0,1,1)	2Y(0)	4Y(1, 1,1,1)	-	4Y(1, 1,1,1)	4Y(0, 0,1,1)	4Y(0, 0,1,1)	4Y(0, 0,1,1)	4Y(1, 1,1,1)	4Y(1, 1,1,1)	4Y(1, 1,1,1)	4Y(1, 1,1,1)	Y(1)
ERE3	-	2Y(0, 1)	2Y(0, 1)	Y(0)	2Y(1, 1)	Y(1)	2Y(1, 1)	2Y(0, 1)	2Y(0, 1)	2Y(0, 1)	2Y(1, 1)	2Y(1, 1)	2Y(1, 1)	2Y(1, 1)	Y(1)
Fp24/IV	-	-	-	-	-	-	-	-	-	-	-	-	-	Y(1)	-

Response elements	**Binding Factor**	***S. lycopersicum* (HT.1)**	***S. pimpinellifolium***	***S. cheesmanii***	***S. habrochaites***	***S. parviflorum***	***S. pennellii***	**HT.2**	**HT.3**	**HT.4**	**HT.5**	**HT.6**	**HT.7**	**HT.8**	**HT.9**
GC-box	-	-	-	-	Y(0)	-	-	-	-	-	-	-	-	-	-
HD-ZIP BS	HD-ZIP	Y(1)	Y(1)	-	-	-	-	Y(1)	Y(1)	Y(1)	-	-	-	-	-
HSE	CPRF1	-	-	-	-	-	-	-	-	-	Y(2)	-	-	-	-
HSE3	tomato HsfA2	Y(2)	Y(2)	Y(2)	Y(2)	-	-	Y(2)	Y(2)	Y(2)	-	-	Y(2)	-	`-
HVH21	-	-	-	-	-	-	-	-	-	-	-	-	-	Y(0)	-
LBS/WBS1	LFU; WUS	-	-	-	-	-	-	-	-	-	-	-	-	Y(1)	-
LRE (activator)	-	Y(2)	Y(2)	Y(2)	-	-	-	Y(2)	Y(2)	Y(2)	Y(2)	-	-	-	-
poly(dA-dT) element	-	2Y(7, 7)	2Y(7, 7)	2Y(7, 7)	-	Y(7)	2Y(7, 7)	2Y(7, 7)	2Y(7, 7)	2Y(7, 7)	2Y(7, 7)	2Y(7, 7)	-	2Y(5, 6,6,7)	2Y(7, 7,7)
PRD motif 1	GmHdl56; GmHdl57	-	-	-	-	Y(0)	-	-	-	-	-	-	-	-	-
RE IV	-	Y(3)	Y(3)	-	-	-	-	Y(3)	Y(3)	Y(3)	-	-	-	-	-
S3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y(0)
S2	S2F	Y(0)	Y(0)	Y(0)	-	-	Y(0)	Y(0)	-	Y(0)	Y(0)	Y(0)	Y(0)	Y(0)	Y(0)
UN U2	-	Y(2)	Y(2)	Y(2)	-	Y(2)	Y(2)	Y(2)	Y(2)	Y(2)	Y(2)	Y(2)	-	-	-
VP1 RE	-	Y(1)	Y(1)	Y(1)	-	Y(1)	Y(1)	Y(1)	Y(1)	Y(1)	Y(1)		Y(1)	Y(1)	-
W-box 6	WRKY40	-	-	-	-	-	-	-	-	-	-	-	-	-	Y(2)
Zmhox1a BS 1	Zmhox1a	-	-	-	-	-	-	-	-	-	-	-	-	-	Y(2)
-314	-	-	-	-	-	-	-	-	-	-	-	-	-	Y(4)	Y(4)

-:unknown nuclear factor, absence of the element

Y:Presence of the element, Digit before Y indicates the number of the particular element predicted in the given sequence, Digit/s given in brackets indicates number of mismatches in the sequence of each element compared to that in the database, as predicted by the software.

Table 14: Functions of various response elements

Response elements	**Binding Factor**	**Function**
ABRE	unknown nuclear factor	ABA-responsive element (Straub *et al*., 1998)
ABRE3	bZIP67 + LEC1/L1L-[NF-YC2]	ABA-responsive element (Busk and Pages, 1995)
AC-II	unknown nuclear factor	Required for vascular-specific gene expression (Patzlaff *et al*., 2003)
AIV	unknown nuclear factor	Auxin responsive element (Walcher and Nemhauser, 2012)
Alfin1 BS2	Alfin1	Zinc-finger protein in alfalfa roots that binds to promoter elements in the salt-inducible MsPRP2 gene (Basto *et al*., 1998)
AT-2b	nuclear proteins from cotyledons	Transcription activator (Lagrange *et al*., 1998)
AT-rich H	unknown nuclear factor	Transcription activator (Forde, 1994)
box 1	GT-1	Involved in organ-specific expression in plant development (Villain *et al*., 1996)
box d	DOF1	Regulate expression of genes responsive to plant hormone and/or stress signals (Yanagisawa, 2000)
CArG (RIN-a)	RIN	Binding site for RIN protein which regulate many ripening related genes (Ito *et al.*, 2008)
C1-box	unknown nuclear factor	Transcription activator (Srinivasan and Kroos, 2004)
CBF-C BS	CBF-C	C-Repeat Binding Factors; low temperature response element (Jaglo-Ottosen, 1998)
C1BS2	C1	Response element present in the promoter of *a2f* gene of *Glycine max**
ERE1	unknown nuclear factor	Ethylene-responsive element (Montgomery *et al*., 1993)
ERE2	unknown nuclear factor	Ethylene-responsive element (Montgomery *et al*., 1993)
ERE3	unknown nuclear factor	Ethylene-responsive element (Montgomery *et al*., 1993)
Fp24/IV	unknown nuclear factor	Response element present in the promoter of *CHS8* gene of *Glycine max**

Response elements	Binding Factor	Function
GC-box	unknown nuclear factor	Transcription enhancer element (Klug *et al*., 2009)
HD-ZIP BS	HD-ZIP	Involved in stress response (Soderman *et al*., 1999)
HSE	CPRF1	Heat-shock responsive element (Rieping and Schoffl, 1992)
HSE3	tomato HsfA2	Heat-shock responsive element (Rieping and Schoffl, 1992)
HVH21	unknown nuclear factor	DNA binding sites recognised in vitro by a knotted class 1 homeodomain protein encoded by the hooded gene, k, in barley (Krusell *et al*., 1997)
LBS/WBS1	LFU; WUS	Enhancer of floral homeotic gene *AGAMOUS* (*AG*) (Hong *et al*., 2003)
LRE (activator)	unknown nuclear factor	Light response element (Lois *et al*., 1989)
poly(dA-dT) element	unknown nuclear factor	A ubiquitous promoter element that stimulates transcription (Iyer and Struhl, 1995)
PRD motif 1	GmHdl56; GmHdl57	Involved in phosphate response (Tang *et al*., 2001)
RE IV	unknown nuclear factor	Response element in the promoter of *ags* gene of *Agrobacterium tumefaciens* (Ti-plasmid)*
S3	unknown nuclear factor	Response element in the promoter of *FT* gene in Arabidopsis*
S2	S2F	Leaf-specific trans-acting factor (Lagrange *et al*., 1997)
UN U2	unknown nuclear factor	Response element in the promoter of *rbcS3C* gene in tomato*
VP1 RE	unknown nuclear factor	Transcriptional activator (Hattori *et al*., 1995)
W-box 6	WRKY40	Stress response element (Yamamoto *et al*., 2004)
Zmhox1a BS 1	Zmhox1a	Feedback control element (Bellmann and Werr, 1992)
-314	unknown nuclear factor	Critical positive cis element in sugar responsive gene expression (Kim and Guiltinan, 1999)

***http://www.softberry.com/berry.phtml?topic=regsitelist**

4.4. Effect of promoter variations in *CYCB* gene expression

In order to see the effect of these promoter variations on *CYCB* expression level, three highly different promoters were selected for further studies. These were WT (Arka vikas) promoter, *Beta* type promoter (*CYCB* promoter in the over expression mutant '*Beta*' which is similar to that in *S. habrochaites*) and *Chilense* type promoter (*CYCB* promoter with Nup fragment). To study the relative efficiency of these promoters, two different strategies were designed. One was based on relative abundance of the transcript of *CYCB* gene. The second strategy was based on transient expression analysis.

4.4.1. Comparison of the promoter efficiencies based on *CYCB* gene expression

The first strategy was to compare the relative efficiencies of the promoters by examining the *CYCB* expression level in fruit tissues of selected accessions carrying these promoters. Since detailed physiological maturity studies were not carried out in all the accessions, the fruit tissues were collected based on their chronological age. 30 DPA (Days Post Anthesis) stage, at which the maximum expression of the *CYCB* gene is reported in the reference cultivar (Arka Vikas), was selected in all the cases. In addition to that 60 DPA stage of *S. habrochaites* was also selected randomly. Quantitative real-time PCR was used to measure the relative expression of *CYCB* gene. In '*Beta*' mutant a five-fold increase in expression was observed compared to ArkaVikas. Similar higher gene expression was observed in the green fruited wild relative, *S. habrochaites*, which has *CYCB* promoter similar to that of '*Beta*' mutant. At 30 DPA stage, *CYCB* expression in *S. habrochaites* was also found to be more than two-fold of that of ArkaVikas, but less than that of '*Beta'* mutant. Interestingly, it was found to be more than eight-fold at 60 DPA

stage of *S. habrochaites*. Similarly, five-fold increase in expression was also observed in the B^{og} accession LA0348 carrying the '*Chilense* type' of promoter compared to its background, whereas, the level of gene expression in another accession of B^{og} (LA4026) with a WT promoter was much lower (Figure 21). Taking in consideration above results, it appears that the '*Beta* type' promoter and '*Chilense* type' promoters are more efficient compared to that of WT.

4.4.2. Comparison of the promoter efficiencies through transient expression studies

Comparing the promoter efficiency based on the expression data in different accessions carrying the promoter has many disadvantages. Firstly, the gene expression might be influenced by other background effects in each accession. Secondly, changes in the physiological age of the fruit vary between species and hence can influence the gene expression profile. In order to find out whether the difference in *CYCB* expression in different accessions and wild relatives of tomato is due to promoter variations, transient expression studies were carried out using promoter-reporter constructs.

In the present study a dual luciferase expression vector pGreenII 0800LUC (kindly provided by Dr. Roger Hellens, Plant and food research, New Zealand) with 'renilla luciferase' (*RLUC*) gene under Cauliflower Mosaic Virus (CaMV) 35S promoter (internal control) and 'firefly luciferase' (*FLUC*) gene with a multiple cloning site upstream, where the promoter of interest can be inserted was used (Figure 5). The promoter fragments of WT (886 bp), '*Beta* type' (933 bp) and '*Chilense* type' (1192 bp) were used to make three different constructs for transient expression studies (Figure 22A) (All the three fragments were amplified with the same set of primers. Insertions in the

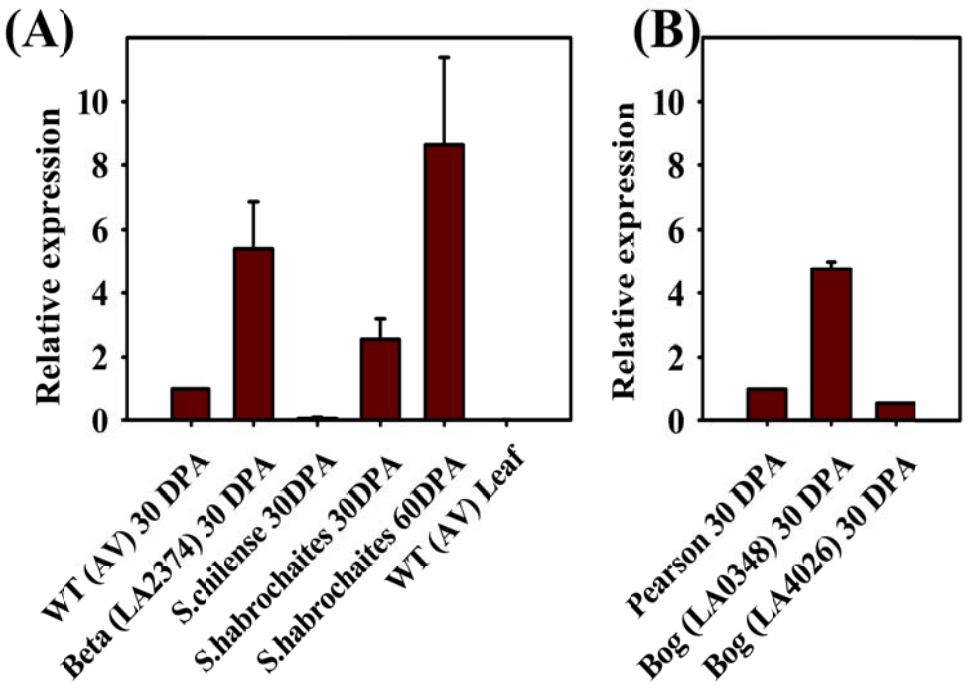


Figure 21: Relative expression of *CYCB* in fruit tissues of different accessions. Quantitative Real-Time PCR was used to measure the relative expression. WT (AV) was used as reference for calculation of relative expression in (A) and Pearson (LA0012) was used as reference in (B). The background of the B^{og} accession LA0348 is Pearson, whereas, the background of the B^{og} accession LA4026 is unknown (http://tgrc.ucdavis.edu). AV: Arka Vikas, DPA: Days Post Anthesis. TGRC accession numbers of *S. challenge* and *S. habrochaites* are LA1969 and LA1362 respectively. Actin was used as the internal control.

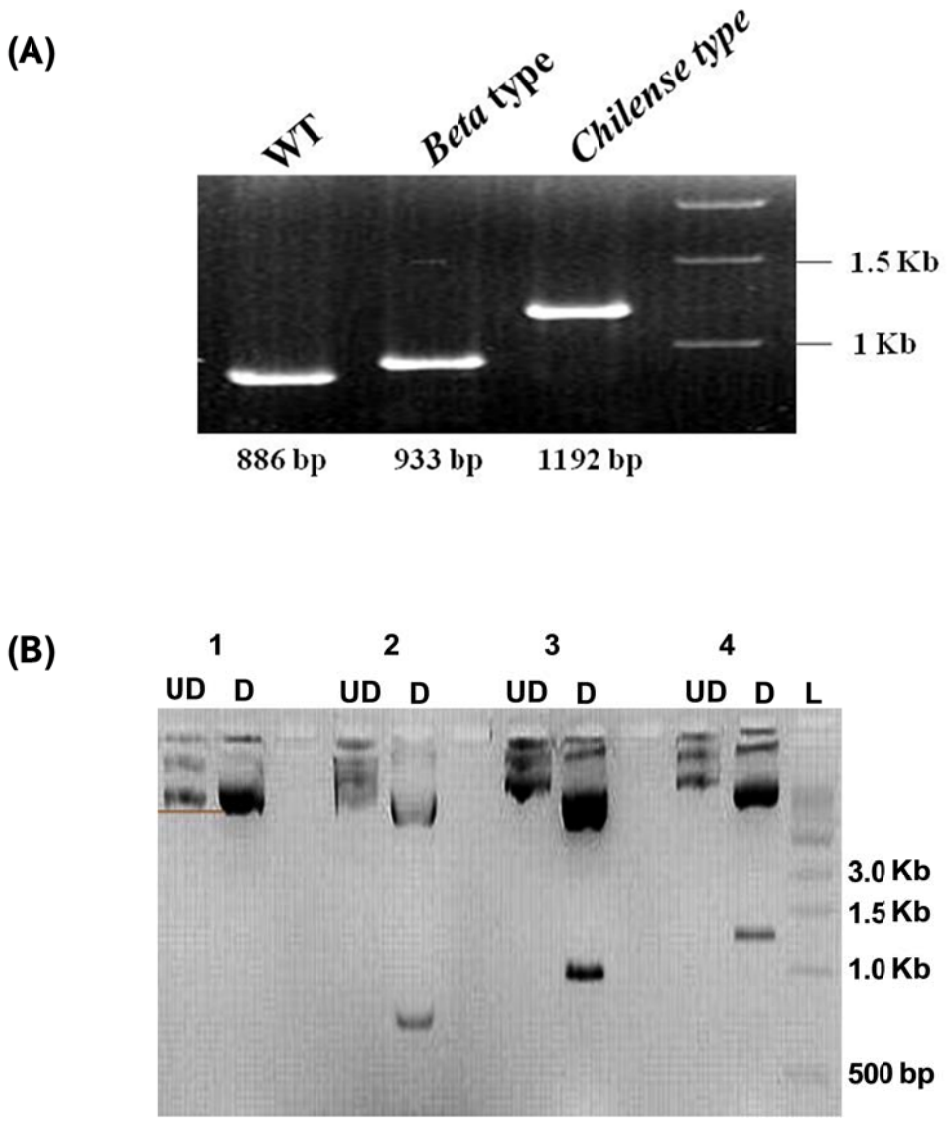


Figure 22: *CYCB* promoter fragments used for construct preparation (A). Promoter fragments amplified from WT (Arka Vikas), *Beta* mutant accession (LA2374) and B^{og} mutant accession (LA0348).

Bam HI and Hind III double digestion of pGREENII 0800-LUC carrying the *CYCB* promoters for *FLUC* (B). 1) pGREENII 0800-LUC vector, 2) pGREENII 0800-LUC containing WT promoter, 3) pGREENII 0800-LUC containing *Beta type* promoter, 4) pGREENII 0800-LUC containing *Chilense type* promoter. UD: Undigested, D: Double digested with Bam HI and Hind III, L: Molecular size ladder. Size of linearised pGREENII 0800-LUC vector is 7.2 Kb. Sizes of WT, *Beta type* and *Chilense type* inserts are 886 bp, 933 bp and 1192 bp respectively.

sequences made the size difference). All the constructs were confirmed by double digestion using Bam HI and Hind III (Figure 22B) and sequencing.

Transient transformation of the constructs was carried out by particle bombardment method into Arka Vikas fruit tissues harvested at 30 DPA stage and fruits were incubated for 24 hr. The promoter efficiencies were compared by measuring the *FLUC* transcript abundance by quantitative real-time PCR and by measuring the relative FLUC activity compared to that of RLUC. On comparing the *FLUC* transcript abundance, a 9-fold increase and 15-fold increase were observed in case of '*Beta* type' and '*Chilense* type' promoters respectively with respect to WT promoter (Figure 23A). This result indicated that both these promoters are more efficient than that of WT and that '*Chilense* type' promoter is more efficient than '*Beta* type'. Similar trend of relative efficiencies was also observed on comparing FLUC/RLUC ratios (Figure 23B). Based on these results it appears that the *CYCB* promoters in wild relatives are more efficient in driving the expression of *CYCB* gene in fruit tissue when compared to cultivated tomato.

4.4.3. Variability of RIN (Ripening INhibitor) binding sites in the *CYCB* promoters

In order to find out the basis for this relative higher efficiency, the promoter sequences were analyzed using PLACE (PLAnt Cis-acting regulatory DNA Elements) software (http://www.dna.affrc.go.jp/PLACE/), which showed the presence of a number of *cis*-elements varying in the type and number of repetitions. Among these elements an interesting difference was observed with respect to the binding sites of a transcription factor-RIN, which controls major events in fruit ripening like ethylene biosynthesis, perception and response, cell wall metabolism, carotenoid biosynthesis and regulation of additional ripening-related transcription factors (Vrebalov *et al*., 2002). It has been shown

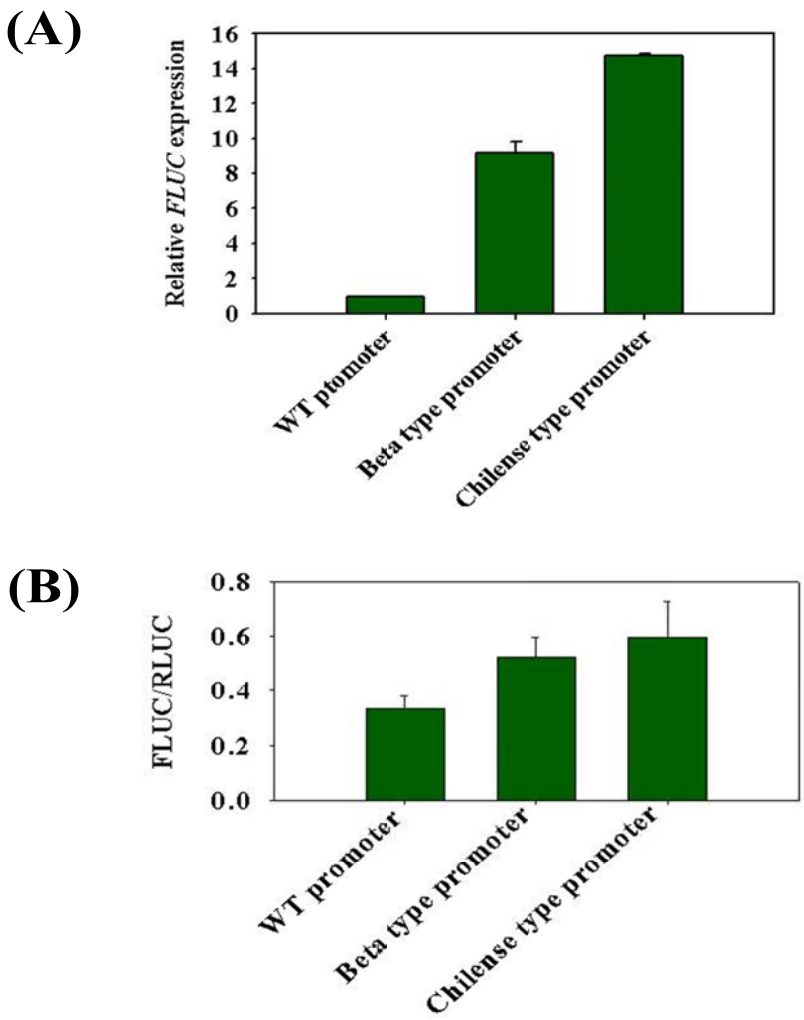


Figure 23: Relative expression of *FLUC* under different promoters in pGreenII 0800-LUC constructs (A). After transient transformation by particle bombardment, the fruit tissues (Arka Vikas at 30 DPA in all cases) were incubated on osmotic medium for 48 hrs and total RNA was extracted. *FLUC* transcripts were quantified by real-time PCR. Expression of actin was used as a control for normalization between samples. The *FLUC* expression in the empty vector pGreenII 0800-LUC (no promoter for *FLUC*) was used to normalize any background level expression. Relative *FLUC* expression under *Beta* type and *Chilense* type promoters were calculated with respect to the *FLUC* expression under WT promoter.

Relative activity of FLUC (Firefly luciferase) compared to RLUC (*Renilla* luciferase) (B). *RLUC* gene is under CaMV 35S promoter. Dual luciferase assay was performed with the lysates prepared from the transiently transformed fruit tissues, to measure the total activity of FLUC and RLUC for each sample. The ratio between FLUC and RLUC activity was taken to compare the relative FLUC expression. Background controls were run with the empty vector.

that RIN specifically binds to the DNA sequences C(C/T)(A/T)6(A/G)G known as 'CArG' box, and the preferable binding sequences start with CCA and ends with TAG (Ito *et al.*, 2008).

Both WT and '*Beta* type' promoter sequences carry four RIN-binding sites; out of them three have similar sequences (CAATATTTTG, CTTTTTTTTG and CCTTTTTTAG). These sequences are located at -98, -140 and -672 respectively in WT promoter and -97, -139 and -684 respectively in '*Beta* type' promoter. The fourth binding site which is located at -623 in WT and -633 in '*Beta* type' promoter, bears an SNP at the second position (CC/TTTTATAGG) (Figure 24). It is believed that the presence of 'C' at the second position ('*Beta* type') allows higher binding efficiency for RIN than 'T' (WT) (Ito *et al.*, 2008). Considering above, this change could be contributing to the relatively higher efficiency of '*Beta* type' promoter.

'*Chilense* type' promoter has all the four RIN binding sites similar to '*Beta* type' located at positions -97, -139, -891 and -940. In addition to that there are three binding sites in the *Nup* fragment. They are CTAAAATATG (-328), CTTTATATTG (-493) and CTATTATTTG (-510) (Figure 24). These additional RIN binding sites might be contributing to its comparatively higher efficiency than '*Beta* type' promoters. Based on the expression studies, '*Chilense* type' promoter could be identified as a gain of function allele. The original sources from which this promoter was identified i.e, *S. chilense* and B^{og} mutant accessions do not show the phenotypic effect in fruits perhaps due to absence of upregulation of carotenoid biosynthesis pathway during fruit ripening and an inactive CYCB enzyme respectively (Ronen *et al*., 2000). Nevertheless, '*Chilense* type' promoter is a probable candidate for breeding a high β-carotene containing tomato cultivar.

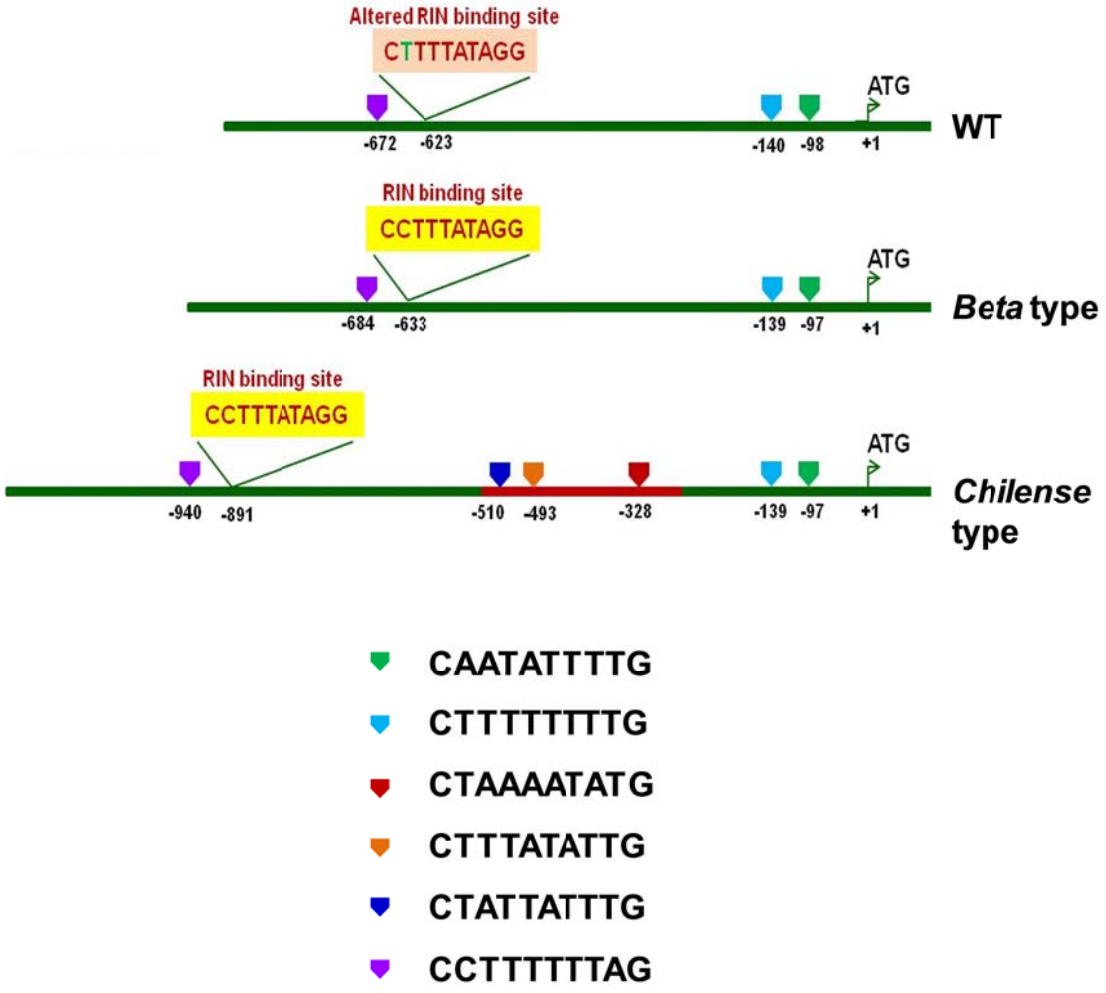


Figure 24: Variability of RIN TF binding sites in *CYCB* promoters. The nucleotide sequences upstream of *CYCB* gene in WT, *Beta* type and *Chilense* type are diagrammatically represented with green bars. The red colored portion in *Chilense* type represents the probable Nup fragment. The RIN binding sites are indicated with arrows above the bars and the exact positions of the first nucleotide in each case are mentioned below the bars. Different RIN binding sequences are indicated with different colors of the arrow. The RIN binding site with an SNP at the second position (T/C) located at positions -623, -633 and -891 in WT, *Beta* type and *Chilense* type promoters respectively might be responsible for the higher efficiency of *Beta* type promoter compared to that of WT. The three additional RIN binding sites identified in the Nup fragment indicated with red (-328), orange (-493) and dark blue (-510) arrows might be the probable cause for much higher efficiency of *Chilense* type promoter compared to that of *Beta* type .

4.5. Carotenoid content in various tomato accessions

Lycopene and *β*-carotene are the major carotenoid pigments in tomato fruit constituting around 95% of the total carotenoids. Hence, estimation of the total carotenoid content would give an idea about the lycopene and *β*-carotene content of a given accession. Total carotenoid content was estimated from the fruits of 183 tomato accessions which showed high variation among the accessions (Table 15). It varied from 833.00 (EC20639) to 22103.33 (EC241446) µg/100g fresh weight (FW) of the fruit tissue. EC241446 (NBPGR109) is a landrace from Israel with orange colored fruit. High amount of *β*-carotene along with other carotenoids like *α*-carotene, *δ*-carotene, lutein etc. might be contributing to the orange color of the fruit. However, the lycopene content might be very low in this accession. Highest frequency of accessions (30 (16.39%)) were having a total carotenoid content in the range of 5000-6000 µg/100g FW, followed by 27 (14.75%) accessions in the range of 6000-7000 µg/100g FW. The reference variety Arka vikas has a medium level (6704.12 µg/100g FW) of total carotenoid content. Only one accession each (0.55%) was having total carotenoid content <1000 µg/100g FW and >20,000 µg/100g FW category. Figure 25 shows graphical representation of frequency of the accessions belonging to different ranges of total carotenoid content.

4.5.1. Preliminary identification of desirable accessions

The biochemically estimated carotenoid contents of the fruits from all the accessions and the reference variety Arka Vikas were analyzed using different statistical methods to study the variability among the accessions and classification of the accessions

to select accessions of interest for further crop improvement programs for carotenoid rich tomatoes.

Several univariate statistical methods were used for studying the variability of the accessions in the population with respect to carotenoid content (Table 16). The mean value of total carotenoid content of all accessions is 6591.5 µg/100g FW, whereas, the standard deviation (SD) is 3450.14 µg/100g FW. Coefficient of variation (CV) (Francis and Kannenberg, 1978) which is the ratio of SD to the mean is irrespective of units of the parameters. It is an indication of the variance of a parameter and is 1.91 for total carotenoid content of the population.

Shakhatreh *et al*. (2010) defined a simple methodology for preliminary selection of desirable and undesirable accessions based on a combination of the mean and SD for each attribute. The accessions were classified into five different categories based on their total carotenoid content (Table 17). According to this classification majority of the accessions (74.32%) are with average performance in their total carotenoid content. 13.66% of the accessions are classified in the desirable range of total carotenoid content, out of which 4.92% are highly desirable. However, 12.02% of accessions are in the undesirable category and none of the accessions are classified in the highly undesirable range. The highly desirable accessions include landraces from Israel and Italy, *Beta old gold* mutant, an introgression line of *S. pennellii*, a non-hybrid cultivar from Greece etc. The details of the accessions classified under desirable and highly desirable category are given in Table 18. Interestingly, few accessions are with orange or reddish orange colored fruits. These accessions might be good sources of carotenoids like *β*-carotene, *α*-carotene,

δ-carotene, lutein etc. and can be used for crop improvement after identification of the major carotenoid in them.

Table 15: Total fruit carotenoid content of various tomato accessions. Desirable (*), highly desirable (**) and undesirable (#) accessions based on univariate statistical method described by Shakhatreh *et al*. (2010) are indicated in the last column. Absence of any symbol indicates accessions with average performance. Total carotenoid contents are expressed in µg/100g FW of fruit tissues.

Code No.	Accession	Total carotenoid	Category
-	Arka Vikas	6704.12	
NBPGR18	EC6192	5673.93	
NBPGR71	EC490141	8284.12	
NBPGR104	EC1177297	3919.92	
NBPGR109	EC241446	22103.33	**
NBPGR110	EC241446A	15698.18	**
NBPGR117	EC251649	7969.42	
NBPGR128	EC320583	6196.89	
NBPGR145	EC339066	5150.96	
NBPGR146	EC338717	9876.15	
NBPGR153	EC362941	6803.16	
NBPGR154	EC362949	7066.96	
NBPGR155	EC363863	13199.06	*
NBPGR167	EC369020	5530.46	
NBPGR172	EC398614	5012.12	
NBPGR173	EC398684	7821.32	
NBPGR174	EC398685	6506.97	
NBPGR176	EC398688	8122.14	
NBPGR178	EC398687	4623.37	
NBPGR181	EC398695	5109.31	
NBPGR182	EC398697	4637.26	
NBPGR184	EC398704	6303.34	
NBPGR185	EC398710	5442.53	
NBPGR186	EC398712	5072.29	
NBPGR188	EC398714	6391.27	
NBPGR189	EC398716	5544.34	
NBPGR190	EC398717	7529.76	
NBPGR196	EC458213	6955.88	
NBPGR219	EC520059	3744.05	
NBPGR224	EC521077	5326.83	
NBPGR229	EC521082	3179.44	#
NBPGR262	EC429	3595.96	
NBPGR298	EC3176-1	11199.76	*
NBPGR311	EC5888	5882.19	
NBPGR312	EC6053-1	6284.82	
NBPGR315	EC6875	2989.69	#
NBPGR319	EC7317	4493.79	

Code No.	Accession	Total carotenoid	Category
NBPGR327	EC10662	3406.21	
NBPGR331	EC12692	5072.29	
NBPGR333	EC13274	6965.14	
NBPGR334	EC13574	3887.52	
NBPGR336	EC13904	5252.78	
NBPGR338	EC15416	4827.00	
NBPGR339	EC16343	6460.69	
NBPGR340	EC16368	9061.62	
NBPGR343	EC20639	833.00	#
NBPGR346	EC27885	7534.38	
NBPGR347	EC25563	8210.00	
NBPGR351	EC26750-1	9339.30	
NBPGR352	EC27251	11607.02	*
NBPGR356	EC27995	7173.40	
NBPGR357	EC28356	1754.01	#
NBPGR359	EC29914	5586.00	
NBPGR361	EC29933	3730.00	
NBPGR362	EC29969	1573.52	#
NBPGR366	EC32019	3915.29	
NBPGR370	EC32287	7696.36	
NBPGR372	EC32557	8237.84	
NBPGR380	EC34480	3822.73	
NBPGR388	EC35252	5359.22	
NBPGR391	EC35272	8571.06	
NBPGR394	EC35322	3396.95	
NBPGR399	EC161645	5636.90	
NBPGR410	EC129602	2452.84	#
NBPGR411	EC129604	5419.39	
NBPGR413	EC144336A	2633.33	#
NBPGR416	IC447708	2244.58	#
NBPGR419	EC357828	9348.56	
NBPGR421	EC381554A	8501.64	
NBPGR424	EC490128	9691.03	
NBPGR427	EC490130	9959.46	
NBPGR435	IC469648	6275.57	
NBPGR437	IC469628	5437.90	
NBPGR438	IC469626	6511.60	
NBPGR439	IC469603	9107.90	
NBPGR441	EC57442	4493.79	
NBPGR443	IC469597	10394.49	*
NBPGR471	IC469714	13624.83	**
NBPGR479	EC531801	9635.50	
NBPGR480	EC529086	5604.51	

Code No.	Accession	Total carotenoid	Category
NBPGR483	EC531802	6914.23	
NBPGR484	EC398710	10375.98	*
NBPGR485	EC6486	12412.30	*
NBPGR486	EC6192	7228.94	
NBPGR490	EC521067B	5979.38	
NBPGR491	EC521068	4243.88	
NBPGR502	EC362958	6395.90	
NBPGR514	EC538139	7668.60	
NBPGR515	EC528362	12393.78	*
NBPGR562	EC2673	2850.85	#
IIVR5	WIR3957	11347.86	*
IIVR14	EC520076	8321.14	
IIVR17	WIR13717	11199.76	*
IIVR19	EC8372	3364.56	
IIVR22	Pusa Rohini	7876.86	
IIVR32	PDT-3-1	2540.77	#
IIVR33	TLBR-2	7728.76	
IIVR35	A.Arka	8274.86	
IIVR36	Riogrande	9080.14	
IIVR37	Castle Rock	6173.75	
IIVR38	KT-15	11542.23	*
IIVR39	Siberia	9468.89	
IIVR40	H-88-87	6432.92	
IIVR42	WIR-3928	7552.90	
IIVR43	EC520077	18400.93	**
IIVR46	EC50-50	3364.56	
IIVR47	EC3414425	7626.94	
IIVR49	LA4003	11718.10	*
IIVR53	LA3995	5724.84	
IIVR54	EC520079	1527.24	#
IIVR56	EC-528365	9043.11	
IIVR73	F-5070	4503.04	
IIVR75	LA3967	3628.35	
IIVR76	LA3971	5262.04	
IIVR78	LA4040	17771.52	**
IIVR86	LA3934	3998.59	
IIVR87	Agata-30	14624.48	**
IIVR89	BL-1208	6613.41	
IIVR98	CLN-2998	5132.45	
IIVR100	TLBR-12	14004.33	**
IIVR101	T.Local	4493.79	
IIVR105	Cerasiformae	2605.56	#
IIVR107	P.Pink	6497.71	

Code No.	Accession	Total carotenoid	Category
IIVR108	IIHR-2201	11810.66	*
IIVR112	Vaibhav	11579.26	*
IIVR114	Feb.4	3813.47	
IIVR115	T-HL	8136.02	
IIVR120	Sel-14	3091.50	#
IIVR128	EC193538	7937.02	
IIVR156	EC009046	3549.68	
IIVR174	EC531800	8228.58	
TGRC4	LA3538	5396.25	
TGRC13	LA3530	2911.01	#
TGRC14	LA3534	6497.71	
TGRC21	LA3539	9061.62	
TGRC22	LA3770	3864.38	
TGRC26	LA3537	3563.56	
TGRC35	LA4025	13828.46	**
TGRC36	LA1795	1786.41	#
TGRC39	LA0806	5609.14	
TGRC47	LA0500	6645.81	
TGRC55	LA0292	4919.56	
TGRC58	LA0276	6932.74	
TGRC59	LA3203	14624.48	**
TGRC60	LA2818	1874.34	#
TGRC61	LA2715	9098.65	
TGRC62	LA3632	11125.71	*
TGRC65	LA0533	4590.98	
TGRC66	LA3129	4415.11	
TGRC70	LA3630	3285.88	
TGRC72	LA0012	2119.62	#
TGRC77	LA3317	6974.40	
TGRC79	LA0502	11246.04	*
TGRC83	LA0180	9191.21	
TGRC84	LA1089	5650.79	
TGRC89	LA1506	6432.92	
TGRC92	LA0266	6719.86	
TGRC93	LA3231	4308.67	
TGRC95	LA1091	4646.51	
TGRC97	LA3903	3498.77	
TGRC99	LA3229	11514.46	*
TGRC101	LA1021	4493.79	
TGRC102	LA3237	4220.74	
TGRC104	LA0516	3720.91	
TGRC107	LA3234	4729.82	
TGRC108	LA3243	4988.98	

Code No.	Accession	Total carotenoid	Category
TGRC109	LA1090	1480.96	#
TGRC117	LA2400	6206.15	
TGRC118	LA0274	4604.86	
TGRC120	LA3554	5812.77	
TGRC121	LA0854	4220.74	
TGRC122	LA3579	4623.37	
TGRC124	LA0215	1351.38	#
TGRC127	LA1016	1170.88	#
TGRC128	LA2999	6071.94	
TGRC129	LA3430	4928.82	
TGRC133	LA3905	2915.64	#
TGRC135	LA3024	5734.09	
TGRC136	LA4104	2202.93	#
TGRC140	LA2921	5155.59	
TGRC141	LA1996	5252.78	
TGRC142	LA1088	6224.66	
TGRC143	LA0744	5548.97	

* >Mean+SD

** >Mean+2SD

<Mean+SD

all others are in the range between Mean±SD

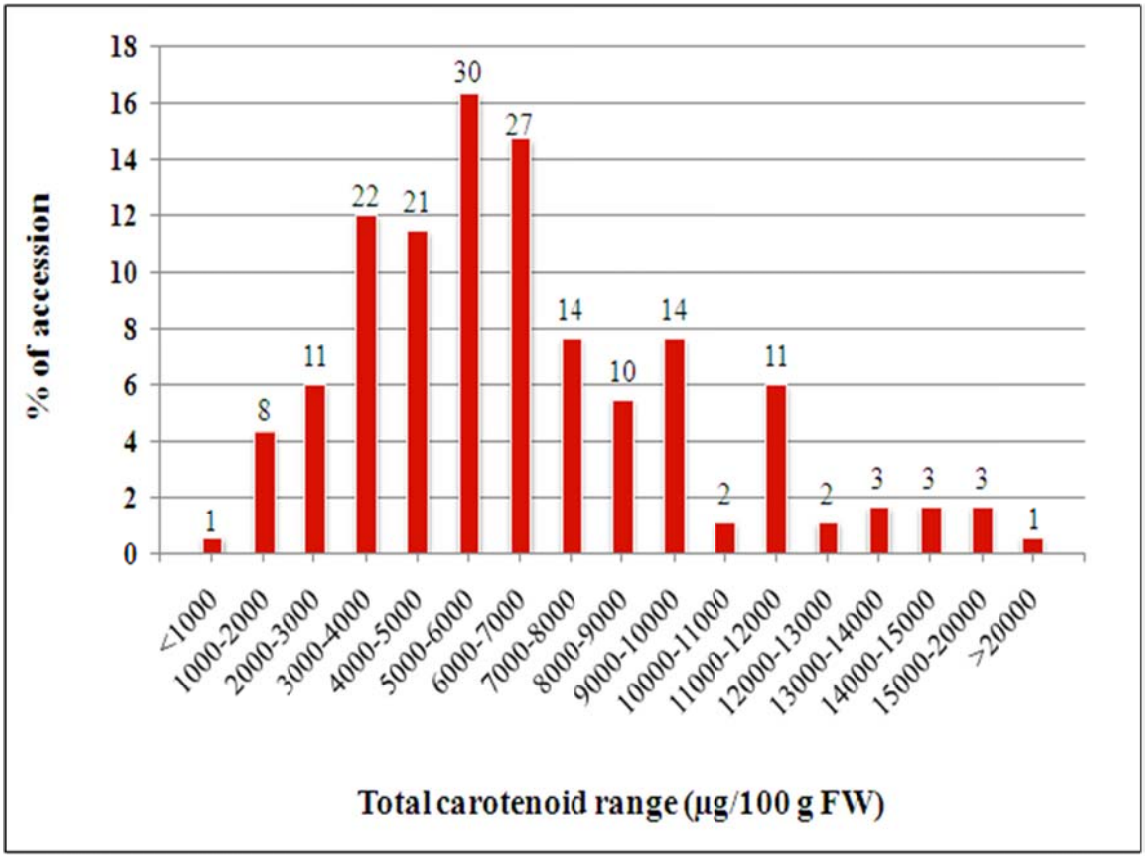


Figure 25: Frequency of accessions belonging to different ranges of total fruit carotenoid content. A method based on mean absorption wavelength and mean absorption coefficients (Biehler *et al*., 2010) was used for estimation. Total carotenoid content was calculated with A_{451} values of the samples as described by (Rodriguez-Amaya and Kimura, 2004).

Table 16: Various univariate statistical parameters and their combinations calculated from total fruit carotenoid content of 183 tomato accessions.

Parameter	Value
Minimum[$]	833.00
Maximum[$]	22103.33
Mean[$]	6591.50
SD[$]	3450.14
CV	1.91
Mean+SD[$]	10041.64
Mean+2SD[$]	13491.79
Mean-SD[$]	3141.36
Mean-2SD[$]	-308.79

[$]µg/100g FW

Table 17: Number of tomato accessions belonging to various statistical categories described by Shakhatreh *et al*. (2010).

Category	**Description**	**Number of accessions**
Between Mean±SD	Average performance	136 (74.32%)
>Mean+SD	Desirable	25 (13.66%)
>Mean+2SD	Highly desirable	9 (4.92%)
<Mean-SD	Undesirable	22 (12.02%)
<Mean-2SD	Highly undesirable	0

Table 18: List of accessions with high total fruit carotenoid content and their details

Sl No.	Accession	Category	Details
1	EC241446	**	M-2495/-; Landrace from Israel; Orange fruit
2	EC241446A	**	M-2495/-; Landrace from Israel; Orange fruit
3	IC469714	**	NA
4	LA4025	**	*Bog* (*Beta old gold*), *L.esculentum*
5	LA3203	**	*L.esculentum* cv. Large Plum; Landrace from Italy
6	LA4040	**	*S. pennellii* introgression line IL 2-5
7	Agata-30	**	Agata, Non-hybrid cultivar of Greece
8	TLBR-12	**	Cultivar with combined resistance to tomato leaf curl virus and bacterial wilt
9	EC520077	**	NA
10	EC363863	*	NA
11	EC3176-1	*	Orangish red colored fruits
12	EC27251	*	Vantage, Landrace from Canada; Orangish red colored fruits
13	IC469597	*	Q161Natti, Traditional variety from Karnataka, India
14	EC6486	*	1271-91 HopeXThatcher3XHindi62X4/W. 307; Landrace from Egypt; Orangish red colored fruits
15	EC398710	*	NA
16	EC528362	*	NA
17	LA3632	*	*L.esculentum* cv. Start 24
18	LA0502	*	*L.esculentum* cv. Marglobe
19	LA3229	*	*L.esculentum* cv. Prospero
20	WIR3957	*	*S. peruvianum*
21	LA4003	*	*S. lycopersicum* - *S. habrochaites* (ex. LA1777) backcross recombinant inbred line in E6203
22	IIHR-2201	*	Water stress tolerent cultivar
23	Vaibhav	*	Tomato Leaf Curl Virus Tolerant Hybrid
24	WIR13717	*	NA
25	KT-15	*	NA

**Highly desirable level of total carotenoid content
*Desirable level of total carotenoid content
NA: Not available

Chapter 5

Discussion

5.1. *CYCB* homologues show high functional diversity in plants

Carotenoids are colored compounds which are highly essential for the living organism and are responsible for contributing many of the beautiful colors displayed by organisms in nature (Britton, 1995). They are synthesized by organisms belonging to the kingdoms Monera, Protista, Fungi and Plantae. Generally organisms under kingdom Animalia do not possess the genes for carotenoid biosynthesis, but they can absorb and accumulate the carotenoids from food. An example of horizontal gene transfer from fungi to pea aphid is reported (Moran and Jarvik, 2010). The carotenoid biosynthesis pathway starts from IPP (isopentenyl pyrophosphate) which is a product of the MEP pathway localized in plastids. The first committed precursor for carotenoid biosynthesis is phytoene. Several oxygenated (xanthophylls) and anoxygenated (carotenes) forms of carotenoids are synthesized as intermediates in the pathway. The deep red colour of ripened tomato is due to accumulation of lycopene which contributes around 90% of the total carotenoids, whereas, *β*-carotene occupies the next position and contributes around 5% of total carotenoids (Kopsell and Kopsell, 2006).

The various enzymatic steps in the carotenoid pathway and the genes coding for them are well elucidated in tomato. Moreover, detailed studies on the chemical conversions of the enzymes have also been carried out in different plant species. The ratio of lycopene to *β*-carotene is determined by the chromoplast specific enzyme lycopene beta cyclase (CYCB) which expresses at very high level during ripening, while the level of expression of its chloroplast counterpart (LCYB) decreases (Ronen *et al*., 2000). Both the enzymes introduce two *β*-ionone rings in the lycopene molecule (Gross, 1991), but they vary in their tissue specificity.

Homology search from NCBI database using CYCB of tomato (AK327886.1) showed similar sequences of lycopene cyclases in various species starting from bacteria. They include other functionally different enzymes like lycopene epsilon cylcase (LCYE) and capsanthin capsorubin synthase (CCS). LCYE introduces one epsilon ring in the lycopene molecule to synthesize *δ*-carotene, whereas, CCS converts antheraxanthin and violaxanthin to capsanthin and capsorubin respectively. It was found that the amino acid sequences of these enzymes are significantly conserved among different plant genera.

The genes coding for LCYB, CYCB, LCYE and CCS present in plants appear to have originated from an ancestral lycopene cyclase gene through gene duplications and evolved to have specialized functions (Bouvier *et al*., 2000). In all the cases the maximum sequence divergence was observed in the N-terminal region of the protein sequence partly due to high divergence in the signal peptide region. Phylogenetic analysis of CYCB homologues showed three major groups. One group consisted mostly of LCYB; another group consisted mostly of LCYE while the third group consisted of enzymes like CYCB and CCS. The CYCB lineage appears to be diverged from the LCYB lineage. A retrotransposon mediated event might have led to the origin of *CYCB* and *CCS* from an ancestral *LCYB* gene as they all lack introns in their sequence unlike *LCYB* (Ronen *et al*., 2000). It is also possible that the *CYCB* lineage might have evolved novel functions like CCS in pepper. CYCB of tomato shows highest similarity with CCS (pepper). The synteny between the locus of *CYCB* and *CCS* (Thorup *et al*., 2000) indicate that in pepper this gene might have evolved to acquire a specialized function with different substrate specificities than that of CYCB. It is interesting to observe that in spite of high sequence divergence, LCYB and CYCB are doing the same function while CYCB and CCS which

share high sequence similarity, are doing different functions. It will be interesting to find how these enzymes maintain their substrate specificity in spite of very high sequence similarity.

It is reported that eventhough there is a decrease in the *LCYB* gene expression and increase in the *CYCB* gene expression in the ripening tomato fruits, the quantity of *LCYB* transcripts are equal to or more than that of *CYCB* transcripts. However, it has been proven that in ripening fruits CYCB plays the prominent role in conversion of lycopene to *β*-carotene (Ronen *et al*., 2000). The high CYCB activity could also be due to post-transcriptional and post-translational regulations of *CYCB*. Moreover, the strong plastid targeting signal of CYCB compared to that of LCYB could contribute to the functional specialization of CYCB in chromoplastic tissues. This indicates the importance of high divergence in the signal peptide sequences compared to the rest of the amino acid sequences in defining their functional specialization.

5.2. Hypervariability in the promoter sequence of *CYCB*

Lycopene and *β*–carotene as essential nutrients are well studied for their beneficiary roles in human body including antioxidant, anticancerous and provitamin A activities (Bendich, 1993). There have been several successful transgenic approaches as a part of crop improvement in various parts of the world to increase the quantity of these carotenoids in different crop plants (Mann *et al*., 2000; Dharmapuri *et al*., 2002; Ralley *et al*., 2004; Davuluri *et al*., 2005). A transgenic tomato with lycopene beta-cyclase (t*Lcy-b*) cDNA driven by the *CaMV 35S* promoter named as 'HighCaro' (HC) produced orange fruits as a result of the complete conversion of lycopene to *β*-carotene (D'Ambrosio *et al*., 2004). The orange coloration and the high *β*-carotene in fruits of the tomato mutant '*Beta*'

results from increased expression of the *CYCB* gene. This increased expression basically arises from changes in the upstream sequences of *CYCB* gene and is attributed to introgression of promoter from wild relatives. Deletions in the *CYCB* gene resulted in the two loss of function mutants '*old-gold'*(B^{og}) and '*old-gold-crimson'* (B^{c}) with an increase in the lycopene content (Ronen *et al*., 2000). Thus mutation in *CYCB* can lead to enhanced lycopene in loss of function situation and more *β*-carotene in gain of the function situation in chromoplast.

Reverse genetic techniques are based on specific variants of gene of interest and they aims to identify function of the gene by correlating the genetic variations with the phenotypic effects, while complementing the classical forward genetics. In this study the genetic variations in *CYCB* gene and its promoter region were examined in a collection of wild relatives of tomato and among tomato accessions from various resources. Since SNPs contribute a major part of genetic variations, the accessions were screened for SNPs using a modified strategy of TILLING known as EcoTILLING (Comai *et al*., 2004). TILLING and EcoTILLING methods are cost effective and can explore functions of several genes of interest by screening for artificially induced random mutations or for naturally occurring genetic variations. Unlike transgenic methods they are not subjected to GMO regulations. In addition to its advantage of high throughput screening, multiple SNPs in an amplicon can also be detected through EcoTILLING since CELI digest only a proportion of the heteroduplexes at a single position.

While the promoter regions has high variation in the form of SNPs with an SNP frequency of 1/1.45 Kb and In-Dels, the coding sequences has very low SNP frequency of 1/7.11 Kb. This result is expected since the accumulation of SNPs and In-Dels in the

coding region is usually not tolerated by the plant because of the drastic effects compared to those in the non-coding region. Based on the number and type of SNPs and In-Dels detected in various accessions, different haplotypes were identified. Eight different haplotypes were identified for the coding region, whereas, for the promoter region nine different haplotypes were identified.

Among the changes, in general transitions have happened with much higher frequency than transversions. This is expected and could be explained by the difference in the molecular mechanism by which they are generated. Interchange of the purine rings (A↔G) happened with maximum frequency. This conversion is a major mutation as amino form of A is known to convert temporarily to imino tautomer of A which pairs with C and in the next round of replication C will get paired with G, resulting in A to G conversion.

The genetic variations observed in the promoter region of CYCB gene in various accessions may or may not have effect on gene expression. While it is apt to examine how these changes in promoter affects expression of *CYCB* gene, it was not found to be feasible to analyze the effect of each of these changes in gene expression as it would take considerable time. As an alternative, the promoter analysis using online softwares was carried out for all the haplotypes and sequences from wild relatives. The analysis revealed that certain elements like ABRE (ABA response element), box1 and boxd (light response elements) and ERE2 and ERE3 (ethylene response elements) were present in all the analyzed sequences indicating their essentiality in the regulation of the pathway. However, the presence of ERE1 (ethylene response element 1), AIV (auxin response element) and LRE (light response element) was exclusively in the colored

fruited species. Considering that fruit ripening is a highly regulated process in colored fruited species, it is likely that these elements might be able to drastically enhance the tomato fruit ripening process. Antisense lines of LeERF1(*Lycopersicum esculentum* ethylene response factor 1), which is a part of binding factor of ERE1, exhibited longer shelf life (Li *et al*., 2007), indicating the crucial role of EREs in inducing fruit ripening. The genetic variations in *CYCB* promoter during the course of evolution might have led to the loss of these elements from the green fruited species because of its non-essentiality or might have led to the gain of this element in the colored fruited species and got selected due to its key role in the fruit ripening process.

The effect of genetic variations in the coding region can vary depending on the type and location of the variation. Nucleotide substitutions may or may not lead to amino acid changes. The change in amino acids for an enzyme may affect substrate binding, initiation of enzyme action, catalytic function or play role in structural integrity etc. Conversely the introduction of stop codons leads to truncation of the enzyme with associated loss of function. Thus the effect of a mutation can vary from no effect on enzymatic activity to a completely nonfunctional enzyme. Online software: PARSESNP (Project Aligned Related Sequences and Evaluate SNPs) (http://www.proweb.org/parsesnp/) was used to analyze the effects of the polymorphisms on the expressed gene product of *CYCB*. It gives result in the form of PSSM score and SIFT score. A PSSM score of >10 means the mutation has deleterious effect while a SIFT score of <0.05 also means deleterious effects of the mutation. The PSSM scores for most SNPs in *CYCB* gene were very low and so these changes were predicted as non-deleterious. There were only two SNPs- G406A and G868A that were predicted as

deleterious based on their SIFT scores (0.00). These mutations may affect the normal functions of the CYCB protein and thus need to be studied further by heterologous expression in *E.coli*. The absence of SNPs in majority of the colored fruited accessions and the less deleterious nature of the SNPs indicate that the *CYCB* gene is under stabilizing selection.

A combined phylogram was prepared based on the nucleotide sequences of *CYCB* from different wild relatives and accessions to see their phylogenic relation. A few of the colored fruited accessions and all the green fruited accessions have higher accumulation of SNPs. Interestingly the pattern of SNPs observed in these accessions were the same as that of the SNPs observed in the green fruited wild relatives. These accessions clustered together with the wild relatives of tomato in the phylogram. It is likely that most probably these accessions have acquired above SNPs through hybridization with wild relatives. In green fruited accessions and wild relatives there is either nil or very little accumulation of carotenoids indicating the absence of upregulation of carotenoid biosynthesis pathway during fruit ripening. Since a fully functional CYCB enzyme in these wild relatives does not impart any evolutionary advantage as they have a carotenoid biosynthesis pathway, where, upregulation requires other regulatory factors controlling ripening, most of the genetic variations are tolerated and this might have led to the accumulation of SNPs in this gene.

5.3. Hints of probable positive translational regulations in colored fruited species of tomato

Regulation of gene expression can occurs at different stages like transcriptional, post-transcriptional, translational and post-translational regulations. There are several

reports on the influence of synonymous codons on translational efficiency (Shao *et al*., 2012; Qian *et al*., 2012). For example, codon optimization enhanced the expression of *MIR (Miraculin)* gene in tomato (Hiwasa-Tanase *et al*., 2011). Most of the base substitutions identified in the *CYCB* gene lead to silent mutations (synonymous/nonsynonymous ratio: 1.7) and thus expected to have no influence on the protein. Nevertheless, even in the case of SNPs resulting in silent mutation (synonymous codons), a probable change in the translational efficiency can be expected due to a difference in the abundance of various tRNA molecules in cell. A comparison of the codon usage efficiencies between different silent mutations observed in CYCB gene of different tomato accessions, using information from codon usage database indicated a lot of variation. The change in the codon usage efficiencies might be altering the overall level of CYCB enzyme and thus may affect the rate of conversion of lycopene to β-carotene.

Few of the synonymous changes were identified exclusively in the color fruited accessions, causing an increase in the codon usage efficiency compared to the green fruited accession or wild relatives. There are two possibilities which may have caused this: 1) The progenitor species may have possessed the efficient codons, however, due to lack of selection pressure, the green fruited accessions tolerated the changes into less efficient codons after the divergence of green fruited species from colored fruited species. On the other hand, the colored fruited ones maintained the efficient codons due to positive selection pressure. 2) Alternatively the progenitor may have possessed less efficient codons which were sustained by the green fruited species due to lack of selection pressure, whereas, positive selection pressure on the colored fruited species allowed selection and maintenance of the more efficient codons. Among these two possibilities,

the presence of multiple less efficient codons in many species of tomato wild relatives strongly supports the second possibility which indicates a less efficient codon usage in the progenitor of these tomato wild relatives. Thus it may indicate an increase in the translational efficiency of *CYCB* gene in colored fruited species compared to the green fruited ones. However, it should also be taken in consideration that codon usage efficiencies may differ in other species.

5.4. A probable loss of function allele of *CYCB* gene identified through EcoTILLING

Eventhough a general pattern of non-deleterious changes were identified in the *CYCB* gene, a higher density of SNPs was observed in case of green fruited accessions compared to the colored fruited ones. Out of them, EC20636 showed the maximum number of genetic variations, most of the SNPs were with no expected effect on the protein. However, a base change G570A causing the conversion of the codon TGG (codes for tryptophan) into TGA (stop codon) would prematurely terminate the protein. Therefore, above SNP would cause a truncation in the enzyme after synthesis of 189 amino acids. Mialoundama *et al*., (2010) reported that 293-FLEET-297 motif in CYCB is highly conserved. They found that the substitution of alanine, lysine, and arginine for glutamate-295 in this motif abolishes the enzyme activity. Therefore, it is ensured that the truncated CYCB would be enzymatically inactive as truncated protein possesses only initial 189 amino acids. Since the above accession is a green fruited one, where the carotenoid biosynthetic pathway is not upregulated during fruit ripening, the functional role of this allele on the fruit phenotype is not discernible. Notwithstanding above, this loss of function allele is a promising candidate for developing a high lycopene containing cultivar and hence could act as a donor plant for cross breeding with commercially grown

cultivars. Moreover, these SNPs are present in a homozygous condition in the accession EC20636 and thus it could be more conveniently used for any crossing experiments in future.

5.5. Introduction of the Nup (Nuclear localized plastid) fragment resulted in a probable gain of function allele of *CYCB* promoter

Chloroplast and mitochondria of eukaryotic cells are believed to have evolved from endosymbiotic organisms (Martin *et al*., 2002). There are several reports on the shuffling of genome fragments between organelles and nucleous (Woischnik and Moraes, 2002; Yuan *et al*., 2002). Many of these exchanged gene fragments can have new functions and create new genes, silence already existing genes or modify the gene expression patterns and thus play key roles in evolution. Upon screening the promoter region of *CYCB* gene, one insertion with a size of 250 bp was detected in the accessions LA0348, LA0458 (accessions of B^{og} mutant) and the wild relative *S. chilense*. This insertion was then identified as a fragment of *rps4 (ribosomal protein small subunit 4)* gene of the tomato chloroplast genome by NCBI BLAST search. The literature search revealed that the source of B^{og} was a plant from the F_2 progenies of a cross between *S. chilense and S. lycopersicum. Old gold* phenotype has been transferred to the latter species by nine successive backcrosses (Rick and Smith, 1953; Rick and Dempsey, 1961). Therefore, it is likely that the *Nup* DNA fragment might have first got integrated into the nuclear genome in *S. chilense* and later got introgressed into *S.lycopersicum* during crossing.

Gene expression studies in selected accessions showed that '*Beta* type promoter' (promoter in the overexpression mutant '*Beta*' which is similar to that of *S. habrochaites*)

and '*Chilense* type promoter' (promoter with Nup fragment) are much efficient as compared to WT *CYCB* promoter. However, comparing the promoter efficiency based on the expression data in different accessions carrying the promoter has many disadvantages. Firstly, the gene expression might be influenced by other background effects in each accession. Secondly, changes in the physiological age can influence the gene expression profile (fruit stages were selected based on their chronological age since exact fruit maturation and ripening stages has not been identified in green fruited accessions). Thus there was a need of a more reliable method of comparison of promoter efficiencies. This was achieved through transient expression analysis of the three different types of promoters using promoter-reporter constructs with a dual luciferase expression vector pGreenII 0800LUC. On comparing the *FLUC* transcript abundance, a 9-fold increase and 15-fold increase were observed in case of '*Beta* type' and '*Chilense* type' promoters respectively compared with the WT promoter, indicating that both these promoters are more efficient than that of WT and that '*Chilense* type' promoter is more efficient than the '*Beta* type'. Similar trend of relative efficiencies was also observed on comparing FLUC/RLUC ratios. Based on these results it appears that the *CYCB* promoters in wild relatives are more efficient than that of cultivated tomato. Moreover, integration of the Nup fragment appears to be contributing to the increased efficiency of the *CYCB* promoter in '*Chilense* type'.

When the promoter elements were compared in all the three promoters, an interesting difference was observed with respect to the binding sites of a transcription factor-RIN (Ripening INhibitor), which controls major events in fruit ripening like ethylene biosynthesis, perception and response, cell wall metabolism, carotenoid

biosynthesis and regulation of additional ripening-related transcription factors (Vrebalov *et al*., 2002). It has been shown that RIN specifically binds to the DNA sequences C(C/T)(A/T)6(A/G)G known as 'CArG' boxes The preferable binding sequences of RIN start with CCA and ends with TAG (Ito *et al.*, 2008). Both WT and '*Beta* type' promoter sequences carry four RIN-binding sites; out of them three have similar sequences. While the beginning of fourth CArG box in '*Beta* type' promoter has the most preferred "CC", the WT promoter has "CT". This change might be contributing to the relatively higher efficiency of '*Beta* type' promoter.

It is of interest to note that '*Chilense* type' promoter had all the four RIN binding sites similar to '*Beta* type' and three additional RIN binding sites in the *Nup* fragment. These extra binding sites might be responsible for the higher efficiency of this promoter compared to '*Beta* type' promoter. Considering above, '*Chilense* type' promoter can be considered as a gain of function allele. It appears that *S. chilense* and B^{og} mutant accessions do not show the phenotypic effect in fruits due to absence of upregulation of carotenoid biosynthetic pathway in case of *S. chilense* and an inactive CYCB enzyme in case of B^{og} mutant. Nevertheless, '*Chilense* type' promoter is a probable candidate for breeding a high *β*-carotene containing tomato cultivar.

5.6. Population of tomato accessions is a good source for crop improvement for carotenoid content

Lycopene and *β*-carotene are the major carotenoid pigments in tomato constituting around 95% of the total carotenoids. Knowledge about the carotenoids in various accessions will be useful in crop improvement. The biochemical estimation shows that 13.66% of the accessions have very high total carotenoid content (>10000µg/100gfresh

weight). Only one accession (EC20639) is with very low total carotenoid content (833 µg/100gfresh weight), whereas, another accession (EC241446) has very high total carotenoid content (22103.33 µg/100gfresh weight).

Preliminary selection of desirable and undesirable accessions was done by a simple methodology based on a combination of the mean and SD as defined by Shakhatreh *et al*. (2010). They defined five different categories of accessions. Those accessions with value of the parameter of interest > (mean + SD) and > (mean + 2SD) can be identified as desirable and highly desirable accessions respectively, whereas, those with values < (mean – SD) and < (mean – 2SD) can be identified as undesirable and highly undesirable ones. All others i.e, those with values between mean ± SD are average in their performance. Around 75% of the accessions have average level of total carotenoid content including the reference variety Arka vikas. The frequency of desirable and undesirable accessions is almost equal in the population. However, none of the accessions are highly undesirable for the total carotenoid content. Around 5% of the accessions are in the highly desirable category, which include a land race from Israel, non-hybrid cultivar of Greece, B^{og} mutant accession, an introgression line of *S. pennellii* etc. Most of the accessions with desirable level of total carotenoid content are exotic collections. Another introgression line of *S. habrochaites* is also with desirable level of total carotenoid. Interestingly, many of them are with orange colored fruits, which might be due to carotenoids like *β*-carotene, *α*-carotene, δ-carotene, lutein etc. In summary, a number of accessions that could act as a good source for breeding for high carotenoid content were identified from the population.

5.7. Why domesticated tomato is red in colour and not orange?

The promoter of *CYCB* might have evolved such that its expression is induced during fruit ripening. In '*Beta*' mutant of tomato *CYCB* gene is highly induced upon fruit ripening giving rise to orange colored tomato. The promoter region of '*Beta*' mutant is similar to that of wild relatives of tomato like *S. habrochaites* and *S. pennellii* indicating that in the progenitor of tomato *CYCB* was evolved to express highly during the fruit ripening stage. It is hypothesized that a change in the promoter sequence of *CYCB* gene had occurred in the lineage that gave rise to tomato (*S. esculentum*) such that the expression level of *CYCB* gene got reduced during fruit ripening resulting in accumulation of lycopene and red colored fruit. This hypothesis is diagrammatically shown and explained in Figure 26.

It is likely that the presence of an SNP in the RIN binding site in wild type tomato (*S. lycopersicum*) may be responsible for red colour of fruit (lycopene) instead of orange color (*β*-carotene). The assumption can be checked by using a promoter-reporter assay and finding out whether by restoring the RIN binding site (by changing a single base pair) in the *S. lycopersicum* promoter could lead to increase in the promoter activity. In case the promoter activity is increased, then the question of whether it is indeed due to RIN binding could be addressed by carrying out the promoter-reporter assay in a RIN mutant fruit.

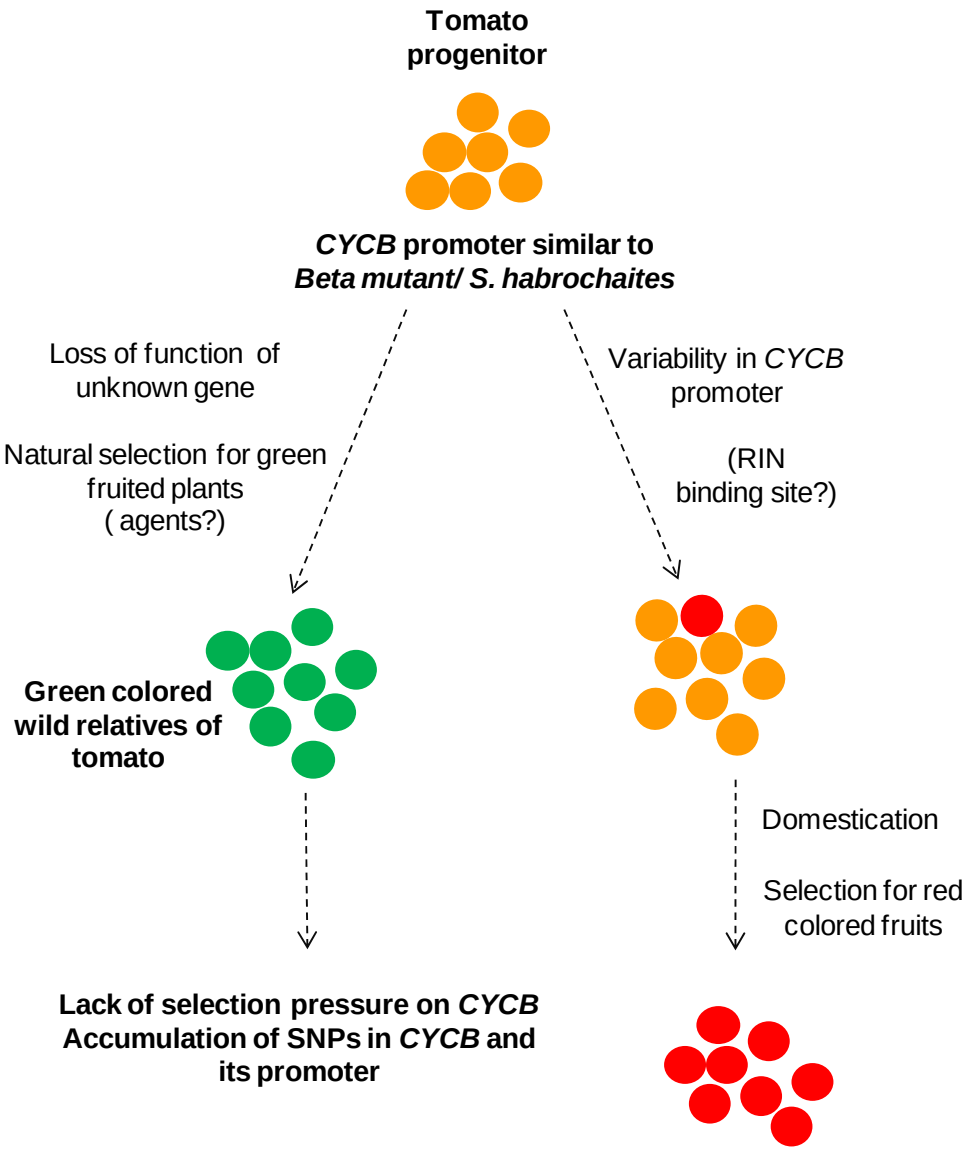


Figure 26: A hypothesis explaining the red color of domesticated tomato rather than orange. The progenitor of domesticated tomato was orange in colour due to proportionally very high amount of β-carotene (orange color) than lycopene (red color). Loss of function mutations in some unknown gene controlling the pathway might have lead to an inefficient carotenoid biosynthesis pathway and conversion of chloroplasts to chromoplasts. Thus the fruits remain green even though it is ripened. There was a natural selection for the green fruited species by some unknown natural agents. In green fruited species a lack of selection pressure on *CYCB* and its promoter lead to accumulation of large number of SNPs. Another probable loss of function mutation in the *CYCB* promoter (may be a variation in RIN binding site) made it less efficient which drastically decreased the amount of CYCB enzyme. Thus the conversion of lycopene to β-carotene got reduced to a much lower level leading to the accumulation of lycopene and red color of the fruit. During domestication, there was a preference for red color rather than orange.

Chapter 6

Summary

Tomato is a widely cultivated and consumed vegetable throughout the world. Its dramatic color changes during ripening and high content of nutritionally important carotenoids like lycopene and *β*-carotene makes it a suitable model system to study about carotenoid biosynthetic pathway. CYCB is the key enzyme that maintains the balance between lycopene and *β*-carotene in tomato fruits.

To study the diversity of *CYCB* gene and its homologues in different plant genera, the amino acid sequences of various homologues were compared. They include enzymes like LCYB (Lycopene beta cyclase, chloroplastic), LCYE (Lycopene epsilon cyclase) and CCS (Capsanthin capsorubin synthase). Phylogenetic analysis of CYCB homologues showed three major groups. One group consists mostly of LCYB; another group consists mostly of LCYE while the third group consists of enzymes like CYCB and CCS. It shows that the amino acid sequences of these enzymes are significantly conserved among different plant genera. The CYCB lineage appears to be diverged from the LCYB lineage and evolved novel functions in different plants like CCS in pepper. CYCB of tomato shows highest similarity with CCS (pepper) and least with LCYE (tomato). Eventhough CYCB and LCYB have the same function, in different tissues; they are highly diverged in their amino acid sequences. The transit peptide region shows high divergence between the three different groups and the transit peptide of CYCB group shows higher plastid targeting scores, probably indicating more effective targeting of this protein towards plastids.

The genetic variations in *CYCB* gene and its promoter region were studied in a collection of tomato accessions from various resources. In addition, few wild relatives

of tomato were also used to examine the diversity. Since SNPs contribute a major part of genetic variations, the accessions were screened for SNPs using a high throughput reverse genetics strategy known as EcoTILLING. An Indian cultivar of *Solanum lycopersicum*, Arka Vikas (Sel 22), was used as a reference variety. Since the mis-match cleavage enzyme, CELI digests only a small proportion of the heteroduplexes at a single position, multiple SNPs in an amplicon can also be detected. The exact location and type of SNPs were identified by sequencing the amplicons.

Seven different haplotypes were identified for the coding region of *CYCB* with an SNP frequency of 1/7.11 Kb. The synonymous/nonsynonymous ratio is 1.7 and the transition/transversion ratio is 5.0. Online software: PARSESNP (Project Aligned Related Sequences and Evaluate SNPs) was used to analyze the effects of the polymorphisms on the expressed gene product. For each amino acid change, the change in their properties like side chain polarity, side chain charge and hydropathy index were studied. A base change G570A identified in the accession EC20636 would cause a truncation in the enzyme after 189 amino acids. Since the above accession is a green fruited one, where the carotenoid biosynthetic pathway is not upregulated in fruit ripening, the functional role of this allele on the fruit phenotype is not discernible. Nevertheless, this loss of function allele is a promising candidate for developing a high lycopene containing cultivar through breeding.

The *CYCB* gene from various wild relatives of tomato was sequenced and a combined phylogram was prepared along with the *CYCB* sequences of various haplotypes using 'GeneBee' software (http://www.genebee.msu.su/services/phtree_reduced.html). The low allele frequencies observed in this gene in colored fruited accessions indicate that

this gene is under stabilizing selection, whereas, green fruited wild relatives of tomato and other accessions showed high amount of variation in *CYCB* locus possibly because of lack of any selection pressure at this locus. Comparison of the codon usage efficiencies in the case of synonymous changes indicates a probability of increase in translational efficiency in the colored fruited species.

The promoter region of *CYCB* showed higher genetic variability between various tomato accessions in the form of SNPs, insertions and deletions. Nine different haplotypes were identified with an SNP frequency of 1/1.45 Kb. A combined phylogram was prepared based on the nucleotide sequences from different haplotypes and wild relatives. The above phylogram revealed the similarity of few haplotypes with the wild relatives indicating that those accessions might have obtained the variations through hybridizations.

Presence of a Nup (Nuclear localized plastid) fragment (250 bp) was observed in the *CYCB* promoter sequences of the accessions LA0348 and LA0500 (accessions carrying the *old-gold* mutation) and the wild relative *S. chilense*. Nucleotide BLAST in NCBI database showed similarity of the sequence to a fragment of *rps4 (ribosomal protein small subunit 4)* gene of the tomato chloroplast genome. It is likely that the Nup DNA fragment might have first got integrated into the nuclear genome in *S. chilense* and later got introgressed into *S.lycopersicum* during crossing.

In order to see the effect of these promoter variations on *CYCB* expression level, three highly different promoters were selected for further studies. These were WT promoter, *Beta* type promoter (promoter in the overexpression mutant '*Beta*' which is similar to that of *S. habrochaites*) and *Chilense* type promoter (promoter with Nup

fragment). The first strategy was to compare the relative efficiencies of the promoters by examining the *CYCB* expression level in fruit tissues of selected accessions carrying these promoters. Quantitative real-time PCR showed higher expression of *CYCB* gene in *Beta* mutant, B^{og} mutant and *S. habrochaites* compared to WT.

In order to find out whether the difference in *CYCB* expression in different accessions and wild relatives of tomato is due to promoter variations, transient expression studies were carried out using promoter-reporter constructs. A dual luciferase expression vector pGreenII 0800LUC with 'renilla luciferase' (*RLUC*) gene under CaMV35S promoter (internal control) and 'firefly luciferase' (*FLUC*) gene with a multiple cloning site upstream, where the promoter of interest can be inserted was used. The promoter fragments of WT, '*Beta* type' and '*Chilense* type' were used to make three different constructs and were transiently transformed to tomato fruit tissues. *FLUC* transcript abundance by quantitative real-time PCR and FLUC/RLUC ratios showed that both *Beta* type and *Chilense* type promoters are efficient than that of WT and that *Chilense* type promoter is more efficient than *Beta* type. It also says that the *CYCB* promoter in wild relatives is more efficient than that of cultivated tomato.

In order to find out the basis for this relative higher efficiency, the promoter elements of these sequences as predicted by PLACE (PLAnt Cis-acting regulatory DNA Elements) software (http://www.dna.affrc.go.jp/PLACE/) were compared. An interesting difference was observed with respect to the binding sites of a transcription factor-RIN (Ripening INhibitor) known as 'CArG' box (Ito *et.al*, 2008). Both WT and '*Beta* type' promoter sequences carry four RIN-binding sites; out of them three have similar sequences, whereas, the fourth binding site bears an SNP at the second position (C/T). It

is believed that the presence of 'C' at the second position ('*Beta* type') allows higher binding efficiency for RIN than 'T' (WT) (Ito *et al*., 2008). Considering above, this change might be the basis for the relatively higher efficiency of '*Beta* type' promoter. On the other hand '*Chilense* type' promoter has all the four RIN binding sites similar to '*Beta* type', and three additional binding sites in the Nup fragment, which might contribute to its comparatively higher efficiency. So *Chilense* type promoter was identified as a gain of function allele. The original sources from which this promoter was identified i.e, *S. chilense* and *Bog* mutant accessions do not show the phenotypic effect in fruits perhaps due to absence of upregulation of carotenoid biosynthetic pathway and/or an inactive CYCB enzyme respectively. Nevertheless, *Chilense* type' promoter is a probable candidate for breeding a high *β*-carotene containing tomato cultivar.

It is likely that the presence of an SNP in the RIN binding site in wild type tomato (*S. lycopersicum*) may be responsible for red color of fruit (lycopene) instead of orange color (*β*-carotene). The assumption can be checked by using a promoter-reporter assay whether by restoring the RIN binding site (by changing a single base pair) in the *S. lycopersicum* promoter could lead to increase in the promoter activity. In case the promoter activity is increased, then the question of whether it is due to RIN binding could be addressed by carrying out the promoter-reporter assay in a RIN mutant fruit.

References

Aggelis A, John I, Karvouni Z, Grierson D (1997) Characterization of two cDNA clones for mRNAs expressed during ripening of melon (*Cucumis melo* L) fruits. *Plant Mol Biol* 1997 33:313-322.

Ahrazem O, Rubio-Moraga A, Lopez RC, Gomez-Gomez L (2010) The expression of a chromoplast-specific lycopene beta cyclase gene is involved in the high production of saffron's apocarotenoid precursors. *Journal of Experimental Botany* 61:105–119.

Alba R, Cordonnier-Pratt M, Pratt LH (2000) Fruit-Localized Phytochromes Regulate Lycopene Accumulation Independently of Ethylene Production in Tomato1. *Plant Physiol*. 123(1): 363–370.

Al-Babili S, Hugueney P, Schledz M, Welsch R, Frohnmeyer H, Laule O, Beyer P (2000) Identification of a novel gene coding for neoxanthin synthase from *Solanum tuberosum*. *FEBS Letters* 485:168–172.

Alque´zar B, Zacarı´as L, Rodrigo MJ (2009) Molecular and functional characterization of a novel chromoplast-specific lycopene *β*-cyclase from Citrus and its relation to lycopene accumulation. *Journal of Experimental Botany* 60(6):1783–1797.

Apel W, Bock R (2009) Enhancement of Carotenoid Biosynthesis in Transplastomic Tomatoes by Induced Lycopene-to-Provitamin A Conversion DOI:10.1104/pp.109.140533.

Araya-Garay JM, Feijoo-Siota L, Veiga-Crespo P, Villa TG (2011) cDNA cloning of a novel gene codifying for the enzyme lycopene *β*-cyclase from *Ficus carica* and its expression in *Escherichia coli*. *Appl Microbiol Biotechnol* 92:769–777.

Azari R, Tadmor Y, Meir A, Reuveni M, Evenor D, Nahon S, Shlomo H, Chen L,Levin I (2010) Light signaling genes and their manipulation towards modulation of phytonutrient content in tomato fruits. *Biotechnol Adv* 28:108–118.

Babu TA, Sharmila V (2010) Vitamin A supplementation in late pregnancy can decrease the incidence of bronchopulmonary dysplasia in newborns. *J Matern Fetal Neonatal Med.* 23(12):1468–1469.

Bai L, Kim EH, DellaPenna D, Brutnell TP (2009) Novel lycopene epsilon cyclase activities in maize revealed through perturbation of carotenoid biosynthesis. *The Plant Journal* 59:588–599.

Bang H, Kim S, Leskovar D, King S (2007) Development of a codominant CAPS marker for allelic selection between canary yellow and red watermelon based on SNP in lycopene beta-cyclase (*LCYB*) gene. *Molecular Breeding* 20:63–72.

Barone A, Chiusano ML, Ercolano MR, Giuliano G, Grandillo S, Frusciante L (2008) Structural and Functional Genomics of Tomato. *International Journal of Plant Genomics*. doi:10.1155/2008/820274.

Bartley GE, Scolnik PA, Beyer P (1999) TwoArabidopsis thaliana carotene desaturases, phytoene desaturase and zeta-carotene desaturase, expressed in Escherichia coli, catalyse a poly-cis pathway to yield pro-lycopene. *European Journal of Biochemistry* 259:396–403.

Bauchet G, Causse M (2012) Genetic Diversity in Tomato (*Solanum lycopersicum*) and Its Wild Relatives, Genetic Diversity in Plants, Prof. Mahmut Caliskan (Ed.).

Beecher GR (1998) Nutrient content of tomatoes and tomato products. *Proc Soc Exp Biol Med*. 218(2):98-100.

Bendich A (1993) Biological functions of dietary carotenoids. *Ann NY Acad Sci* 691:61–67.

Biehler E, Mayer F, Hoffmann L, Krause E, Bohn T (2010) Comparison of 3 Spectrophotometric Methods for Carotenoid Determination in Frequently Consumed Fruits and Vegetables. *Journal of Food Science* 75:1.

Biodiversity-International (2006) "Crop wild relatives." 25.

Blas AL, Ming R, Liu ZY, Veatch OJ, Paull RE, *et al.* (2010) Cloning of the Papaya Chromoplast-Specific Lycopene beta-Cyclase, CpCYCB, Controlling Fruit Flesh Color Reveals Conserved Microsynteny and a Recombination Hot Spot. Plant Physiology 152:2013–2022.

Boileau TWM, Boileau AC, Erdman JW (2002) Bioavailability of alltrans and cis-isomers of lycopene. *Exp. Biol. Med*. 227:914-919.

Bouvier F, D'Harlingue A, Backhaus RA, Kumagai MH, Camara B (2000) Identification of neoxanthin synthase as a carotenoid cyclase paralog. *Eur J Biochem* 267:6346–6352.

Bouvier F, Hugueney P, d'Harlingue A, Kuntz M, Camara B (1994) Xanthophyll biosynthesis in chromoplasts: isolation and molecular cloning of an enzyme catalyzing the conversion of 5,6-epoxycarotenoid into ketocarotenoid. *Plant J* 6:45-54.

Bouvier F, Isner JC, Dogbo O, Camara B (2005) Oxidative tailoring of carotenoids: a prospect towards novel functions in plants. *Trends in Plant Science* 10:187–194.

Bramley P (1997) The regulation and genetic manipulation of carotenoid biosynthesis in tomato fruit. *Pure Appl Chem* 69:2159-2162.

Britton G. (1995) UV/visible spectroscopy. In Britton G, Liaaen-Jensen S, Pfander H (eds), carotenoids: Spectroscopy, Birkhäuser Verlag, Basel 1B:13-63.

Burbidge A, Grieve TM, Jackson A, Thompson A, McCarty DR, Taylor IB (1999) Characterization of the ABA-deficient tomato mutant *notabilis* and its relationship with maize *Vp14*. *The Plant Journal* 17(4):427-31.

Burton GW, Ingold KU (1984) beta-Carotene: an unusual type of lipid antioxidant. *Science* 224:569–573.

Cazzonelli CI, Cuttriss AJ, Cossetto SB, Pye W, Crisp P, Whelan J, Finnegan EJ, Turnbull C, Pogson BJ (2009a) Regulation of carotenoid composition and shoot branching in Arabidopsis by a chromatin modifying histone methyltransferase, SDG8. *Plant Cell* 21: 39-53.

Cazzonelli CI, Millar T, Finnegan EJ, Pogson BJ (2009*b*) Promoting gene expression in plants by permissive histone lysine methylation. *Plant Signalling & Behaviour* 4:484–488.

Cazzonelli CI, Yin K, Pogson BJ (2009c) Potential implications for epigenetic regulation of carotenoid biosynthesis during root and shoot development. *Plant Signalling & Behaviour* 4, 339–341.

Chen Y, Li F, Wurtzel ET (2010) Isolation and characterisation of the *Z-ISO* gene encoding a missing component of carotenoid biosynthesis in plants. *Plant Physiology* 153:66–79.

Chetelat RT, Ji Y (2007) Cytogenetics and evolution. *In* M. K. Razdan and A.

K.Mattoo, eds. Genetic improvement of Solanaceous crops. *Science Publishers*, *Enfield, NH*. 77–112.

Clagett-Dame M, DeLuca HF (2002) The role of vitamin a in mammalian reproduction and embryonic development. *Annu Rev Nutr.* 22:347–381.

Cohen JD (1996) *In vitro* tomato fruit cultures demonstrate a role for indole-3-acetic acid in regulating fruit ripening. *J. Am. Soc. Hortic. Sci*. 121:520-524.

Colbert T, Till BJ, Tompa R, Reynolds S, Steine MN, Yeung AT, McCallum CM, Comai L, Henikoff S (2001) High-throughput screening for induced point mutations. *Plant Physiol*. 126:480-484.

Comai L, Young K, Till BJ, Reynolds SH, Greene EA (2004) Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. *Plant J.* 37(5):778–786.

Cooper JL, Till BJ, Laport RG, Darlow MC, Kleffner JM, *et al.* (2008) TILLING to detect induced mutations in soybean. *BMC Plant Biol* 8:9.

Cseri A, Cserhati M, von Korff M, Nagy B, Horvath GV, *et al.* (2011) Allele mining and haplotype discovery in barley candidate genes for drought tolerance. *Euphytica* 181: 341–356.

Cunningham FX, Gantt E (2001) One ring or two? Determination of ring □number in carotenoids by lycopene *β*-cyclases. *Proc Natl Acad Sci USA* 98:2905-2910.

Cuttriss AJ, Chubb A, Alawady A, Grimm B, Pogson B (2007) Regulation of lutein biosynthesis and prolamellar body formation in *Arabidopsis*. *Functional Plant Biology* 34:663–672.

D'Ambrosio C, Giorio G, Marino I, Merendino A, Petrozza A, Salfi L, Stigliani AL, Cellini F (2004) Virtually complete conversion of lycopene into beta-carotene in fruits of tomato plants transformed with the tomato lycopene *β*-cyclase (*tlcy-b*) cDNA. *Plant Sci* 166:207–214.

da Silva DJH, Abreu FB, Caliman FRB, Antonio AC, Patel VB (2008) Chapter 1: Tomatoes: Origin, Cultivation Techniques and Germplasm Resources, Tomatoes and Tomato Products Nutritional, Medicinal and Therapeutic Properties, Victor R. Preedy and Ronald R. Watson (Ed.), *Science Publishers* 3-25.

Davuluri GR, van Tuinen A, Fraser PD, Manfredonia A, Newman R, Burgess D, Brummell DA, King SR, Palys J, Uhlig J, Bramley PM, Pennings HM, Bowler C (2005) Fruit-specific RNAi-mediated suppression of DET1 enhances carotenoid and flavonoid content in tomatoes. *Nat Biotechnol* 23:890–895.

de CandolleAP (1882). L'origine des plantes cultivées. Paris, Germer Baillière, VIII-379.

De Pascale S, Maggio A, Fogliano V, Ambrosino P, Ritieni A (2001) Irrigation with saline water improves carotenoids content and antioxidant activity of tomato. *J. Hortic. Sci. Biotechnol.* 76:447-453.

de Vos RCH, Hall RD, *et al.* (2011) Metabolomics of a Model Fruit: Tomato. *Annual Plant Reviews* 109-155.

Devitt LC, Fanning K, Dietzgen RG, Holton TA (2010) Isolation and functional characterization of a lycopene *β*-cyclase gene that controls fruit colour of papaya (*Carica papaya* L.). *Journal of Experimental Botany* 61(1):33–39.

Dharmapuri S, Rosati C, Pallara P, Aquilani R, Bouvier F, Camara B, Giuliano G (2002) Metabolic engineering of xanthophyll content in tomato fruits. *FEBS Lett* 22:30–34.

Ding Y, Avramova Z, Fromm M (2011) The *Arabidopsis* trithorax-like factor ATX1 functions in dehydration stress responses via ABA-dependent and ABA-independent pathways. *The Plant Journal* 66:735–744.

Dinneny JR, Weigel D, Yanofsky MF (2005) A genetic framework for fruit patterning in *Arabidopsis thaliana*. *Development*, 132:4687–4696.

Diretto G, Tavazza R, Welsch R, Pizzichini D, Mourgues F, Papacchioli V, Beyer P, Giuliano G (2006) Metabolic engineering of potato tuber carotenoids through tuber-specific silencing of lycopene epsilon cyclase. *BMC Plant Biology* 6:13.

Dorais M, Gosselin A, Papadopoulos AP (2001) Greenhouse tomato fruit quality. *Hortic. Rev*. 26:239-306.

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

Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP, and related tools. *Nature Protocols* 2:953-971.

Emmanuel E, Levy AA (2002) Tomato mutants as tools for functional genomics. *Curr. Biol.* 5:112–117.

Enfissi EMA, Barneche F, Ahmed I, Lichtle C, Gerrish C, McQuinn RP, Giovannoni JJ, Lopez-Juez E, Bowler C, Bramley PM, Fraser PD (2010) Integrative transcript and metabolite analysis of nutritionally enhanced DE-ETIOLATED1 downregulated tomato fruit. *Plant Cell* 22:1190–1215.

Fanasca S, Colla G, Maiani G, Venneria E, Rouphael Y, Azzini E, Saccardo F (2006) Changes in antioxidant content of tomato fruits in response to cultivar and nutrient solution composition. *J. Agric. Food Chem*. 54:4319-4325.

Fang J, Chai CL, Qian Q, Li CL, Tang JY, *et al*. (2008) Mutations of genes in synthesis of the carotenoid precursors of ABA lead to pre-harvest sprouting and photo-oxidation in rice. *Plant Journal* 54: 177–189.

FAO (2010) Plant genetic ressource for food and agriculture. *Rome, Food and Agriculture Organisation of the United Nations.*

Fei Z, Joung J, Tang X, Zheng Y, Huang M, Lee JM, McQuinn R, Tieman DM, Alba R, Klee HJ, Giovannoni JJ (2011) Tomato Functional Genomics Database: a comprehensive resource and analysis package for tomato functional genomics. *Nucleic Acids Research* 39.

Fei Z, Tang X, Alba R, Giovannoni J (2006) Tomato Expression Database (TED): a suite of data presentation and analysis tools. *Nucleic Acids Res*. 34: D766–D770.

Fernie AR, Klee H (2011) "The use of natural genetic diversity in the understanding of metabolic organization and regulation." *Frontiers in Plant Science* 2.

Fernie AR, Schauer N (2009) "Metabolomics-assisted breeding: a viable option for crop improvement?" *Trends in Genetics* 25(1):39-48.

Finer JJ, Vain P, Jones MW, McMullen MD (1992) Development of the particle inflow gun for DNA delivery to plant cells. *Plant Cell Rep*. 11(7):323-8.

Finkers R, de Weerd H, *et al.* (2011) *BreeDB:* A database suppporting quantitative aspects of plant breeding. *Plant and Animal Genomes XIX, San Diego*.

Francis TR, Kannenberg LW (1978) Yield stability studied in short-season maize. I. A descriptive method for grouping genotypes. *Can J Plant Sci*. 58:1029–1034.

Fraser PD, Enfissi EM, Halket JM, Truesdale MR, Yu D, Gerrish C, Bramley PM

(2007) Manipulation of phytoene levels in tomato fruit: effects on isoprenoids, plastids and intermediary metabolism. *The Plant Cell* 19:3194–3211.

Fraser PD, Truesdale MR, Bird CR, Schuch W, Bramley PM (1994) Carotenoid biosynthesis during tomato fruit development. *Plant Physiol*. 105:405-413.

Fuentes P, Pizarro L, Moreno JC, Handford M, Rodriguez-Concepcion M, *et al.* (2012) Light-dependent changes in plastid differentiation influence carotenoid gene expression and accumulation in carrot roots. *Plant Molecular Biology* 79:47–59.

Fujisawa M, Nakano T, Ito Y (2011) Identification of potential target genes for the tomato fruit-ripening regulator RIN by chromatin Immunoprecipitation. *BMC Plant Biology* 11:26.

Fusari CM, Lia VV, Hopp HE, Heinz RA, Paniego NB (2008) Identification of single nucleotide polymorphisms and analysis of linkage disequilibrium in sunflower elite inbred lines using the candidate gene approach. *BMC Plant Biol* 8:7.

Gady A, Hermans F, *et al.* (2009) "Implementation of two high through-put techniques in a novel application: detecting point mutations in large EMS mutated plant populations." *Plant Methods* 5(1):13.

Galpaz N, Wang Q, Menda N, Zamir D, 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.

Garcia-Closas R, Berenguer A, Tormo MJ, Sanchez MJ, Quiors JR (2004) Dietary sources of vitamin C, vitamin E and specific carotenoids in Spain. *Brit. J. Nutr*. 91:1005-1011

Gärtner C, Stahl W, Sies H (1997). Lycopene is more bioavailable from tomato paste than from fresh tomatoes. *Am. J. Clin. Nutr*. 66:116-122.

Gerster H (1997) The potential role of lycopene for human health. *J Am Coll Nutr* 16:109–126.

Gilchrist EJ, Haughn GW, Ying CC, Otto SP, Zhuang J, Cheung D, Hamberger B, Aboutorabi F, Kalynyak T, Johnson L, Bohlmann J, Ellis BE, Douglas CJ, Cronk QCB (2006) Use of Ecotilling as an efficient SNP discovery tool to survey genetic variation in wild populations of *Populus trichocarpa*. *Molecular Ecology* 15: 1367–1378.

Giordano LB, Silva JBC (2000) Tomatoes for industrial processing . Brasília:DF: Embrapa Communication for Transfer of Technology 168.

Giovannoni JJ (2007) Fruit ripening mutants yield insights into ripening control. *Curr. Opin. Plant Biol.* 10:283–289.

Giuliano G, Tavazza R, Diretto G, Beyer P, Taylor MA (2008) Metabolic engineering of carotenoid biosynthesis in plants.*Trends in Biotechnology* 26:139–145.

Gray J, Picton S, Shabbeer J, Schuch W, Grierson D (1992) "Molecular biology of fruit ripening and its manipulation with antisense genes," *Plant Molecular Biology* 19(1): 69–87.

Greene EA, Codomo CA, Taylor NE, Henikoff JG, Till BJ, Reynolds SH, Enns LC, Burtner C, Johnson JE, Odden AR, Comai L, Henikoff S (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. *Genetics* 164:731–740.

Gross J (1991) Pigments in Vegetables. van Nordstrand Reinhold, New York, USA.

Gupta SK, Sharma S, Santisree P, Kilambi HV, Appenroth K, Sreelakshmi Y, Sharma R (2014) Complex and shifting interactions of phytochromes regulate fruit development in tomato. *Plant Cell Environ*. 37(7):1688-702.

Guzman I, Hamby S, Romero J, Bosland PW, O'Connell MA (2010) Variability of Carotenoid Biosynthesis in Orange Colored *Capsicum* spp. *Plant Sci* 179(1-2):49–59.

Ha SH, Kim JB, Park JS, Lee SW, Cho KJ (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.

Havaux M, Niyogi KK (1999) The violaxanthin cycle protects plants from photo-oxidative damage by more than one mechanism. Proc Natl Acad Sci *USA* 96:8762–8767.

Hedden Walter MH, Floss DS, Strack D (2010) Apocarotenoids: hormones, mycorrhizal metabolites and aroma volatiles. *Planta* 232:1–17.

Hildering GJ, Verkerk K (1965) Chimeric structure of the tomato plant after seed treatment with Ems and X-rays. In: *The Use of Induced Mutations in Plant Breeding*. Pergamon Press, Oxford 317–320.

Hiwasa-Tanase K, Nyarubona M, Hirai T, Kato K, Ichikawa T, Ezura H (2011) High-level accumulation of recombinant miraculin protein in transgenic tomatoes expressing a synthetic miraculin gene with optimized codon usage terminated by the native miraculin terminator**.** *Plant Cell Rep* **30**(1)**:**113-124.

http://faostat.fao.org/

http://linux1.softberry.com/berry.phtml?topic=nsitep&group=programs&subgroup=promoter

http://multalin.toulouse.inra.fr/multalin/

http://tgrc.uctavis.edu

http://tilling.ucdavis.edu/index.php/Tomato_Tilling; ;

http://www .sgn.cornell.edu

http://www.ateronon.com/

http://www.dna.affrc.go.jp/PLACE/

http://www.ebi.ac.uk/Tools/msa/clustalw2/

http://www.proweb.org/parsesnp/

http://Zamir.sgn.cornell.edu/mutants/

www.ars.usda.gov

www.avrdc.org

www.tgrc.ucdavis.edu

Hugueney P, Badillo A, Chen H C, Klein A, Hirschberg J, Camara B, Kuntz M (1995) Metabolism of cyclic carotenoids: a model for the alteration of this biosynthetic pathway in *Capsicum annuum* chromoplasts. *Plant J* 8:417-424.

Ikoma Y, Komatsu A, Kita M, Ogawa K, Omura M, Yano M, Moriguchi T (2001) Expression of a phytoene synthase gene and characteristic carotenoid accumulation during citrus fruit development. *Physiol Plant* 111:232-238.

Inbaraj S, Chen BH (2008) Chapter 7: Carotenoids in Tomato Plants, Tomatoes and Tomato Products Nutritional, Medicinal and Therapeutic Properties, Victor R. Preedy and Ronald R. Watson (Ed.), *Science Publishers* 133-164.

Inoue H, Nojima H, Okayama H (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene* 96 (1):23–28.

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

Itkin M, Seybold H, Breitel D, Rogachev I, Meir S, 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, Kasumi T (2008). DNA-binding specificity, transcriptional activation potential, and the *rin* mutation effect for the tomato fruit-ripening regulator RIN. *Plant J.* 55:778–786.

Jackson H, Braun CL, Ernst H (2008) The chemistry of novel xanthophylls carotenoids. *American Journal of Cardiology* 101:S50–S57.

Jeknic Z, Morre JT, Jeknic S, Jevremovic S, Subotic A, Chen THH (2012) Cloning and Functional Characterization of a Gene for Capsanthin-Capsorubin Synthase from Tiger Lily (*Lilium lancifolium* Thunb. 'Splendens') *Plant Cell Physiol* 53(11):1899–1912.

Kachanovsky DE, Filler S, Isaacson T, Hirschberg J (2012) Epistasis in tomato color mutations involves regulation of *phytoene synthase 1* expression by cis-carotenoids. *Proc Natl Acad Sci* 109(46):19021-6.

Karlova R, Rosin FM, Busscher-Lange J, Parapunova V, Do PT, Fernie AR, Fraser PD, Baxter C, Angenent GC, de Maagd RA (2011) Transcriptome and metabolite profiling show that APETALA2a is a major regulator of tomato fruitripening. *Plant Cell* 23:923–941.

Kaur D, Sharma R, Wani AA, Gill BS, Sogi DS (2006) Physicochemical changes in seven tomato (*Lycopersicon esculentum*) cultivars during ripening. *Int. J. Food Protect*. 9:747-757.

Khachik F, Beecher G, Smith JC (1995) Lutein, lycopene, their oxidative metabolites in chemoprevention of cancer. *J. Cell Biochem. Suppl*. 22:236-246.

Kim S, Park M, Yeom S *et al.* (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. *Nature Genetics* 46:270-279.

Kling DE, Schnitzer JJ (2007) Vitamin A deficiency (VAD), teratogenic, and surgical models of congenital diaphragmatic hernia (CDH). *Am J Med Genet C Semin Med Genet.* 145C(2):139–157.

Kopsell DA, Kopsell DE (2006). Accumulation and bioavailability of dietary carotenoids in vegetable crops. *Tr. Plant Sci*. 11:499-507.

Kramer MG, Redenbaugh K (1994) "Commercialization of a tomato with an antisense polygalacturonase gene: The FLAVR SAVR™ tomato story." *Euphytica* 79(3):293-297.

Krubasik P, Sandmann G (2000) Molecular evolution of lycopene cyclases involved in the formation of carotenoids with ionone end groups. *Biochemical Society Transactions* 28:806–810.

Kuntz M, Römer S, Suire C, Hugueney P, Weil JH, Schantz R, Camara B (1992) Identification of a cDNA for the plastid-located geranylgeranyl pyrophosphate synthase from *Capsicum annuum*: correlative increase in enzyme activity and transcript level during fruit ripening. *Plant J* 2:25-34.

Lange BM, Ghassemian M (2003) Genome organization in Arabidopsis thaliana: a survey for genes involved in isoprenoid and chlorophyll metabolism. *Plant Molecular Biology* 51: 925–948.

Lee HS (2001) Characterization of carotenoids in juice of red navel orange (Cara Cara). *Journal of Agricultural and Food Chemistry* 49:2563–2568.

Lenucci MS, Cadinu D, Taurino M, Piro G, Dalessandro G (2006) Antioxidant composition in cherry and high-pigment tomato cultivars. *J. Agric. Food Chem*. 54:2606-2613.

Leonardi C, Ambrosino P, Esposito F, Fogliano V (2000) Antioxidative activity and carotenoid and tomatine contents in different typologies of fresh consumption tomatoes. *J Agric Food Chem* 48:4723–4727.

Lessin WJ, Catigani GI, Schwartz SJ (1997) Quantification of cis-trans isomers of provitamin A carotenoids in fresh and processed fruits and vegetables. *J Agric Food Chem* 45:3728–3732.

Li L, Van Eck J (2007) Metabolic engineering of carotenoid accumulation by creating a metabolic sink. *Transgenic Research* 16:581–585.

Li X, Song Y, Century K, Straight S, Ronald P, Dong X, Lassner M, Zhang Y (2001) A fast neutron deletion mutagenesis-based reverse genetics system for plants. *Plant J.* 27:235–242.

Lieberman M, Segev O, *et al.* (2004) "The tomato homolog of the gene encoding UVdamaged DNA binding protein 1 (DDB1) underlined as the gene that causes the *high pigment-1* mutant phenotype." *Theoretical and Applied Genetics* 108(8):1574-1581.

Lieberman M, Segev O, Gilboa N, Lalazar A, Levin I (2004) The tomato homolog of the gene encoding UV-damaged DNA binding protein 1 (DDB1) underlined as the gene that causes the *high pigment-1* mutant phenotype. *Theor Appl Genet* 108:1574–1581.

Lin Z, Hong Y, Yin M, Li C, Zhang K, Grierson D. (2008) A tomato HD-Zip homeobox protein, LeHB-1, plays an important role in floral organogenesis and ripening. *Plant J.* 55: 301–310.

Liu Q, Xu J, Liu Y, Zhao X, Deng X, Guo L, Gu J (2007). A novel bud mutation that confers abnormal patterns of lycopene accumulation in sweet orange fruit (*Citrus sinensis* L. Osbeck). *Journal of Experimental Botany* 58:4161–4171.

Liu Y, Roof S, Ye Z, Barry C, van Tuinen A, Vrebalov J, Bowler C, 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.

Mann V, Harker M, Pecker I, Hirschberg J (2000) Metabolic engineering of astaxanthin production in tobacco flowers. *Nat Biotechnol* 18:888–892.

Manning K, Tor M, Poole M, Hong Y, Thompson AJ, King GJ, Giovannoni JJ, Seymour GB (2006) A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat. Genet*. 38:948–952.

Marin E, Nussaume L, Quesada A, Gonneau M, Sotta B, Hugueney P, Frey A, Marion-PollA (1996) Molecular identification of zeaxanthin epoxidase of *Nicotiana plumbaginifolia*, a gene involved in abscisic acid biosynthesis and corresponding to the ABA locus of *Arabidopsis thaliana*. *EMBO Journal* 15:2331–2342.

Martel C, Vrebalov J, Tafelmeyer P, Giovannoni JJ (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.

Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D (2002). Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. USA* 99:12246–12251.

Matsuzoe N, Zushi K, Johjima T (1998) Effect of soil water deficit on coloring and carotene formation in fruits of red, pink, and yellow type cherry tomatoes. *Journal of the Japanese Society of Horticultural Science* 67(4):600–606.

McCallum CM, Comai L, Greene EA, Henikoff S (2000a) Targeted screening for induced mutations. *Nature Biotechnol.* 18:455–457.

McCallum CM, Comai L, Greene EA, Henikoff S (2000b) Targeting induced local lesions in genomes (TILLING) for plant functional genomics. *Plant Physiol.* 123:439–442.

Meissner R, Chague V, Zhu Q, Emmanuel E, Elkind Y, Levy AA (2000) "Technical advance: a high throughput system for transposon tagging and promoter trapping in tomato," *The Plant Journal* 22(3):265–274.

Meissner R, Jacobson Y, Melamed S, Levyatuv S, Shalev G, Ashri A, Elkind Y, Levy A (1997) A new model system for tomato genetics. *Plant J.* 12:1465–1472.

Mejlhede N, Kyjovska Z, Backes G, Burhenne K, Rasmussen SK, *et al.* (2006) Ecotilling for the identification of allelic variation with the powdery mildew resistance genes *mlo* and *Mla* of barley. *Plant Breeding* 125: 461–467.

Menda N, Semel Y, Peled D, Eshed Y, Zamir D (2004) *In silico* screening of a saturated mutation library of tomato. *Plant J.* 38(5): 861-872.

Mialoundama AS, Heintz D, Jadid N, Nkeng P, Rahier A, Deli J, Camara B, Bouvier F (2010) Characterization of Plant Carotenoid Cyclases as Members of the Flavoprotein Family Functioning with No Net Redox Change. *Plant Physiology* 153:970–979.

Minoia S, Petrozza A, *et al*. (2010) "A new mutant genetic resource for tomato crop improvement by TILLING technology." *BMC Research Notes* 3(1): 69.

Minoia S, Piron F, Piednoir E, Caillieux E, Mosca G, Petrozza F, Cellini F, Carriero F, Bendahmane A (2007) TILLING in tomato: A reverse genetic approach for tomato crop improvement. *Sol Genome Workshop*- Jeju Island.

Monselise SP, Halevy AH (1961) Detection of lycopene in pink orange fruit. *Science* 133:1478.

Moran N, Jarvik T (2010). Lateral Transfer of Genes from Fungi Underlies Carotenoid Production in Aphids. *Science, 328* (5978):624-627.

Mortensen A, Skibsted LH, Truscott TG (2001) The interaction of dietary carotenoids with radical species. *Arch. Biochem. Biophys*. 385:13-19.

Moyle LC (2008) Ecological and evolutionary Genomics in the wild tomatoes (*Solanum* sect. Lycopersicon) *Evolution* 62(12): 2995–3013.

Mueller LA, Tanskley SD, Giovannoni JJ, *et al.* (2005a) "The tomato sequencing project, the first cornerstone of the International Solanaceae Project (SOL)," *Comparative and Functional Genomics*, 6(3):153–158.

Mueller TH, Solow N, Taylor, *et al*. (2005b) "The SOL genomics network. A comparative resource for Solanaceae biology and beyond," *Plant Physiology* 138(3):1310–1317.

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

Naito K, Kusaba M, Shikazono N, Takano T, Tanaka A, Tanisaka T, Nishimura M (2005) Transmissible and nontransmissible mutations induced by irradiating *Arabidopsis thaliana* pollen with γ-rays and carbon ions. *Genetics* 169:881–889.

Neuman H, Galpaz N, Cunningham Jr FX, Zamir D, Hirschberg J (2014) The tomato mutation *nxd1* reveals a gene necessary for neoxanthin biosynthesis and demonstrates that violaxanthin is a sufficient precursor for abscisic acid biosynthesis. *The Plant Journal* 78: 80–93.

Nielsen SS (1994) Introduction to the Chemical Analysis of Foods. Fruits and vegetables products. *Jones and Bartlett Publishers. Boston London* 202-203.

Nieto C, Piron F, Dalmais M, Marco CF, Moriones E, *et al.* (2007) EcoTILLING for the identification of allelic variants of melon eIF4E, a factor that controls virus susceptibility. *BMC Plant Biol* 7:34.

North HM, De AA, Boutin JP, Frey A, To A, Botran L, Sotta B, 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.

Okabe Y, Asamizu E, *et al.* (2011) "Tomato TILLING Technology: development of a reverse genetics tool for the efficient isolation of mutants from Micro-Tom mutant libraries." *Plant and Cell Physiology* 52(11):1994-2005.

Oleykowski CA, Bronson Mullins CR, Godwin AK, Yeung AT (1998) Mutationdetection using a novel plant endonuclease. *Nucl Acid Res* 26:4597-4602.

Paran I, van der Knaap E (2007) "Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper." *Journal of Experimental Botany* 58(14):3841-3852.

Park H, Kreunen SS, Cuttriss AJ, DellaPenna D, Pogson BJ (2002) Identification of the carotenoid isomerase provides insight into carotenoid biosynthesis, prolamellar body formation and photomorphogenesis.*The Plant Cell* 14:321–332.

Parry AD, Babiano MJ, Horgan R (1990) The role of ciscarotenoids in abscisic acid biosynthesis. *Planta* 182:118-128.

Pecker I, Gabbay R, Cunningham FX, Hirschberg J (1996) Cloning and characterization of the cDNA for lycopene *β*-cyclase from tomato reveals decrease in its expression during fruit ripening. *Plant Mol Biol* 30:807–819.

Peralta IE, Spooner D, *et al.* (2007) "Taxonomy of wild tomatoes and their relatives (*Solanum* sect. *Lycopersicoides*, sect. *Juglandifolia*, sect. *Lycopersicon*; *Solanaceae*)."*Systematic Botany Monographs* 84.

Perkel J (2008) SNP genotyping: Six technologies that keyed a revolution. *Natur Meth* 5: 447–453.

Perry JA, Wang TL, Welham TJ, Gardner S, Pike JM, *et al.* (2003) A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume *Lotus japonicus*. *Plant Physiol* 131(3): 866–871.

Pfundel EE, Renganathan M, Gilmore AM, Yamamoto HY, Dilley RA (1994) Intrathylakoid pH in isolated pea chloroplasts as probed by violaxanthin de-epoxidation. *Plant Physiology* 106:1647–1658.

Piron F, Nicolaï M, *et al.* (2010) "An Induced Mutation in Tomato eIF4E Leads to Immunity to Two potyviruses." *PLoS ONE* 5(6):e11313.

Pirrello J, Prasad BCN, Zhang W, Chen K, Mila I, Zouine M, Latché A, Pech CJ, Ohme-Takagi M, Regad F, Bouzayen M (2012) Functional analysis and binding affinity of tomato ethylene response factors provide insight on the molecular bases of plant differential responses to ethylene. *BMC Plant Biology* 12:190.

Qian W, Yang JR, Pearson NM, Maclean C, Zhang J (2012) **Balanced codon usage optimizes eukaryotic translational efficiency.** *PLoS Genet* **8**(3)**:**e1002603.

Qin G, Gu H, Ma L, Peng Y, Deng XW, Chen Z, Qu LJ (2007) Disruption of phytoene desaturase gene results in albino and dwarf phenotypes in *Arabidopsis* by impairing chlorophyll, carotenoid and gibberellins biosynthesis. *Cell Research* 17, 471–482.

Quinet M, Kinet JM, *et al.* (2011 "Flowering response of the uniflora:blind:self-pruning and jointless:uniflora:self-pruning tomato (*Solanum lycopersicum*) triple mutants." *Physiologia Plantarum* 141(2):166-176.

Ralley L, Enfissi EM, Misawa N, Schuch W, Bramley PM, Fraser PD (2004) Metabolic engineering of ketocarotenoid formation in higher plants. *Plant J* 39:477–486.

Reboul E, Borel P, Mikail C, Abou L, Charbonnier M, Caris-Veyrat C, Goupy P, Portugal H, Lairon D, Amiot MJ (2005) Enrichment of tomato paste with 6% tomato peel increases lycopene and *β*-carotene bioavailability in men. *J. Nutr.* 135:790-794.

Rice-Evans CA, Sampson J, Bramley PM, Holloway DE (1997) Why do we expect carotenoids to be antioxidants in vivo? *Free Radic Res* 26:381–398.

Rick CM (1979) Biosystematic studies in Lycopersicon and closely related species of Solanum. *Biol. Taxon. Solanaceae* 7:667–679.

Rick CM, Dempsey WH (1961) Gold struck on chromosome 6. *Report of the Tomato Genetics Cooperative* 11:21.

Rick CM, Smith PG (1953) ***Novel Variation in Tomato Species*** Hybrids. *American Naturalist* 87:359-373.

Rockholm DC, Yamamoto HY (1996) Violaxanthin de-epoxidase. *Plant Physiology* 110: 697–703.

Rodriguez-Amaya DB, Kimura M (2004) Harvestplus Handbook for Carotenoid Analysis.HarvestPlus Technical Monograph 2. Washington, DC and Cali: International Food Policy Research Institute (IFPRI) and International Center for Tropical Agriculture (CIAT).

Rodríguez-Villalón A, Gas E, Rodríguez-Concepción M (2009) Phytoene synthase activity controls the biosynthesis of carotenoids and the supply of their metabolic precursors in dark-grown *Arabidopsis* seedlings. *The Plant Journal* 60:424–435.

Römer S, Fraser PD, Kiano JW, Shipton CA, Misawa N, Schuch W, Rosati C, Aquilani R, Dharmapuri S, Pallara P, Marusic C, Tavazza R, Bouvier F, Camara B, Giuliano G (2000) Metabolic engineering of betacarotene and lycopene content in tomato fruit. *Plant J* 24:413-420.

Ronen G, Carmel-Goren L, Zamir D, Hirschberg J (2000) The regulation of 1-aminocycloprpane-1-carboxylate synthase gene expression during the transition from system-1 to system-2 ethylene synthesis in tomato. *Proc. Natl. Acad. Sci.* 97(20):11102-11107.

Ronen G, Cohen M, Zamir D, 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*. *The Plant Journal* 17(4): 341–351.

Sadler G, Davis J, Dezman D (2006) Rapid Extraction of Lycopene and *β*-Carotene from Reconstituted Tomato Paste and Pink Grapefruit Homogenate. *Journal of Food Science* 55(5):1460 – 1461.

Sanders RA, Hiatt W (2005) "Tomato transgene structure and silencing." *Nature Biotechnology* 23(3):287-289.

Sandmann G, Romer S, Fraser PD (2006) Understanding carotenoid metabolism as a necessity for genetic engineering of crop plants. *Metabolic Engineering* 8:291–302.

Sato Y, Shirasawa K, Takahashi Y, Nishimura M, Nishio T (2006) Mutant selection from progeny of gamma-ray-irradiated rice by DNA heteroduplex cleavage using *Brassica* petiole extract. *Breeding Sci.* 56:179–183.

Saze H (2008) Epigenetic memory transmission through mitosis and meiosis in plants. *Seminars in Cell & Developmental Biology* 19:527–536.

Sesso HD, Buring JE, Norkus EP, Gaziano JM (2004) Plasma lycopene, other carotenoids, and retinol and the risk of cardiovascular disease in women. *Am J Clin Nutr*. 79(1):47-53.

Sesso HD, Buring JE, Norkus EP, Gaziano JM (2005) Plasma lycopene, other carotenoids, and retinol and the risk of cardiovascular disease in men. *Am J Clin Nutr*. 81(5):990-997.

Seymour GB, Manning K, Poole M, King GJ (2008) The genetics and epigenetics of fruit development and ripening. *Curr. Opin. Plant Biol*. 11:58–63.

Shahbazi M, Gilbert M, Labouré A, Kuntz M (2007) Dual Role of the Plastid Terminal Oxidase in Tomato. *Plant Physiol*. 145(3): 691–702.

Shakhatreh Y, Haddad N, Alrababah M, Grando S and Ceccarelli S (2010) Phenotypic diversity in wild barley (Hordeum vulgare L. ssp. spontaneum (C. Koch) Thell.) accessions collected in Jordan. *Genet Resour Crop Evol*. 57:131–146.

Shao ZQ, Zhang YM, Feng XY, Wang B, Chen JQ (2012) **Synonymous codon ordering: a subtle but prevalent strategy of bacteria to improve translational efficiency.** *PLoS One* **7**(3).

Sies H, Stahl W (1997) Carotenoids and intercellular communication via gap junctions. *Int J Vitam Nutr Res* 67:364–367.

Slade J, Knauf VC (2005) "TILLING moves beyond functional genomics into crop improvement," *Transgenic Research*. 14(2):109–115.

Smith JS, Watson CF, Bird CR, Ray J, Schuch W, Grierson D (1990) "Expression of a truncated tomato polygalacturonase gene inhibits expression of the endogenous gene in transgenic plants," *Molecular and General Genetics* 224(3):477–481.

Sneath PHA, Sokal RR (1973) *Numerical Taxonomy*. Freeman, San Francisco.

Sood R, English MA, Jones M, Mullikin J, Wang DM, *et al.* (2006) Methods for reverse genetic screening in zebrafish by resequencing and TILLING. *Methods* 39(3): 220–227.

Spooner DM, Anderson JG, Jansen RK (1993) Chloroplast DNA evidence for the interrelationships of tomatoes, potatoes, and cucumber (Solanaceae). *Am. J. Bot*. 80:676-688.

Sreelakshmi Y, Gupta S, Bodanapu R, Chauhan VS, Hanjabam M, Thomas S, Mohan V, Sharma S, Srinivasan R, Sharma R (2010) NEATTILL: A simplified procedure for nucleic acid extraction from arrayed tissue for TILLING and other high-throughput reverse genetic applications. *Plant Methods* 6:3.

Srinivasan R, Chauhan VS, Hussain JA, Panigrahy M, Reddaiah B, Negi S, Sharma S, Santisree P, Kharshiing E, Tyagi W, Sharma R. TILLING (Targeting Induced Local lesions IN Genome), a reverse genetic tool for high-throughput screening of tomato mutants. *International workshop in Tomato Genomics*, Hyderabad, India.

Srivastava A, Handa AK (2005) Hormonal regulation of tomato fruit development: a molecular perspective. *J. Plant Growth Regul.* 24:67-82.

Stahl W, Sies H (1996) Lycopene: a biologically important carotenoid for humans? *Arch Biochem Biophys* 336:1–9.

Stevens AM, Rick CM (1986) The Tomato Crop: A Scientific Basis for Improvement. *In:* Genetics and breeding, J.G. Atherton and J. Rudich. [eds.]. Chapman and Hall, London, UK 35-110.

Strack D, FesterT (2006) Isoprenoid metabolism and plastid reorganisation in arbuscular mycorrhizal roots. *New Phytologist* 172:22–34.

Sun HJ, Uchii S, Watanabe S, Ezura H (2006) "A highly efficient transformation protocol for Micro-Tom, a model cultivar for tomato functional genomics," *Plant & Cell Physiology* 47(3):426–431.

Szymkowiak EJ, Irish EE (2005) "JOINTLESS suppresses sympodial identity in inflorescence meristems of tomato." *Planta* 223(4):646-658.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* 30:2725-2729.

Télef N, Stammitti-Bert L, Mortain-Bertrand A, Maucourt M, Carde JP, Rolin D,. Gallusci P (2006) Sucrose deficiency delays lycopene accumulation in tomato fruit pericarp discs. *Plant Mol. Biol*. 62:453-469.

Teodoro AJ, Oliveira FL, Martins NB, Maia GA, Martucci RB, Borojevic R (2012) Effect of lycopene on cell viability and cell cycle progression in human cancer cell lines. *Cancer Cell Int.* 12(1):36.

The Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485:635-641.

Thomas CM, Jones D, English JJ, Carroll BJ, Bennetzen JL, Harrison K, Burbidge A, Bishop GJ, Jones JDG (1994) Analysis of the chromosomal distribution of transposon-carrying T-DNAs in tomato using the inverse polymerase chain reaction. *Mol. Gen. Genet.* 242:573–585.

Thorup TA, Tanyolac B, Livingstone KD, Popovsky S, Paran I, Jahn M (2000) Candidate gene analysis of organ pigmentation loci in the Solanaceae. *Proc Natl Acad Sci USA* 97:11192–11197.

Till BJ, Colbert T, Codomo C, *et al.* (2006) "High-throughput TILLING for *Arabidopsis*," *Methods in Molecular Biology* 323:127–135.

Till BJ, Reynolds SH, Weil C, Springer N, Burtner C (2004) Discovery of induced point mutations in maize genes by TILLING. *BMC Plant Biol.* 4:12.

Till BJ, Zerr T, Comai L, Henikoff S (2006) A protocol for TILLING and Ecotilling in plants and animals. *Nat Prot* 1:2465-2477.

Tomato Crop Germplasm Committee Report (Tomato CGC) (2003) USDA National Plant Germplasm System, Washington, DC, USA.

Toor RK, Savage GP, Heeb A (2006) Influence of different types of fertilizers on the major antioxidant components of tomatoes. *J. Food Comp. Anal*. 19:20-27.

Uppala PT, Dissmore T, Lau BH, Andacht T, Rajaram S (2012) Selective Inhibition of Cell Proliferation by Lycopene in MCF-7 Breast Cancer Cells In vitro: A Proteomic Analysis. *Phytother Res*. doi: 10.1002/ptr.4764.

Vardhini BV, Rao SS (2002) Acceleration of ripening of tomato pericarp discs by brassinosteroids. *Phytochemistry* 6:843-847.

Vrebalov LJ, Pan IL, Arroyo AJ, McQuinn R, Chung M, Poole M, Rose J, Seymour G, Grandillo S, Giovannoni J, Irish VF (2009) Fleshy fruit expansion and ripening are regulated by the tomato *SHATTERPROOF* gene *TAGL1*. *Plant Cell* 21:3041–3062.

Vrebalov LJ, Ruezinsky D, Padmanabhan V, White R, Medrano D, Drake R, Schuch W, Giovannoni J (2002) A MADS-Box Gene Necessary for Fruit Ripening at the Tomato *Ripening-Inhibitor* (*Rin*) *Science* 296:343.

Wang N, Shi L, Tian F, Ning H, Wu X, Long Y, Meng J (2010) Research article Assessment of FAE1 polymorphisms in three Brassica species using EcoTILLING and their association with differences in seed erucic acid Contents. *BMC Plant Biology* 10:137.

Wassef L, Wirawan R, Chikindas M, Breslin PA, Hoffman DJ, Quadro L (2014) *β*-Carotene–Producing Bacteria Residing in the Intestine Provide Vitamin A to Mouse Tissues In Vivo. *J. Nutr.* 144(5): 608-613.

Watanabe S, Mizoguchi T, Aoki K, Kubo Y, Mori H, Imanishi S, Yamazaki Y, Shibata D, Ezura H (2007) Ethylmethanesulfonate (EMS) mutagenesis of *Solanum lycopersicum* cv. Micro-Tom for large-scale mutant screens. *Plant Biotech.* 24(1): 33-38.

Welsch R, Beyer P, Hugueney P, Kleinig H, von Lintig J (2000) Regulation and activation of phytoene synthase, a key enzyme in carotenoid biosynthesis, during photomorphogenesis. *Planta* 211:846–854.

Welsch R, Maass D, Voegel T, Dellapenna D, Beyer P (2007) Transcription factor RAP2.2 and its interacting partner SINAT2: stable elements in the carotenogenesis of *Arabidopsis* leaves. *Plant Physiology* 145:1073–1085.

Welsch R, Wust F, Bar C, Al-Babili S, Beyer P (2008) A third phytoene synthase is devoted to abiotic stress-induced abscisic acid formation in rice and defines functional diversification of phytoene synthase genes. *PlantPhysiology* 147:367–380.

West KP Jr. (2002) Extent of vitamin A deficiency among preschool children and women of reproductive age. *J Nutr.* 132(9 Suppl):2857S–2866S.

Winkler S, Schwabedissen A, Backasch D, Bokel C, Seidel C, *et al.* (2005) Target-selected mutant screen by TILLING in Drosophila. *Genome Res*. 15(5): 718–723.

Woischnik M, Moraes CT (2002) Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. *Genome Res.* 12:885–893.

Woodall AA, Britton G, Jackson MJ (1997) Carotenoids and protection of phospholipids in solution or in liposomes against oxidation by peroxyl radicals: Relationship between carotenoid structure and protective ability. *Biochim Biophys Acta* 1336:575–586.

www.eu-sol.wur.nl

Xia Y, Ning Z, Bai G, Li R, Yan G, *et al.* (2012) Allelic Variations of a Light Harvesting chlorophyll A/B-Binding Protein Gene (*Lhcb1*) Associated with Agronomic Traits in Barley. *PLoS ONE* 7(5):e37573.

Yano K, Aoki K, Shibata D (2007) Genomic databases for tomato. *Plant Biotechnol* 24:17-25.

Young KR, Ulloa CU, Luteyn JL, Knapp S (2002) Plant evolution and endemism in Andean South America: an introduction. *Bot. Rev*. 68:4–21.

Young PR, Lashbrooke JG, Alexandersson E, Jacobson D, Moser C, Velasco R, Vivier MA (2012) The genes and enzymes of the carotenoid metabolic pathway in *Vitis vinifera* L. *BMC Genomics* 13:243.

Yu S, Liao F, Wang F, Wen W, Li J, *et al.* (2012) Identification of Rice Transcription Factors Associated with Drought Tolerance Using the Ecotilling Method. *PLoS ONE* 7(2).

Yuan Q, Hill J, Hsiao J, Moffat K, Ouyang S, Cheng Z, Jiang J, Buell CR (2002) Genome sequencing of a 239-kb region of rice chromosome 10L reveals a high frequency of gene duplication and a large chloroplast DNA insertion. *Mol. Genet. Genomics* 267:713–720.

Zhang L, Ma G, Shirai Y, Kato M, Yamawaki K, Ikoma Y, Matsumoto H (2012) Expression and functional analysis of two lycopene *β*-cyclases from citrus fruits. *Planta* 236:1315–1325.

Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. Edited in *Evolving Genes and Proteins* by V. Bryson and H.J. Vogel, Academic Press, New York 97-166.

Annexure

Annexure I: List of tomato accessions used in the study

S.No	Code No.	Accession	Details	Source
1	TGRC4	LA3538	*hp-1* (*high pihment-1*), *L.esculentum*,AC	TGRC
2	TGRC5	LA3247	*u*(*uniform ripening*), *L.esculentum* cv.Craigella,AC	TGRC
3	TGRC8	LA2089	*Epi* (*Epinastic*), *L.esculentum*, VFN8	TGRC
4	TGRC13	LA3530	*gs* (*green stripe*), *L.esculentum*, AC	TGRC
5	TGRC14	LA3534	*gf* (*green flesh*), *L.esculentum*, AC	TGRC
6	TGRC17	LA2529	*alc*(*alcobaca*), *L.esculentum* cv. alcobaca	TGRC
7	TGRC21	LA3539	*ug*(*uniform grey green*), *L.esculentum*, AC	TGRC
8	TGRC22	LA3770	*nor* (*Non ripening*), *L.esculentum* ,AC	TGRC
9	TGRC24	LA3255	*L.esculentum*	TGRC
10	TGRC26	LA3537	*Nr* (*Never ripe*), *L.esculentum*	TGRC
11	TGRC29	LA3532	*r*(*Yellow flesh*), *L.esculentum*, AC	TGRC
12	TGRC31	LA2056	*r*(*Yellow flesh*), *L.esculentum*	TGRC
13	TGRC32	LA2997	*r*(*Yellow flesh*), *L.esculentum*	TGRC
14	TGRC33	2-141	*r*(*Yellow flesh*), *L.esculentum*	TGRC
15	TGRC35	LA4025	*Bog* (*Beta old gold*), *L.esculentum*	TGRC
16	TGRC36	LA1795	*rin* (*ripening inhibitor*), *L.esculentum*	TGRC
17	TGRC39	LA0806	*Bc* (*Beta crimson*), *L.esculentum* cv. High Crimson	TGRC
18	TGRC40	LA3754	*rin* (*ripening inhibitor*), *L.esculentum*	TGRC
19	TGRC43	LA2374	*B* (*Beta*), *L.esculentum* cv. Caro Red	TGRC
20	TGRC47	LA0500	*rin* (*ripening inhibitor*), *L.esculentum*	TGRC
21	TGRC55	LA0292	*Bog* (*Beta old gold*), *L.esculentum*	TGRC
22	TGRC56	LA0505	*Od*, *L.esculentum* var. Cerasiforme	TGRC
23	TGRC57	LA3465	*Lx*, *L.esculentum* cv. Laketa	TGRC
24	TGRC58	LA0276	*L.esculentum* cv. Walter	TGRC
25	TGRC59	LA3203	*L.esculentum* Red Top VF	TGRC
26	TGRC60	LA2818	*L.esculentum* cv. Large Plum	TGRC
27	TGRC61	LA2715	*L.esculentum* cv. Porphyre	TGRC
28	TGRC62	LA3632	*L.esculentum* cv. Start 24	TGRC
29	TGRC65	LA0533	*L.esculentum* cv. Condine Red	TGRC
30	TGRC66	LA3129	*L.esculentum* cv. Ehovot 13	TGRC
31	TGRC70	LA3630	*L.esculentum* cv. Vrbikanske nizke	TGRC
32	TGRC71	LA0503	*L.esculentum* cv. Roumanian Sweet	TGRC
33	TGRC72	LA0012	*L.esculentum* cv. Pearson	TGRC
34	TGRC73	LA1022	*L.esculentum* cv. VFN8	TGRC
35	TGRC75	LA3030	*L.esculentum* cv. Gardener	TGRC
36	TGRC76	LA3246	*L.esculentum* cv. Vagabond	TGRC
37	TGRC77	LA3317	*L.esculentum* cv. Campvell 28, Canada	TGRC
38	TGRC78	LA3233	*L.esculentum* cv. Pritchard	TGRC
39	TGRC79	LA0502	*L.esculentum* cv. Marglobe	TGRC
40	TGRC82	LA4024	*L.esculentum* cv. E-6203	TGRC
41	TGRC83	LA0180	*L.esculentum* cv.San Marzano	TGRC
42	TGRC84	LA1089	*L.esculentum* cv. John Baer	TGRC
43	TGRC85	LA3121	*L.esculentum* cv. Chico Grande	TGRC
44	TGRC87	LA1504	*L.esculentum* cv. Marmande	TGRC
45	TGRC89	LA1506	*L.esculentum* cv. Stone	TGRC
46	TGRC92	LA0266	*L.esculentum* cv. Earli Pak	TGRC
47	TGRC93	LA3231	*L.esculentum* cv. Gulf State Market	TGRC
48	TGRC95	LA1091	*L.esculentum* cv. Stokedale	TGRC
49	TGRC97	LA3903	*L.esculentum* cv. Prima Bel	TGRC

S.No	Code No.	Accession	Details	Source
50	TGRC99	LA3229	*L.esculentum* cv. Prospero	TGRC
51	TGRC100	LA3905	*L.esculentum* cv. Vantage	TGRC
52	TGRC101	LA1021	*L.esculentum* cv. Santa Cruz	TGRC
53	TGRC102	LA3237	*L.esculentum* Homestead 24	TGRC
54	TGRC103	LA3343	*L.esculentum* Rio Grande	TGRC
55	TGRC104	LA0516	*L.esculentum* Ace	TGRC
56	TGRC106	LA3122	*L.esculentum* Vendor	TGRC
57	TGRC107	LA3234	*L.esculentum* Sioux	TGRC
58	TGRC108	LA3243	*L.esculentum* Platense	TGRC
59	TGRC109	LA1090	*L.esculentum* Rutgers	TGRC
60	TGRC117	LA2400	*sp u L. esc*. Cv. Castlemart	TGRC
61	TGRC118	LA0274	*d l w L. esc*.	TGRC
62	TGRC119	LA0842	*glu L. esc*.	TGRC
63	TGRC120	LA3554	*yv L.esc.*	TGRC
64	TGRC121	LA0854	*fa L esc.*	TGRC
65	TGRC122	LA3579	*Xa L. esc.*	TGRC
66	TGRC123	LA2065	*spl* W6 *L. esc*.	TGRC
67	TGRC124	LA0215	*at u y L. esc*.	TGRC
68	TGRC125	LA0458	*Cmr Lv* (*L. chilense*)	TGRC
69	TGRC126	LA2133	*L. parviflorum*	TGRC
70	TGRC127	LA1016	*dps L. esc.*	TGRC
71	TGRC128	LA2999	*gf L. esc.*	TGRC
72	TGRC129	LA3430	*Xa-3 L. esc*	TGRC
73	TGRC133	LA3905	*L.esculentum* cv. Vantage	TGRC
74	TGRC134	LA0517	*L.esculentum* cv. Early Santa Clara	TGRC
75	TGRC135	LA3024	*L.esculentum* cv. Fireball	TGRC
76	TGRC136	LA4104	*sucrL.esculentum*	TGRC
77	TGRC138	LA3022	*d L.esculentum*	TGRC
78	TGRC140	LA2921	*Del L.esculentum*	TGRC
79	TGRC141	LA1996	*Aft L.esculentum*	TGRC
80	TGRC142	LA1088	*y L.esc*.cv. Ohio Globe A	TGRC
81	TGRC143	LA0744	*L.esculentum* cv. VF11	TGRC
82	TGRC145	LA0300	*bk d o p r s g L.esculentum*	TGRC
83	TGRC146	LA1500	*lp L.esculentum*	TGRC
84	IIVR1	EC520078		IIVR
85	IIVR2	WIR4361		IIVR
86	IIVR3	EC520065		IIVR
87	IIVR4	H-24		IIVR
88	IIVR5	WIR3957		IIVR
89	IIVR6	DVRT-2		IIVR
90	IIVR8	LA3957		IIVR
91	IIVR12	EC565216		IIVR
92	IIVR14	EC520076		IIVR
93	IIVR15	WIR3956		IIVR
94	IIVR17	WIR13717		IIVR
95	IIVR18	WIR3928		IIVR
96	IIVR19	EC8372		IIVR
97	IIVR21	Chiku Grande		IIVR
98	IIVR22	Pusa Rohini		IIVR
99	IIVR23	Type-1		IIVR
100	IIVR24	Pusa Sadabahar		IIVR
101	IIVR25	Shalimar-2		IIVR

S.No	Code No.	Accession	Details	Source
102	IIVR26	M.local		IIVR
103	IIVR27	Arka Alok		IIVR
104	IIVR28	TLBR-3		IIVR
105	IIVR29	TLBR-5		IIVR
106	IIVR30	Mutant		IIVR
107	IIVR31	ALT-9797		IIVR
108	IIVR32	PDT-3-1		IIVR
109	IIVR33	TLBR-2		IIVR
110	IIVR34	H-88-78-5		IIVR
111	IIVR35	A.Arka		IIVR
112	IIVR36	Riogrande		IIVR
113	IIVR37	Castle Rock		IIVR
114	IIVR38	KT-15		IIVR
115	IIVR39	Siberia		IIVR
116	IIVR40	H-88-87		IIVR
117	IIVR42	WIR-3928		IIVR
118	IIVR43	EC520077		IIVR
119	IIVR44	EC521080		IIVR
120	IIVR45	LA3996		IIVR
121	IIVR46	EC50-50		IIVR
122	IIVR47	EC3414425		IIVR
123	IIVR48	EC521078		IIVR
124	IIVR49	LA4003		IIVR
125	IIVR50	LA3956		IIVR
126	IIVR52	DT-10		IIVR
127	IIVR53	LA3995		IIVR
128	IIVR54	EC520079		IIVR
129	IIVR55	LA3980		IIVR
130	IIVR56	EC528365		IIVR
131	IIVR57	F-6102		IIVR
132	IIVR58	F-6030		IIVR
133	IIVR59	LA3928		IIVR
134	IIVR60	Gujrat Tomato		IIVR
135	IIVR61	WIR3969		IIVR
136	IIVR62	CHRT-4		IIVR
137	IIVR64	FLA7421		IIVR
138	IIVR66	H-86		IIVR
139	IIVR68	PDT-3-1		IIVR
140	IIVR69	VLT-34		IIVR
141	IIVR70	Mont favet		IIVR
142	IIVR71	IIVR-2200		IIVR
143	IIVR72	EC528367		IIVR
144	IIVR73	F-5070		IIVR
145	IIVR74	Cherry Red		IIVR
146	IIVR75	LA3967		IIVR
147	IIVR76	LA3971		IIVR
148	IIVR77	EC398405		IIVR
149	IIVR78	LA4040		IIVR
150	IIVR81	LA4024		IIVR
151	IIVR82	WIR5032		IIVR
152	IIVR83	LA3999		IIVR
153	IIVR84	BT-111-3-2-3		IIVR

S.No	Code No.	Accession	Details	Source
154	IIVR85	LA4040-2		IIVR
155	IIVR86	LA3934		IIVR
156	IIVR87	Agata-30		IIVR
157	IIVR89	BL-1208		IIVR
158	IIVR90	DVKT-1		IIVR
159	IIVR91	Azad No.1		IIVR
160	IIVR92	H-88-78-2		IIVR
161	IIVR93	Arka Vikas		IIVR
162	IIVR94	P.Rohit		IIVR
163	IIVR95	PBC(Punjab Chuhara)		IIVR
164	IIVR96	BT-12		IIVR
165	IIVR97	Nandi		IIVR
166	IIVR98	CLN-2998		IIVR
167	IIVR99	TLBR-4		IIVR
168	IIVR100	TLBR-12		IIVR
169	IIVR101	T.Local		IIVR
170	IIVR102	S-2-95-1-3-1		IIVR
171	IIVR103	Arka Saurabh		IIVR
172	IIVR104	M-88-78-3		IIVR
173	IIVR105	Cerasiformae		IIVR
174	IIVR106	PKM-1		IIVR
175	IIVR107	P.Pink		IIVR
176	IIVR108	IIHR-2201		IIVR
177	IIVR109	S.local		IIVR
178	IIVR111	P.Gaurav		IIVR
179	IIVR112	Vaibhav		IIVR
180	IIVR113	Sankranti		IIVR
181	IIVR114	Feb.4		IIVR
182	IIVR115	T-HL		IIVR
183	IIVR116	Feb.2		IIVR
184	IIVR117	Superbug		IIVR
185	IIVR118	FLA7171		IIVR
186	IIVR120	Sel-14		IIVR
187	IIVR121	?		IIVR
188	IIVR122	EC007785		IIVR
189	IIVR124	EC273966		IIVR
190	IIVR128	EC193538		IIVR
191	IIVR129	EC494372		IIVR
192	IIVR130	EC251643-3		IIVR
193	IIVR132	EC339058-A		IIVR
194	IIVR134	EC29933		IIVR
195	IIVR136	EC5627		IIVR
196	IIVR137	PUSA-SHEETAL		IIVR
197	IIVR138	PUSA-GAURAV		IIVR
198	IIVR143	EC2997		IIVR
199	IIVR152	EC362948		IIVR
200	IIVR153	EC241446-A		IIVR
201	IIVR156	EC009046		IIVR
202	IIVR158	EC5358139		IIVR
203	IIVR163	EC007345		IIVR

S.No	Code No.	Accession	Details	Source
204	IIVR166	EC33878		IIVR
205	IIVR168	EC370867		IIVR
206	IIVR169	EC538153		IIVR
207	IIVR170	PUSA-UPKAR		IIVR
208	IIVR172	EC3176		IIVR
209	IIVR173	EC241446		IIVR
210	IIVR174	EC531800		IIVR
211	IIVR175	EC1914		IIVR
212	IIVR176	EC562073		IIVR
213	IIVR178	EC251068		IIVR
214	IIVR182	EC2791		IIVR
215	IIVR184	EC57442		IIVR
216	IIVR185	EC016786		IIVR
217	IIVR186	EC6053-1		IIVR
218	IIVR188	EC52106-B		IIVR
219	NBPGR2	EC252		NBPGR
220	NBPGR3	EC487		NBPGR
221	NBPGR4	EC1087		NBPGR
222	NBPGR5	EC1154		NBPGR
223	NBPGR7	EC2673		NBPGR
224	NBPGR8	EC2790		NBPGR
225	NBPGR9	EC2798		NBPGR
226	NBPGR11	EC2977		NBPGR
227	NBPGR12	EC3176		NBPGR
228	NBPGR13	EC3216		NBPGR
229	NBPGR14	EC3668		NBPGR
230	NBPGR16	EC4506		NBPGR
231	NBPGR17	EC5863		NBPGR
232	NBPGR18	EC6192		NBPGR
233	NBPGR19	EC6486		NBPGR
234	NBPGR20	EC6845		NBPGR
235	NBPGR21	EC7912		NBPGR
236	NBPGR23	EC8591		NBPGR
237	NBPGR24	EC8822		NBPGR
238	NBPGR25	EC8936		NBPGR
239	NBPGR26	EC12689		NBPGR
240	NBPGR27	EC13736		NBPGR
241	NBPGR28	EC14073		NBPGR
242	NBPGR29	EC15127		NBPGR
243	NBPGR31	EC16786		NBPGR
244	NBPGR32	EC16790		NBPGR
245	NBPGR34	EC25265		NBPGR
246	NBPGR37	EC27910		NBPGR
247	NBPGR39	EC35244		NBPGR
248	NBPGR46	EC279088		NBPGR
249	NBPGR48	EC373378		NBPGR
250	NBPGR49	EC381554		NBPGR
251	NBPGR50	EC383117		NBPGR
252	NBPGR54	EC398701		NBPGR
253	NBPGR55	EC398707		NBPGR
254	NBPGR56	EC398711		NBPGR
255	NBPGR57	EC398715		NBPGR

S.No	Code No.	Accession	Details	Source
256	NBPGR60	EC433607		NBPGR
257	NBPGR62	EC439542		NBPGR
258	NBPGR63	EC443369		NBPGR
259	NBPGR64	EC470413		NBPGR
260	NBPGR67	EC490128		NBPGR
261	NBPGR71	EC490141		NBPGR
262	NBPGR75	EC520046		NBPGR
263	NBPGR76	EC520075		NBPGR
264	NBPGR77	EC521039		NBPGR
265	NBPGR78	EC521043		NBPGR
266	NBPGR79	EC521046		NBPGR
267	NBPGR80	EC521083		NBPGR
268	NBPGR81	EC521086		NBPGR
269	NBPGR84	EC528365		NBPGR
270	NBPGR85	EC528372		NBPGR
271	NBPGR86	EC528373		NBPGR
272	NBPGR87	EC528374		NBPGR
273	NBPGR88	EC528388		NBPGR
274	NBPGR89	EC529083		NBPGR
275	NBPGR91	EC538141		NBPGR
276	NBPGR92	EC538146		NBPGR
277	NBPGR93	EC538148		NBPGR
278	NBPGR94	EC538149		NBPGR
279	NBPGR96	EC538156		NBPGR
280	NBPGR97	EC546727		NBPGR
281	NBPGR101	EC170047		NBPGR
282	NBPGR104	EC1177297		NBPGR
283	NBPGR109	EC241446		NBPGR
284	NBPGR110	EC241446 A		NBPGR
285	NBPGR115	EC251646		NBPGR
286	NBPGR117	EC251649		NBPGR
287	NBPGR119	EC251581		NBPGR
288	NBPGR128	EC320583		NBPGR
289	NBPGR145	EC339066		NBPGR
290	NBPGR146	EC338717		NBPGR
291	NBPGR153	EC362941		NBPGR
292	NBPGR154	EC362949		NBPGR
293	NBPGR155	EC363863		NBPGR
294	NBPGR156	EC362933		NBPGR
295	NBPGR167	EC369020		NBPGR
296	NBPGR171	EC385654		NBPGR
297	NBPGR172	EC398614		NBPGR
298	NBPGR173	EC398684		NBPGR
299	NBPGR174	EC398685		NBPGR
300	NBPGR176	EC398688		NBPGR
301	NBPGR178	EC398687		NBPGR
302	NBPGR181	EC398695		NBPGR
303	NBPGR182	EC398697		NBPGR
304	NBPGR183	EC398699		NBPGR
305	NBPGR184	EC398704		NBPGR
306	NBPGR185	EC398710		NBPGR
307	NBPGR186	EC398712		NBPGR

S.No	Code No.	Accession	Details	Source
308	NBPGR188	EC398714		NBPGR
309	NBPGR189	EC398716		NBPGR
310	NBPGR190	EC398717		NBPGR
311	NBPGR194	EC433607		NBPGR
312	NBPGR196	EC458213		NBPGR
313	NBPGR219	EC520059		NBPGR
314	NBPGR221	EC521048		NBPGR
315	NBPGR223	EC521076		NBPGR
316	NBPGR224	EC521077		NBPGR
317	NBPGR225	EC521078		NBPGR
318	NBPGR226	EC521079		NBPGR
319	NBPGR227	EC521080		NBPGR
320	NBPGR229	EC521082		NBPGR
321	NBPGR231	WIR1378		NBPGR
322	NBPGR232	WIR3768		NBPGR
323	NBPGR241	EC6488		NBPGR
324	NBPGR261	EC155		NBPGR
325	NBPGR262	EC429		NBPGR
326	NBPGR273	EC742		NBPGR
327	NBPGR275	EC1191		NBPGR
328	NBPGR282	EC2347		NBPGR
329	NBPGR285	EC25D		NBPGR
330	NBPGR289	EC2053		NBPGR
331	NBPGR293	EC2765		NBPGR
332	NBPGR295	EC2802		NBPGR
333	NBPGR296	EC2990		NBPGR
334	NBPGR298	EC3176-1		NBPGR
335	NBPGR311	EC5888		NBPGR
336	NBPGR312	EC6053-1		NBPGR
337	NBPGR314	EC8630		NBPGR
338	NBPGR315	EC6875		NBPGR
339	NBPGR319	EC7317		NBPGR
340	NBPGR320	EC7345		NBPGR
341	NBPGR321	EC7785		NBPGR
342	NBPGR324	EC9046		NBPGR
343	NBPGR327	EC10662		NBPGR
344	NBPGR330	EC11309		NBPGR
345	NBPGR331	EC12692		NBPGR
346	NBPGR333	EC13274		NBPGR
347	NBPGR334	EC13574		NBPGR
348	NBPGR336	EC13904		NBPGR
349	NBPGR337	EC14181		NBPGR
350	NBPGR338	EC15416		NBPGR
351	NBPGR339	EC16343		NBPGR
352	NBPGR340	EC16368		NBPGR
353	NBPGR341	EC16788		NBPGR
354	NBPGR342	EC20636		NBPGR
355	NBPGR343	EC20639		NBPGR
356	NBPGR346	EC27885		NBPGR
357	NBPGR347	EC25563		NBPGR
358	NBPGR349	EC26150		NBPGR
359	NBPGR350	EC26676		NBPGR

S.No	Code No.	Accession	Details	Source
360	NBPGR351	EC26750-1		NBPGR
361	NBPGR352	EC27251		NBPGR
362	NBPGR353	EC27911		NBPGR
363	NBPGR354	EC27960		NBPGR
364	NBPGR356	EC27995		NBPGR
365	NBPGR357	EC28356		NBPGR
366	NBPGR359	EC29914		NBPGR
367	NBPGR361	EC29933		NBPGR
368	NBPGR362	EC29969		NBPGR
369	NBPGR363	EC30303		NBPGR
370	NBPGR364	EC31764		NBPGR
371	NBPGR366	EC32019		NBPGR
372	NBPGR370	EC32287		NBPGR
373	NBPGR371	EC32481		NBPGR
374	NBPGR372	EC32557		NBPGR
375	NBPGR374	EC32614		NBPGR
376	NBPGR378	EC33878		NBPGR
377	NBPGR379	EC34477		NBPGR
378	NBPGR380	EC34480		NBPGR
379	NBPGR384	EC35236		NBPGR
380	NBPGR386	EC35240		NBPGR
381	NBPGR387	EC35242		NBPGR
382	NBPGR388	EC35252		NBPGR
383	NBPGR391	EC35272		NBPGR
384	NBPGR392	EC35293		NBPGR
385	NBPGR394	EC35322		NBPGR
386	NBPGR395	EC35360		NBPGR
387	NBPGR397	EC36238		NBPGR
388	NBPGR398	EC89248		NBPGR
389	NBPGR399	EC161645		NBPGR
390	NBPGR401	EC163598		NBPGR
391	NBPGR403	EC164660		NBPGR
392	NBPGR404	EC164665		NBPGR
393	NBPGR406	EC27336		NBPGR
394	NBPGR410	EC129602		NBPGR
395	NBPGR411	EC129604		NBPGR
396	NBPGR412	EC141887		NBPGR
397	NBPGR413	EC144336 A		NBPGR
398	NBPGR415	EC490122		NBPGR
399	NBPGR416	IC447708		NBPGR
400	NBPGR417	EC496124		NBPGR
401	NBPGR419	EC357828		NBPGR
402	NBPGR421	EC381554 A		NBPGR
403	NBPGR424	EC490128		NBPGR
404	NBPGR427	EC490130		NBPGR
405	NBPGR431	IC447706		NBPGR
406	NBPGR432	IC469682		NBPGR
407	NBPGR433	IC469653		NBPGR
408	NBPGR434	IC469633		NBPGR
409	NBPGR435	IC469648		NBPGR
410	NBPGR436	IC469629		NBPGR
411	NBPGR437	IC469628		NBPGR

S.No	Code No.	Accession	Details	Source
412	NBPGR438	IC469626		NBPGR
413	NBPGR439	IC469603		NBPGR
414	NBPGR441	EC57442		NBPGR
415	NBPGR442	IC469525		NBPGR
416	NBPGR443	IC469597		NBPGR
417	NBPGR469	EC520052		NBPGR
418	NBPGR471	IC469714		NBPGR
419	NBPGR477	EC398716		NBPGR
420	NBPGR479	EC531801		NBPGR
421	NBPGR480	EC529086		NBPGR
422	NBPGR482	EC398691		NBPGR
423	NBPGR483	EC531802		NBPGR
424	NBPGR484	EC398710		NBPGR
425	NBPGR485	EC6486		NBPGR
426	NBPGR486	EC6192		NBPGR
427	NBPGR490	EC521067 B		NBPGR
428	NBPGR491	EC521068		NBPGR
429	NBPGR495	EC368883		NBPGR
430	NBPGR502	EC362958		NBPGR
431	NBPGR504	EC363942		NBPGR
432	NBPGR514	EC538139		NBPGR
433	NBPGR515	EC528362		NBPGR
434	NBPGR516	EC538153		NBPGR
435	NBPGR517	EC538455		NBPGR
436	NBPGR518	EC529085		NBPGR
437	NBPGR520	EC526146		NBPGR
438	NBPGR521	EC531805		NBPGR
439	NBPGR523	EC529081		NBPGR
440	NBPGR526	EC368943		NBPGR
441	NBPGR533	EC168290		NBPGR
442	NBPGR537	EC164660		NBPGR
443	NBPGR539	EC16788		NBPGR
444	NBPGR542	EC177371		NBPGR
445	NBPGR543	EC25265		NBPGR
446	NBPGR547	EC16790		NBPGR
447	NBPGR548	EC16780		NBPGR
448	NBPGR549	EC168283		NBPGR
449	NBPGR550	EC26684		NBPGR
450	NBPGR551	EC12692		NBPGR
451	NBPGR552	EC12689		NBPGR
452	NBPGR553	EC7912		NBPGR
453	NBPGR554	EC14078		NBPGR
454	NBPGR555	EC16654		NBPGR
455	NBPGR557	EC13736		NBPGR
456	NBPGR558	EC29933		NBPGR
457	NBPGR559	EC8936		NBPGR
458	NBPGR561	EC114375		NBPGR
459	NBPGR562	EC2673		NBPGR
460	NBPGR565	EC2798		NBPGR
461	NBPGR566	EC5888		NBPGR
462	NBPGR567	EC2977A		NBPGR
463	NBPGR568	EC1914		NBPGR

S.No	Code No.	Accession	Details	Source
464	NBPGR569	EC3668		NBPGR
465	NBPGR570	EC3261		NBPGR
466	NBPGR571	EC31767		NBPGR
467	NBPGR572	EC2791		NBPGR
468	NBPGR573	EC3176		NBPGR
469	NBPGR574	EC3216		NBPGR
470	NBPGR579	EC35240		NBPGR
471	NBPGR581	EC1087		NBPGR
472	NBPGR582	EC129604		NBPGR
473	NBPGR584	EC14073		NBPGR
474	NBPGR585	EC27910		NBPGR
475	NBPGR586	EC251581		NBPGR
476	NBPGR589	EC35293		NBPGR
477	NBPGR590	EC27960		NBPGR
478	NBPGR592	EC141887		NBPGR
479	NBPGR596	EC2990		NBPGR
480	NBPGR597	EC34480		NBPGR
481	NBPGR598	EC2977-A		NBPGR
482	NBPGR600	EC8591		NBPGR
483	NBPGR601	EC8822		NBPGR
484	NBPGR602	EC135580		NBPGR
485	NBPGR603	EC2790		NBPGR
486	NBPGR605	EC52077		NBPGR
487	NBPGR607	EC276		NBPGR
488	NBPGR609	EC5863		NBPGR
489	NBPGR610	EC252		NBPGR
490	NBPGR611	EC372		NBPGR
491	NBPGR614	EC170047		NBPGR
492	NBPGR615	EC50055		NBPGR
493	BSS1	N2219	NA	BSS
494	BSS2	N2231	NA	BSS
495	BSS3	N2241	NA	BSS
496	BSS4	N2244	NA	BSS
497	BSS5	N2263	NA	BSS
498	BSS6	N2266	NA	BSS
499	BSS7	N2267	NA	BSS
500	BSS8	N2268	NA	BSS
501	BSS9	N2269	NA	BSS
502	BSS10	N2279	NA	BSS
503	BSS11	N2280	NA	BSS
504	BSS12	N2292	NA	BSS
505	BSS13	N2298	NA	BSS
506	BSS14	N2369	NA	BSS
507	BSS15	N2370	NA	BSS
508	BSS16	N2387	NA	BSS
509	BSS17	N2389	NA	BSS
510	BSS18	N2403	NA	BSS
511	BSS19	N2411	NA	BSS
512	BSS20	N2414	NA	BSS
513	BSS21	N2448	NA	BSS
514	BSS22	N2457	NA	BSS
515	BSS23	N2465	NA	BSS

S.No	Code No.	Accession	Details	Source
516	BSS24	N2466	NA	BSS
517	BSS25	N2469	NA	BSS
518	BSS26	N2480	NA	BSS
519	BSS27	N2481	NA	BSS
520	BSS28	N2485	NA	BSS
521	BSS29	N2523	NA	BSS
522	BSS30	N2535	NA	BSS
523	BSS31	N2536	NA	BSS
524	BSS32	N2537	NA	BSS
525	BSS33	N2673	NA	BSS
526	BSS34	N2717	NA	BSS
527	BSS35	N2730	NA	BSS
528	BSS36	N2731	NA	BSS
529	BSS37	N2733	NA	BSS
530	BSS38	N2735	NA	BSS
531	BSS39	N2764	NA	BSS
532	BSS40	N2812	NA	BSS
533	BSS41	N2829	NA	BSS
534	BSS42	N2832	NA	BSS
535	BSS43	4218	NA	BSS
536	BSS44	4219	NA	BSS
537	BSS45	4225	NA	BSS
538	BSS46	4228	NA	BSS
539	BSS47	4229	NA	BSS
540	BSS48	4240	NA	BSS
541	BSS49	4241	NA	BSS
542	BSS50	4243	NA	BSS
543	BSS51	4257	NA	BSS

TGRC: Tomato Genetics Resource Center at University of California, Davis (tgrc.ucdavis.edu/).
IIVR : Indian Institute of Vegetable Research, Varanasi, U. P., India (www.iivr.org.in/).
NBPGR: National Bureau of Plant Genetic Resources, New Delhi, India (www.**nbpgr**.ernet.in/).
BSS: Bejo Sheetal Seeds Pvt. Ltd. Jalna, India (http://www.bejosheetalseeds.com/).
NA: Not available

NB: The details of NBPGR accessions are available at the PGR (Plant Genetic Resources) portal of NBPGR site (http://www.nbpgr.ernet.in:8080/PGRPortal/(S(o3cy12bkoz5s5e55w0dacf2a))/AdvancePassportSearch.aspx).