Harnessing Natural Diversity in Tomato to Improve Folate Levels Using Metabolic Profiling and Genic Polymorphism

Thesis submitted for the award of the degree of

DOCTOR OF PHILOSOPHY

By

Pallawi Upadhyaya

Supervisor

Prof. R. P. Sharma

Repository of Tomato Genomics Resources

Department of Plant Sciences

School of Life Sciences

University of Hyderabad

September, 2016

ENROLLMENT No. 10LPPH06

Repository of Tomato Genomics Resources
Department of Plant Sciences
School of Life Sciences
University of Hyderabad

DECLARATION

I, **Pallawi Upadhyaya**, hereby declare that the work described in this thesis entitled "**Harnessing Natural Diversity in Tomato to Improve Folate Levels Using Metabolic Profiling and Genic Polymorphism**" submitted by me under the supervision of **Professor R. P. Sharma**, Department of Plant Sciences, is an original research work. I also declare that it has not been submitted previously in part or in full to this University or any other University or Institution for the award of any degree or diploma. A report on plagiarism statistics from the University Librarian is enclosed.

Date
Place: Hyderabad

Pallawi Upadhyaya
Enrol. No. 10LPPH06

Prof. R. P. Sharma
Supervisor

Repository of Tomato Genomics Resources
Department of Plant Sciences
School of Life Sciences
University of Hyderabad

CERTIFICATE

This is to certify that the thesis entitled **"Harnessing Natural Diversity in Tomato to Improve Folate Levels Using Metabolic Profiling and Genic Polymorphism"** is based on the results of the work done by **Miss. Pallawi Upadhyaya** for the degree of **Doctor of Philosophy** under my supervision. This work presented in this thesis is original and has not been submitted for any degree or diploma of any other University. A report on plagiarism statistics from the University Librarian is enclosed.

Head
Dept. of Plant Sciences

Dean
School of Life Sciences

Prof. R. P. Sharma
Supervisor

Dedicated to
My adorable Parents
& Family

Acknowledgements

I owe my gratitude to all those people who have made this thesis possible and because of whom I could accomplish this work.

My sincere gratitude is to my mentor, Prof. R. P. Sharma. It has been a great experience working under his guidance and support. I appreciate your scientific disposition and all the discussion sessions with you have been a great learning experience. Thank you very much for all the support through these years.

I would like to express my special appreciation and thanks to Dr. Y. Sreelakshmi, you have been an inspiring figure for me. I would like to thank you for encouraging my research and for constant inspiration.

My sincere thanks to present and former Deans Prof. P. Reddanna, Prof. A. S. Raghavendra (Dean, School of Life Sciences), Prof R. P. Sharma, Prof M. Ramnadham (Former Deans, School of Life Sciences); Prof Ch. Venkata Ramana (Head, Dept. of Plant Sciences), and Prof A. R. Reddy, for allowing me to utilize the facilities of the School and Departments.

I also thank my doctoral committee members Prof A. R. Reddy and Dr. Y. Sreelakshmi for their guidance and timely suggestions throughout the research.

I thank all my lab members for their love, help and guidance. I sincerely thank Dr. Sulabha, Dr. Vineeta, Dr. Soni, Mr. Reddaiah, Dr. Vijee, Dr. Santisree, Dr. Sherin, Dr. Mickey, **Dr. Vajir**, Dr. Vineet, Dr. Bharti and Dr. Rahul for their support and suggestions. I extend my gratitude to my lab seniors Dr. Rakesh, Dr. Supriya, Rachna, Dr. Sapana, Dr. Kapil and Dr. Kamal. I am especially grateful to Dr. Alka, **Dr. Suresh and Dr. Chaitanya** who have taught me more than just science and have been a great moral

support throughout. **Dr. Alka**, who stayed with me for the latter duration of my Ph.D. has been a true senior and friend and has been a great support. I am very grateful to **Prateek**, who helped me in most of the bioinformatics work during my research, especially in the *in silico* work. I extend my love and thankfulness to Hyma, Anusha, Himabindu, Manjari, Swathi, Gayathri and Sumona for their help and live company in the lab. My sincere thanks to Dr. Arun Kumar Pandey for his various suggestions. I thank all my well-wishers for their wishes.

I extend my thanks to field assistant (Narshimha, Zameer, Yadagiri) and lab accountants Venkat Reddy, Bhaskar garu and Anil for maintaining my plants in green house and taking care of various administrative issues.

I now thank my parents who have loved and supported me in all the ups and downs during my Ph.D. work and have been a tremendous cause in maintaining my mental sanity. Their encouragement keeps me going and I am very fortunate in being born to such amazing parents. I am thankful to my siblings who always boosted my mood and saved me from melancholic plunges! My sister has been like my other self and I always look forward to her in both, good and bad times.

Finally, I acknowledge the financial support provided by the joint CSIR-UGC New Delhi, in the form of JRF and SRF.

Last, but by no means least I would like to express my gratitude and thanks to almighty God for his kindness, blessing and strength which has helped me to successfully achieve this goal.

Pallawi Upadhyaya

2016

ABBREVIATIONS

ADCL	*Aminodeoxychorismate lyase*
ADCS	*Aminodeoxychorismate synthase*
ATP	Adenosine triphosphate
AV	Arka-Vikas
BLAST	Basic Local Alignment Search Tool
bp	Base pair
cDNA	Complementary DNA
cm	Centimeter
cv.	Cultivar
DEPC	Diethyl pyrocarbonate
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxyribonucleotide triphosphate
EMS	Ethyl Methyl Sulfonate
Et-Br	Ethidium bromide
FW	Fresh weight
GCH I	*GTP cyclohydrolase* I
GGH	*γ-glutamyl hydrolase*
HPLC	High Performance Liquid Chromatography
HT	Haplotype
HPPK–DHPS	Dihydropterin pyrophosphokinase - dihydropteroate synthase
IIVR	Indian Institute of Vegetable Research
KbP	Kilobase pairs
LC-MS	Liquid Chromatography Mass Spectrometry
MA	Microbiological assay
Mb	Megabase
MG	Mature green
mg	Milligram

mRNA	Messenger RNA
NBPGR	National Bureau of Plant Genetic resources
NCBI	National Centre for Bioinformatics Information
nt	Nucleotide
O.D.	Optical Density
pABA	p –Aminobenzoate
PCR	Polymerase chain reaction
QTL	Quantitative trait locus
qRT-PCR	Quantitative real-time RT PCR
RNA	Ribonucleic acid
RNAse	Ribonuclease
RR	Red ripe
RT	Room temperature
SDS	Sodium dodecyl sulphate
SIFT	Sorting Intolerant From Tolerant
TGRC	Tomato Genetic Resource Center
THF	Tetrahydrofolate
v/v	volume/volume
w/v	weight/volume
WT	Wild type

TABLE OF CONTENTS

Declaration i

Certificate ii

Acknowledgment iii

Abbreviation v

Chapter 1: Introduction 1-5

Chapter 2: Review of literature 6-28

2.1 Folates:- as vitamins and role in metabolism

2.1.1 Folate in plant metabolism

2.2 Rich sources of folate

2.3 Folate deficiency in humans

2.3.1 Methylation cycle and folates

2.4 Structure and biosynthesis of folates

2.4.2 Folate breakdown

2.5 Transport of folate precursors and derivatives

2.6 Folate estimation in crops and vegetables

2.6.1 Folate fortification and enhancement in crop plants

2.7 Introgression lines as useful resource

2.8 Mutants in the folate biosynthesis pathway

2.8.1 Folylpolyglutamate synthase and their roles

2.8.2 Gamma glutamyl hydrolases

2.8.3 TILLING/EcoTILLING as a tool to isolate folate pathway gene variants

2.8.3.4 Next generation sequencing

2.9 EcoTILLING

Chapter 3: Materials and method 29-43

3.1 Plant material

3.2 TILLING Population development

3.3 PCR and mutation detection: LI-COR

3.4 EcoTILLING

3.5 Folate extraction and estimation

3.8-3.9 extraction and estimation of folate precursors pABA and pterin

3.10 RNA extraction, cDNA synthesis and real time PCR

Chapter 4: Folate estimation in natural accessions of tomato **44-57**

4.1.1 Microbiological assay of folate estimation using Lactobacillus casei

4.2 Results

4.3 Discussion

4.4 Folate estimation using LC-MS

4.5 Results

4.6 Discussion

Chapter 5: Characterization of high and low folate accessions **58-65**

5.1 Introduction

5.2 Results

5.3 Total folate content in various tissues of Sel.14

5.4 Characterization of low folate accession EC27910

5.5 Comparative transcript profiling among high and low accessions

5.6 Discussion

Chapter 6: Search for folate regulatory bins in introgression lines **66-69**

6.1 Introduction

6.2 Results

6.3 Discussion

Chapter 7: Assessment of genic polymorphism in folate turnover genes **70-88**

7 Introduction

7.1 EcoTILLING results

7.2 TILLING results

7.2.1 Mutants identified in *FPGSm* and *GGH*

7.2.2.1 Characterization of *GGH2* mutant line

7.2.3 TILLING by sequencing

7.3 *In silico* survey of natural polymorphism in 444 natural accession

Chapter 8: Summary **89-94**

References **95-113**

Annexure 1

Annexure 2

Annexure 2.1

Annexure 3

Annexure 3.1

Chapter 1

Introduction

INTRODUCTION

Micronutrients play a vital role in the maintenance of a healthy metabolism. Vitamins and minerals are micronutrients which are required in very small amount but are highly crucial for the maintenance of cellular processes. Vitamins act as cofactors in enzymatic reactions. Their deficiency results in perturbed metabolism eventually leading to manifestation of disease symptoms. Currently six vitamins are known viz. A, B, C, D, E and K. Among these, B vitamins constitute a big group comprising B1, B2, B3, B5, B6, B7, B9 and B12 commonly known as vitamin B complex. These different classes of B vitamins are involved in the metabolism of carbohydrates, fatty acids and proteins, exchange of hydrogen in energy transfer reactions and provide various chemical units required for these reactions.

One of the important vitamins of B complex group is vitamin B9 which is also popularly known as folates. The name 'Folate' is derived from Latin word *folium*, which means "leaf". The name is specifically linked to their abundance in dark green leafy vegetables among the commonly consumed foods. *In vivo* folates exist as various chemical derivatives of a common molecule namely tetrahydro folate (THF) and are collectively known as folate forms or folate vitamers. The synthetic form of folate is the oxidized form called as folic acid. The synthetic form as pills is prescribed to pregnant women or to the patients facing folate deficiency.

Folate is synthesized only by plants, certain bacteria, and fungi. The animals do not synthesize it due to lack of enzymes of folate biosynthesis pathway. Though folate is required by both, animals and plants, animals obtain it only through dietary sources. This implies that a regular intake of folate is needed to avoid the folate deficiency.

In living organisms folates act as a cofactor by transferring (donating or receiving) one carbon units (for example methyl, methylene, and formyl) in the biochemical reactions which different folate vitamers carry. These reactions are known as one carbon metabolism or C1 metabolism in its entirety. Though every micronutrient has its unique role in the maintenance of a healthy metabolism, the folate plays a crucial role due to its involvement in the synthesis of DNA from its precursors (thymidine and purine) and metabolism of many amino acids (methionine, glycine, serine, cysteine, and histidine). These molecules are directly related to DNA synthesis and subsequently cell division and therefore are of principal importance in a cell. Among other amino acids methionine is particularly important since S-adenosylmethionine (SAM) is produced from it. SAM is the key donor of methyl group in several methylation reactions of DNA, RNA, proteins and

phospholipids. These methylation reactions lead to regulated expression of gene as per cellular metabolic needs.

Apart from these folate roles that are common for animals and plants; plants also have their specific requirements. In plants, folates are necessary for photorespiration, biosynthesis of lignin, alkaloids, betaines, and chlorophyll (Hanson and Roje, 2001), early seedling development (Srivastava et al., 2011), chromatin remodeling (Zhou et al., 2013) and in N metabolism (Jiang et al., 2013). Role of folates in photorespiration was established by Collakova et al. (2008) in recycling of excess glycine generated through photorespiration through interconversions of folate species. In this process, THF, 5,10-Methenyl THF, 5,10-Methylene THF are mainly involved. Since, plants can synthesize folates to cater for its requirements plants do not face folate deficiency like animals.

Though plants are most important source of dietary folate, most of the commonly consumed staple food items of daily diet viz. corn, rice, wheat, cassava etc. are not rich in folate leading to folate deficiency. Insufficient intake of folate impairs the C1 metabolism and results in several diseases such as neural tube defects (NTDs for example anencephaly, spina bifida), cardiovascular diseases, and several psychiatric anomalies like depression and irritability in humans. NTDs is the major indicator of folate deficiency and its prevalence worldwide (expressed as number of NTDs per 10,000 births) is used to represent the severity of the problem. Several other diseases like colorectal cancer, dementia and cognitive dysfunction are inversely related to the adequate folate intake. To overcome the problems arising due to chronic folate deficiency, developed countries fortify the staple food like bread, rice and breakfast cereals with folic acid. However this approach is expensive and cannot be practiced in developing countries due the absence of a national food supply system. It can also result in excess intake of folate but may miss the target population (women of childbearing age) (Osterhues et al., 2013). In addition, developing countries cannot afford food fortification programs. An alternate approach for ensuring folate intake is, biofortification by making the crops naturally rich in folate content. This can be achieved via two main approaches, 1) - by finding folate rich foods existing naturally and characterizing them and 2) - enhancing the folate levels of staple food crops by metabolic engineering. The latter approach has some extent of rejection and apprehension about safety of genetically modified (GM) food. Considering this, identification of the varieties/ germplasms of common food crops which can offer appropriate amount of folate through daily serving is a better alternative.

There are many fruits, vegetables, pulses and animal food sources that can provide adequate amounts of folate when consumed regularly. Among these rich folate sources are spinach, green leafy vegetables, avocado, citrus fruits, legumes (beans, pulses, lentils), broccoli, yeast, asparagus, Brussels sprouts, liver and fermented dairy products (https://ods.od.nih.gov/factsheets/Folate-HealthProfessional/). Fermented dairy products contain more folate than milk which is attributed to the bacteria causing the fermentation (Iyer and Tomar, 2009). Though these food items are rich in folates yet the daily diet mostly consists of cereal products which are poor source of folates (Bekaert et al., 2007). Therefore, many countries including United States, Canada, Costa Rica, Chile and South Africa have started mandatory addition of folic acid (food fortification) in breads, cereals, cornmeal, flours, rice, pasta and other grain based products.

The staple crops like rice, wheat and corn were the first targets of biofortification efforts to mitigate the nutrient deficiency of large population in developing countries. The biofortification was attempted by engineering of pathway genes for important metabolites. Overexpression of the key genes either from the same family or different organisms was tried to enhance the accumulation of metabolites of interest. The first successful example was golden rice (GR) by enriching it with provitamin A (Ye et al., 2000). Currently several crops have been biofortified such as corn (Naqvi et al., 2009), wheat (Cakmak et al., 2010), potato (Diretto et al., 2007) and tomatoes (Muir et al., 2001; Butelli et al., 2008) for nutrients vitamin A, C and B9 (folate), iron and zinc, carotenoids, anthocyanins and flavonols etc. with excellent to moderate success. Several efforts have been also made to improve folate levels in Arabidopsis, tomato, rice, corn, lettuce and potato (Hossain et al., 2004; Diaz de la Garza et al., 2004; Diaz de la Garza et al., 2007; Storozhenko et al., 2007a; Naqvi et al., 2009; Nunes et al., 2009 and Blancquaert et al., 2013). However the crop specific variations were seen with respect to folate accumulation. While the overexpression of GTP cyclohydrolase I (GCHI) and Aminodeoxychorismate synthase (ADCS) resulted in enrichment of total folate in rice and tomato, it was not as successful in Arabidopsis and potato. This variation may arise as the regulatory mechanism of folate biosynthesis may be different in every crop.

Among the vegetable crops, tomato is the most commonly consumed. It is grown and consumed worldwide. Tomato is rich in vitamin A, vitamin C and E. Other important nutrients found in tomato are carotenoids, mainly β-carotene, which is a provitamin A. Tomato also contains potassium, sugars, flavonols and folates (Canene-Adams et al., 2005). Most importantly it contains high amount of lycopene which has anticancer properties.

Tomato is consumed in various forms. As raw or slightly blanched/roasted forms are eaten as side dish and salads, processed forms of tomato is consumed as sauce, juice, puree, pickles and chutneys. Considering wide consumption of tomato makes it a suitable target for biofortification. Biofortification is usually achieved through conventional breeding techniques or through genetically engineering the pathway gene expression. First genetic engineering efforts to enhance the folate content of tomato involved overexpression of GTP cyclohydrolase I=GCHI (the first committed step of pterin biosynthesis) (Diaz de la Garza et al., 2004) which resulted in limited success with only two fold enhancement in total folate level. However, overexpression of aminodeoxychorismate synthase (ADCS), the first step in pABA synthesis and crossing these plants with the GCHI overexpression line (Diaz de la Garza et al., 2007) resulted in a 25 fold increment in total folate content. However, these tomatoes are not yet released for consumers owing to the general disagreement over genetically modified (GM) foods. There is little information about the variation of folate levels in the natural accessions and landraces of tomato (Iniesta et al., 2009). This is an area which is not explored and lines enriched in folate can be identified by screening large number of tomato germplasms. The folate enriched lines can be used for breeding high folate cultivars.

In addition of identifying the naturally occurring variations in the folate biosynthesis pathway genes (EcoTILLING, gain or loss of function), the alternate approach of enhancing folate levels by mutagenesis (TILLING) can also help in improving folate levels. The variations observed in total folate content and in folate biosynthesis pathway genes at genomic level (single nucleotide polymorphism=SNP) can be used for further characterization and breeding work. Thus, these approaches can improve the folate levels in tomato.

The identification of folate regulatory mechanisms can also be carried out by examination of folate levels in tomato introgression lines (ILs). ILs, which contain well defined chromosomal segments of a wild relative into the common cultivar's genome are otherwise near isogenic to the introgressed parent's genome. Therefore, screening ILs for folate levels can lead to identification of quantitative trait loci (QTLs) for folate biosynthesis. Tomato cultivar M82 and *Solanum pennellii* IL population has been used extensively since its creation (Eshed and Zamir, 1994a, b). Various traits have been explored like fruit antioxidants (Rousseaux et al., 2005), root morphology and cellular development (Ron et al., 2013), vitamin E determinants (Quadrana et al., 2014; Raiola et al., 2015) and QTLs affecting the root response under chilling stress in *S. lycopersicum* X *S.*

habrochaites ILs (Arms et al., 2015). ILs are a useful resource as wild varieties are often genetically richer than the cultivated ones and clearly defined segments from wild relative may help in ascertaining the regulation of a specific trait.

It is evident from foregoing that though folate is an important cofactor for plant metabolism little is known about its regulation in higher plants. To decipher how folate level is regulated, a multi-pronged approach was outlined in this study using tomato as a model species. At the outset the following research objectives were outlined-

- Examination of variation in folate levels in natural accessions of tomato and introgression lines of tomato crossed with *S. pennellii* (and *S. lycopersicum* X *S. pennellii* ILs).
- Identification of genetic variations (SNPs) in γ-glutamyl hydrolase (*GGH*) and folylpolyglutamate synthase (*FPGS*) genes in tomato natural accessions using Eco-TILLING.
- Identification of mutation by using TILLING in *GGH* and *FPGS* genes in an EMS mutagenized M_2 population of tomato.

Chapter 2

Review of literature

REVIEW OF LITERATURE

Vitamins are micronutrients which serve as cofactor in the metabolism and are equally important for both, plants and animals. They carry out almost the similar functions in plants and animals. The only difference is that animals cannot synthesize them and depend on plants for their vitamin needs while plants synthesize most of the vitamins except B12 whose functions in plant metabolism has been suitably substituted (Smith et al., 2007).

Though many roles of vitamins in plants and animals are shared, yet they also have unique functions. For example, characterization of Arabidopsis salt overly sensitive (*sos4*) mutant showed that *SOS4* gene coded for pyridoxal kinase which regulated Na^+ and K^+ homeostasis in plants by modulating the ion transporter activity (Shi et al., 2002). Most of the actions of vitamins are common between animals and plant metabolism. For example, vitamins of B complex group perform carboxylation, decarboxylation reactions in central carbon metabolism, electron transfer redox reactions, acyl group transfer in lipid metabolism and C1 group transfer reactions. Similarly, vitamin C is an antioxidant and functions in ascorbate-glutathione cycle (Katoh and Hashimoto, 2004; Fischer and Bacher, 2006; Drewke and Leistner, 2001; Matamoros et al., 2006; Alban et al., 2000; Webb et al., 2004; Rébeillé et al., 2006).

One distinct difference between plants and animal is that the vitamin deficiency symptoms are only observed in animals. This relates to the fact that animals are dependent on plants for their regular vitamin intake. The plants experience such perturbations only when there are lesions in vitamins biosynthesis genes. Many of the plants metabolites, vitamins and their precursors have unique function in animals. The carotenoids are known for their photoprotective role in plants, act as precursor of abscisic acid and also serve as attractants in flowers and fruits for animals. In animals these play very diverse and important roles such as regulation of immune functions, reproduction and cellular communication (Bartley and Scolnik, 1995). It is known that carotenoids are vital for vision as a critical part of rhodopsin in retinal receptors that absorb light. The difference between plant and animal derived carotenoids is that the plants provide provitamin A which is converted to vitamin A in human body while animal based diet sources like fish meat, dairy and poultry products provide pre-formed vitamin A. It is evident that despite sharing basic molecular mechanisms, involvement of vitamins in plants and animals can differ in different metabolic pathways.

Among the various vitamins, B_9 or folates are very important due to their multifarious roles particularly for their role in very initial metabolic reactions required for the cell division. In the ensuing sections the various aspects of folate biology are reviewed in details.

2.1 Folates:- As vitamins and role in metabolism

Folates are water soluble vitamins which participate in C1 metabolism (Figure 2.1, 2.5) and occur as various one carbon (methyl, formyl, methenyl, and methylene) chemical derivatives of *tetrahydrofolate* (THF). THF is made up of pterin, pABA and a single glutamate molecule (H_4PteGlu) (Figure 2.2 and 2.3). The various single carbon groups of THF derivatives are received or donated in chemical reactions, thus performing one carbon metabolic reactions (Cossins, 2000). Its fully oxidized monoglutamate form, folic acid is the synthetic form which is used for food fortification and for treating the deficiency conditions.

2.1.1 Folates in plant metabolism

Folates are involved in key reactions like synthesis of DNA and several amino acids which are the known functions in both prokaryotes and eukaryotes (and plants and animals). At the same time there are many specific roles that folates play in plants and more roles are also being discovered. Apart from basic functions of folates to carry out one carbon transfer, there are various processes/reactions which are plant specific. There is differential abundance of different folate forms among plants which is governed by temporal and spatial requirements of the cell or tissue. Among the folate forms, 5-methylTHF is generally the most abundant form (Shohag et al., 2011; Van Daele et al., 2014) though this is not the universal trait. The internal cues and metabolic requirements of a plant determines the accumulation of a particular folate form. The extraction procedures also have a profound effect on the observed folate profiles (discussed later in chapter 4). The abundance of 5-methylTHF could be due to higher stability of THF form or due to the critical role of folate for donating a methyl group to homocysteine during the formation of methionine. Methionine is a precursor of S-adenosyl methionine (SAM). In turn SAM as a methyl and sulfate donor is important for synthesis of a wide range of molecules including the methylation of DNA leading to chromatin silencing (Zhou et al., 2013 and Groth et al., 2016; Chiang et al., 1996).

Several of plant specific metabolic reaction requires participation of folates such as photorespiration (Hanson and Roje, 2001), chlorophyll and lignin biosynthesis as well (Van Wilder et al., 2009 Srivastava et al., 2015).

2.2 Rich sources of folate and recommended daily allowance (RDA)

Plant-based foods are the main dietary sources of folate for humans and other animals. Among plant-based foods; fruits, nuts, and vegetables provide about 30% of requirement of folate in the American diet (Kader et al., 2004). The leafy vegetables such as spinach, lettuce, broccoli, asparagus, and fruits such as citrus are the enriched sources of dietary folate (Kader et al., 2004; Delchier et al. 2016) (Figure 2.4, Table 2.1). The natural forms of folate (from dietary sources) are considered better for intestinal absorption than the synthetic form (folic acid). However, staple food items consumed in developing countries such as wheat, maize, and rice contain very low amount of folate (Table 2.2) which is insufficient to meet its RDA and results in deficiency. Folate deficiency is a serious problem as it not only affects a person during its lifetime but also induces abnormalities in fetal development thus affecting the next generation as well. Prevention of folate deficiency is a serious concern worldwide and it is important to take cognizance of its deficiency symptoms to ensure a healthy intake of vitamins. Taking cognizance of prevalent folate deficiency in routinely consumed foods in USA the RDA of 400 µg/day for adults and 600 µg/day for pregnant women has been recommended (https://ods.od.nih.gov/factsheets/Folate-HealthProfessional/#h10).

2.3 Folate deficiency in humans and perturbed C1 metabolism

Folates are one of the various vitamins which take part in metabolism as cofactors (Figure 2.1). Since humans are unable to synthesize folate synthesis its requirement is exclusively met from dietary intakes mostly from food derived from plants. The folate involvement ranges from biosynthesis and metabolism of nucleotides and amino acids (serine, glycine, histidine, and methionine) to pantothenate (vitamin B5) (Blancquaert et al., 2010). The concentration of folates differs in various body tissues and highest amount of folate is in liver (Table 2.3). The liver serves as storage site for folate and in general, the amount present in the liver is sufficient to meet the body's requirement of folate for 4-5 months in the event of consumption of folate deficient diet (Hercberg and Galan, 1992). However, prolonged suboptimal intake of folate or heavy alcohol consumption depletes the overall folate storage in liver (Steinberg and Campbell, 1979; Steinberg, 1984). The folate deficiency disturbs the metabolic homeostasis resulting in complications like cardiovascular diseases, dementia, cognitive dysfunction, depression and neural tube defects (Ansari et al., 2014). Other problems associated with deficiency are megaloblastic anemia (Li et al., 2003), higher risk on major depressive disorder (Papakostas et al., 2012), Alzheimer's disease (Seshadri et al., 2002), cardiovascular (Scott and Weir, 1996) and

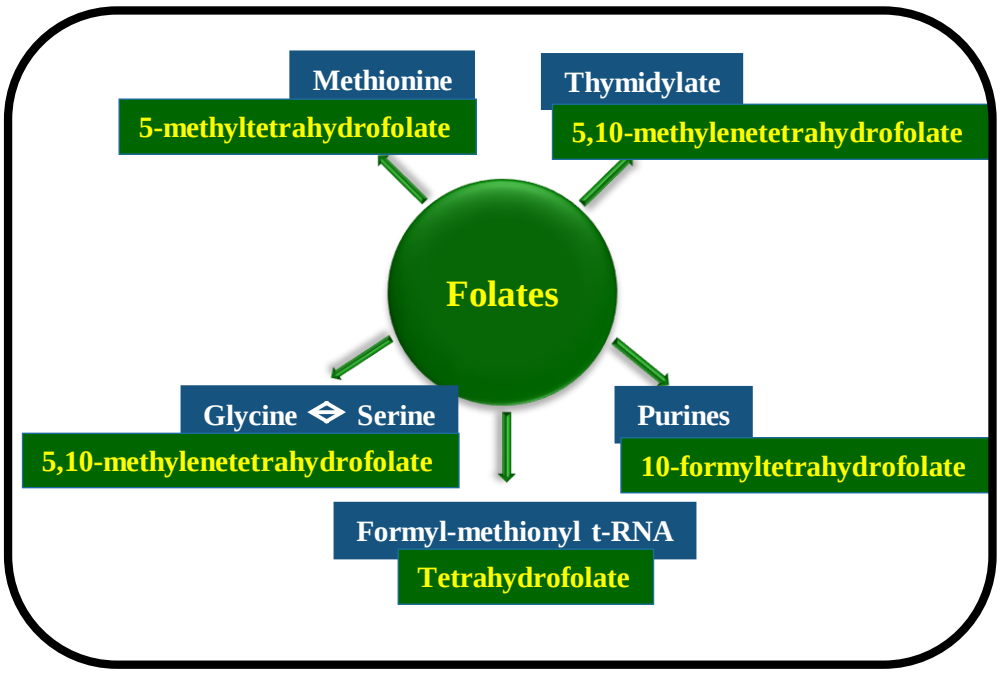


Figure 2.1. Schematic representation of role and participation of THF and its C1 derivatives in various one carbon transfer reaction.
Note:- The white fonts show the reaction or molecule requiring folate cofactors. Yellow fonts indicate the C1 substituted THF derivative involved in respective metabolic reactions.

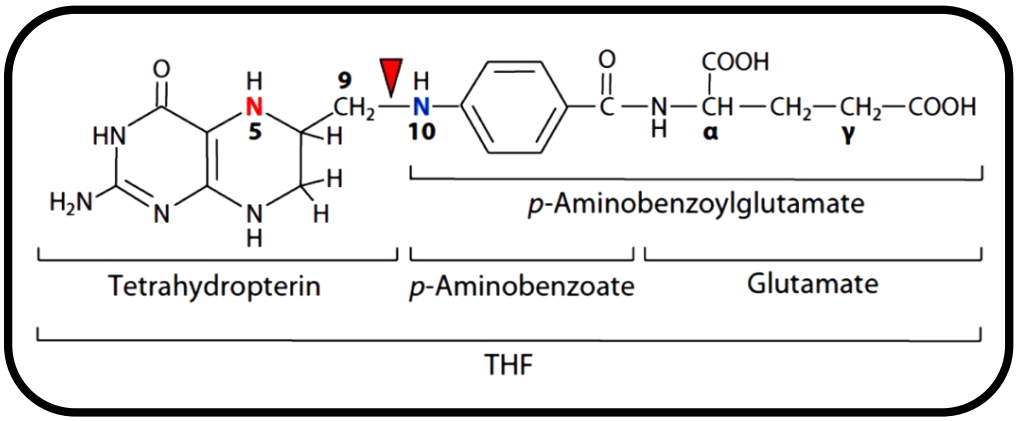


Figure 2.2. Structure of Tetrahydrofolate (THF). The red arrowhead at C9–N10 shows the labile bond prone to oxidative breakage. A polyglutamyl tail can be attached via the γ-carboxyl group of the glutamate moiety.

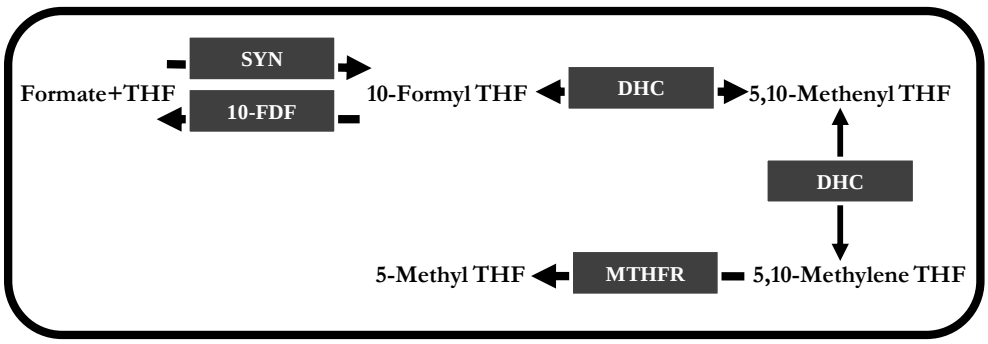


Figure 2.3. Enzymes involved in interconversions of various folate species. Abbreviations: SYN=10-formyl THF synthetase; 10-FDF=10-formyl THF deformylase (PurU in *E. coli*); DHC=5,10-methylene THF dehydrogenase/5,10-methenyl THF cyclohydrolase; MTHFR, 5,10-methylene THF reductase.
Figure adopted from Collakova et al. (2008).

Figure 2.4. Various dietary sources rich in folates. Tomato is reportedly not an enriched folate source. Therefore attempts are being made to biofortify it with folate.

Table 2.1 Average folate levels of superfoods (folate rich foods) and their contribution in providing the RDA of folate in standard daily serving.
Note:- data obtained from the comprehensive list by National Institutes of Health, Office of Dietary Supplements: **(http://ods.od.nih.gov/factsheets/Folate-HealthProfessional/)**

Food source	Folate per serving (µg)	Contribution to daily value (%)
Spinach, boiled, ½ cup	131	33
Brussels sprouts, frozen, boiled, ½ cup	78	20
Orange, fresh, 1 small	29	7
Black-eyed peas (cowpeas), boiled, ½ cup	105	26
Beef liver, braised, 3 ounces	215	54
Fish, halibut, cooked, 3 ounces	12	3
Broccoli, chopped, frozen, cooked, ½ cup	52	13

Table 2.2 Staple food items and their folate content. Data retrieved from Bekaert et al. (2008).
Note:- Daily value for folate is 400 µg per day for adults.

Food item	Folate content (µg/100 g) dry weight
Rice seeds	6-8
Wheat seeds	38-43
Maize seeds	19

Table 2.3 Normal folate levels in various body tissues
Data retrieved from **https://ods.od.nih.gov/factsheets/Folate-HealthProfessional/#h10,** **http://apps.who.int/iris/bitstream/10665/75584/1/WHO_NMH_NHD_EPG_12.1_eng.pdf** and Daly et al. (1995)
Note:- Folic acid conversion factor: 1 ng/mL = 2.265 nmol/L.

Tissue type	Folate content	Overall folate content in body
Liver	5-15 mg per liver	10-30 mg
Serum	6-20 ng/mL	
RBC	140-570 ng/mL	

coronary diseases (Stanger, 2004), stroke (Endres et al., 2005), and several (mainly colorectal) cancers (Friso and Choi, 2005).

2.3.1 Methylation cycle and folates

In the food, folate is predominantly present in polyglutamylated form. Post ingestion of food, polyglutamylated folates are converted to monoglutamate forms for proper absorption and distribution by the action of gamma glutamyl hydrolases (GGH) present in the jejunal brush borders. This is followed by the conversion of different folate vitamers to the main circulating folate form detected in blood plasma, i.e. 5-methylTHF by the action of methylenetetrahydrofolate reductase (MTHFR). 5-methylTHF is also most important folate vitamer as it provides methyl units to methyltransferases, which use a broad range of substrates, such as hormones, DNA, proteins, and lipids, as part of the 'methyl cycle' as reviewed by Scott et al., 2000 (Figure 2.5, 2.6). The folate deficiency results in disturbance of methyl cycle, leading to Hyperhomocysteinuria due to the insufficient activity of methionine synthase which converts homocysteine to methionine (with the help of both 5-methyltetrahydrofolate and vitamin B12). Thus, elevated homocysteine levels in the blood (which is indicative of not only folate deficiency but vitamin B_{12} as well) is believed to be linked with coronary and cardiovascular diseases.

Similar to folate deficiency, the excess intake of folate can also cause health hazards. The upper safe limit of folate intake is 1 mg/day and indiscriminate use of folate supplements can result in complications like cognitive decline, progression (if not repression) in certain types of cancers, reduced cytotoxicity of natural killer cells of immune system and masking of vitamin B_{12} deficiency symptoms (Ulrich and Potter, 2006). In *Caenorhabditis elegans*, folate, synthesized by its gut microbes was shown to reduce its life span (Virk et al., 2012). However, the studies documenting the health effects of excess folate intake are still preliminary in nature.

2.4 Structure and biosynthesis of Folate in plants

Folates are only synthesized by the bacteria and plants. Natural folate exists as *t*etra*h*ydro*f*olate (THF) and its derivatives. A THF molecule contains three moieties namely pterin, *p*ara-*a*mino*b*enzoic *a*cid (pABA), and glutamate (Figure 2.2) (Hanson and Gregory III, 2011). Folate biosynthesis is distributed in different compartments of a plant cell beginning with synthesis of pterin and pABA the primary precursors, their condensation into dihydropteroate (DHP) through multiple steps. It is followed by the addition of first glutamate molecule and subsequent reduction, generation of first tetrahydrofolate (THF) molecule (Figure 2.7).

The pterin is synthesized in the cytosol from GTP. GTP cyclohydrolase I (GCHI) carries out the first committed step of pterin biosynthesis (Basset et al., 2002) converting GTP to dihydroneopterin triphosphate (DHN-P3) which through further dephosphorylation and the action of DHN aldolase, is converted to 6-hydroxymethyldihydropterin (HMDHP). HMDHP is transported into the mitochondria and phosphorylated to produce HMDHP-P2 which is condensed with pABA. The GTP cyclohydrolase I (GCHI) activity is modulated by GTP and high GTP levels its catalytic activity in plants.

The synthesis of pABA is localized in plastids starting from conversion of chorismate to aminodeoxychorismate (ADC) and ultimately to pABA. These steps are catalyzed by aminodeoxychorismate synthase (ADCS) and aminodeoxychorismate lyase 1 and 2 (ADCL1, ADCL2) (Basset et al., 2004). *In viv*o the pABA molecule is largely present in its glucose esterified forms than the free pABA however the enzymes and reaction involved in reconversion remains to be deciphered (Hanson and Gregory III, 2011).

The HMDHP from cytosol and pABA from plastid are then transported into the mitochondria where HMDHP is converted to HMDHP-P2 and coupled with pABA in a reaction catalyzed by dihydropteroate synthase (DHPS) and forms dihydropteroate (DHP). DHPS is a bifunctional enzyme and is also known as hydroxymethyldihydropterin pyrophosphokinase/dihydropteroate synthase (HPPK/DHPS). The activity of DHPS is inhibited by the sulfa drugs (Yun et al., 2012). Since mammals do not synthesize folate sulfa drugs are used to prevent bacterial infection by blocking their folate biosynthesis. DHP is further coupled with a glutamate molecule with the help of dihydrofolate (DHF) synthase and forms DHF which is reduced to tetrahydrofolate (THF), a reaction catalyzed by bifunctional dihydrofolate reductase/thmidylate synthase (DHFR/TS) (Luo et al., 1993; Cox et al., 1999; Ravanel et al., 2001). In Arabidopsis, a cytosolic form of (HPPK/DHPS) is also reported which helps in conferring stress resistance in germinating seeds (Navarrete et al., 2012).

Post THF synthesis, THF and its derivatives are polyglutamylated by folylpolyglutamate synthases (FPGS), an enzyme which is present in both, plants and animals. FPGS adds one glutamate residue at a time through γ-carboxyl peptide bonds. There are 3 isoforms of FPGS in Arabidopsis which are localized in plastid, mitochondria and cytosol while in other higher plants mostly have only two isoforms, plastidial and mitochondrial (Ravanel et al., 2001; Mehrshahi et al., 2010). Folate polyglutamylation increases the suitability of folates during one carbon transfer reactions as they are preferred

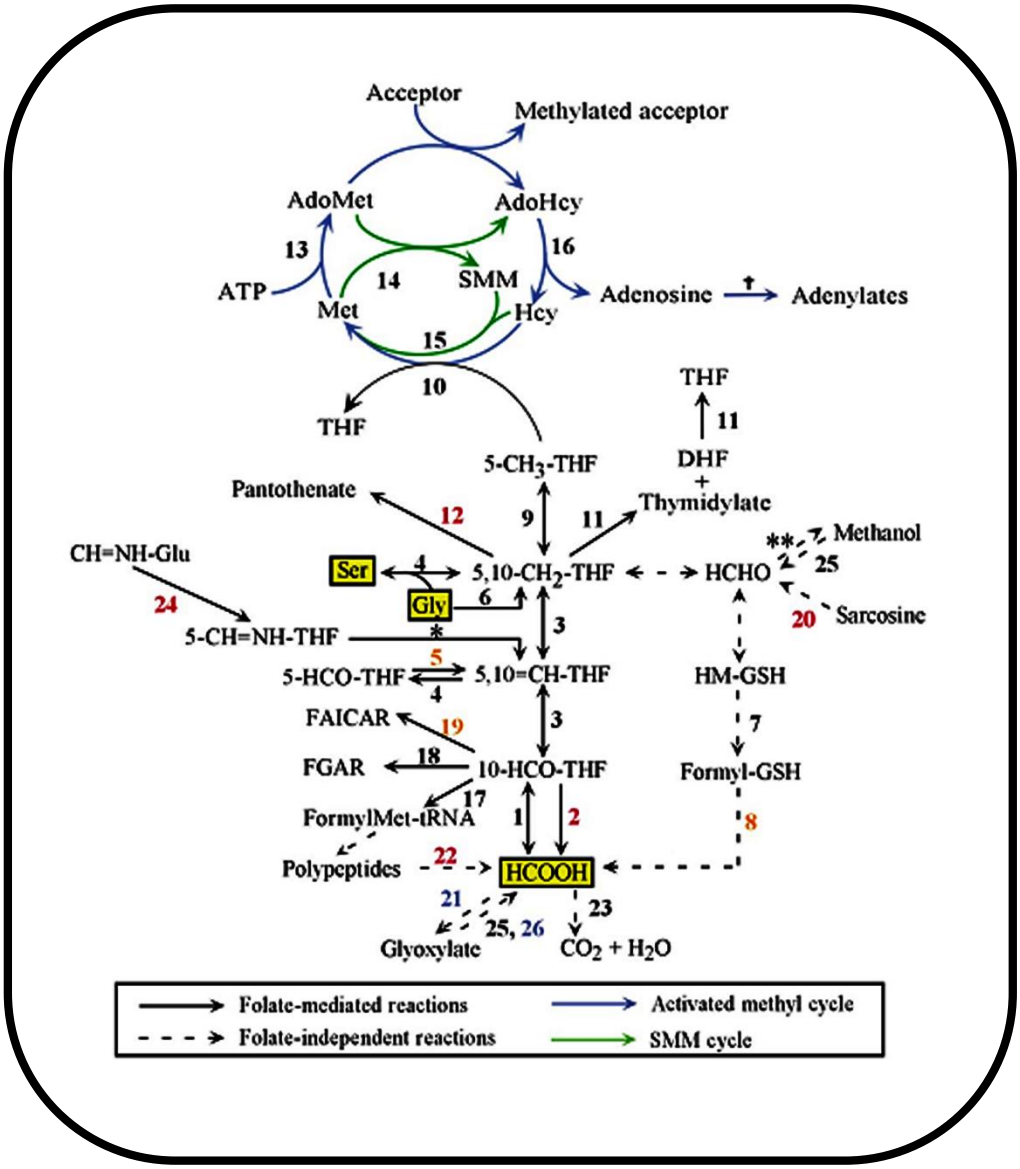

Acceptor
Methylated acceptor
AdoMet
AdoHcy
13
16
ATP
14
SMM
Adenosine
Adenylates
Met
Hcy
15
10
THF
THF
11
DHF
+
Thymidylate
5-CH3-THF
Pantothenate
12
9
11
CH=NH-Glu
Ser
4
5,10-CH2-THF
HCHO
Methanol
25
Gly
6
24
20
Sarcosine
5-CH=NH-THF
3
5
5-HCO-THF
5,10=CH-THF
HM-GSH
4
FAICAR
19
3
7
FGAR
18
10-HCO-THF
Formyl-GSH
17
FormylMet-tRNA
1
2
8
Polypeptides
22
HCOOH
21
23
25, 26
Glyoxylate
CO2 + H2O
Folate-mediated reactions
Activated methyl cycle
Folate-independent reactions
SMM cycle

Figure 2.5. The reactions of plant C1 metabolism as deduced from biochemical and DNA sequence evidence. The principal sources of C1 units are boxed and highlighted in yellow. Enzymes are numbered and are as follows- 1. 10-FormylTHF synthetase, 2. 10-FormylTHF deformylase, 3. 5,10-MethyleneTHF dehydrogenase, 5,10-MethenylTHF cyclohydrolase, 4. Serine hydroxymethyltransferase, 5. 5-FormylTHF cycloligase, 6. Glycine decarboxylase complex-P-protein, H-protein, T-protein, L-protein, 7. Formaldehyde dehydrogenase (GSH), 8. S-Formylglutathione hydrolase, 9. 5,10-MethyleneTHF reductase, 10. Methionine synthase (B12 independent), 11. Dihydrofolate reductase/thymidylate synthase, 12. Ketopantoate hydroxymethyltransferase, 13. S-Adenosylmethionine synthetase, 14. Methionine S-methyltransferase, 15. Homocysteine S-methyltransferase, 16. S-Adenosylhomocysteine hydrolase, 17. Methionyl-tRNA transformylase, 18. GAR Transformylase, 19. AICAR Transformylase/IMP cyclohydrolase, 20. Sarcosine oxidase, 21. Glyoxylate synthetase, 22. Polypeptide deformylase, 23. Formate dehydrogenase, 24. Glutamate formiminotransferase, 25. Catalase (peroxidatic activity), 26. Glyoxylate decarboxylase. The asterisk marks the reaction catalyzed by 5-formimino-THF cyclodeaminase (EC 4.3.1.4), for which there is as yet no biochemical or DNA evidence in plants. The double asterisk marks the reduction of formaldehyde to methanol, which has been shown to occur *in vivo* but for which the enzymatic basis in uncertain. The dagger denotes adenosine salvage reactions. Abbreviations: 5-CH=NHTHF, 5-forminoTHF; DHF, dihydrofolate; Hcy, homocysteine; Met, methionine; GSH, glutathione; HM-GSH, S-hydroxymethylglutathione; FGAR, formylglycinamide ribonucleotide; FAICAR, formamidoimidazolecarboxamide ribonucleotide; CH=NH-Glu, formiminoglutamate (Hanson and Roje, 2001).

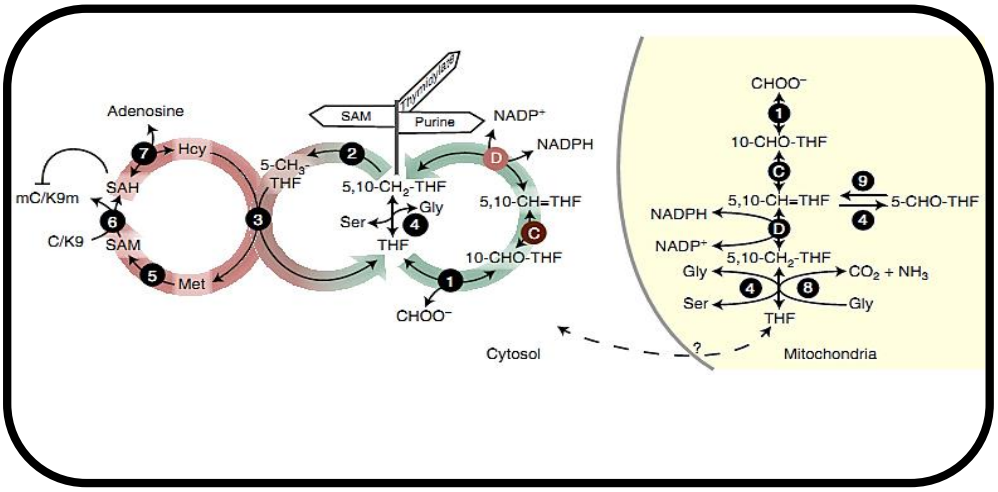


Figure 2.6. Schematic representation of plant SAM and folate metabolism in the cytosol and mitochondria. One-carbon enters the cytoplasmic folate cycle (green) either through formyltetrahyrofolate synthetase (1) or serine hydroxyl methyl transferase (SHMT) (4); Methylenetetrahydrofolate Dehydrogenase 1 (MTHFD1) reversibly interconverts 10-formyl-THF (10-CHO-THF) to 5,10-methylene-THF (5,10-CH_2-THF) by cyclohydrolase (C) and NADPþ-dependent dehydrogenase (D) activity. 5,10-CH_2-THF serves for thymidylate synthesis or is converted by methylenetetrahyrofolate reductase (2) to 5-methyl-THF (5-CH_3-THF), which enters the (methionine (Met) cycle (red) and serves for homocysteine (Hcy) remethylation to Met by methionine synthase (3). S-adenosylmethionine (SAM) synthetase (5) converts Met to SAM, which is further converted to S-adenosylhomocysteine

(SAH) (6) during methylation of cytosines, H3K9 and so on. SAH is a competitive inhibitor of methyltransferases (6) and is recycled to Hcy by SAH hydrolase (7). In mitochondria, one-carbon is transferred to THF during the oxidation of Gly by the glycine decarboxylase complex (8), but surplus of Gly due to photorespiration leads to consumption of one-carbon by SHMT during serine production30. 10-formyl-THF (5-CHO-THF), a byproduct of SHMT, is metabolized by mitochondrial 5-formyltetrahyrofolate cycloligase in order to re-enter the folate cycle (9). Shuttle of THF between mitochondria and the cytosol has been described in other organisms, but remains uncharacterized in plants. The numbers in the bracket indicates the enzyme reactions steps.
Figure adopted from Groth et al. (2016).

substrates over monoglutamate folates (Shane, 1989; Wagner, 2001). Also, protein binding is assumed to enhance the stability of folates and prevents their degradation. There are reports of 2-7 glutamate residues generally present in plant folates (Orsomando et al., 2005 and Akhtar et al., 2008); however, recently up to 17 glutamate molecules were detected attached to 5-methylTHF in papaya (Ramos-Parra et al., 2013). Polyglutamylated folates are not favored by folate carriers (Shane, 1989; Orsomando et al., 2005) and need to be monoglutamylated for movement across membranes which is achieved by the activity of γ-glutamyl hydrolase (GGH).

Removal of glutamate molecule from THF derivatives and pABA-Glu(n) (a folate degradation product) is achieved by γ-glutamyl hydrolase (GGH). Mammalian GGH are lysosomal and show highest activity in liver and kidneys; tissues with maximum folate storage (Galivan et al., 2000), where turnover of folates is high and folate need to be transported to various tissues as required. Lin et al. (1993) showed GGH in the extracts of six-days old pea cotyledon extracts act as an endopeptidase and noted that activity was localized in the cytosol. Later, Orsomando et al. (2005) showed that GGH has both endopeptidase and exopeptidase activity and is localized in the vacuole. Plants have one or more *GGH* genes, Arabidopsis and tomato have three *GGH* isoforms. It is assumed that GGH plays a role in maintaining folate homeostasis in plants (Akhtar et al., 2008). While, GGH1, GGH2 in tomato form active homodimers and heterodimers, GGH3 supposedly codes a non-functional protein. Also, GGH1 and GGH2 can complement the function of each other but only GGH2 is able to form monoglutamates while GGH1 forms di- or tri-glutamates.

2.4.1 Major regulatory points of folate biosynthesis

There are studies which postulate various models of regulation of folate biosynthesis (Scott et al., 2000; Waller et al., 2010; Hanson and Gregory III, 2011) (Figure 2.8). Biosynthesis of tetrahydrofolate must be highly regulated to maintain a balance between its demand and supply since many of the biosynthesis reactions are ATP dependent, none of it could be surplus (Scott et al., 2000). While many of the reactions of folate biosynthesis pathway are not yet fully understood, there are evidences that some of the intermediates of the pathway may regulate synthesis feedback or feed forward regulation (Waller et al., 2010). For example, dihydropteroate (DHP), product of the DHPS (dihydropteroate synthase) inhibits its activity indicating that DHPS activity may be closely governed by DHP concentration (Rébeillé et al., 1997). *In vitro* DHP also inhibits ADCS. However, since DHP and p-ABA pools are localized in mitochondria and plastid,

it is not likely that *in vivo* DHP inhibits ADCS activity (Hanson and Gregory III, 2011). It is reported that methionine acts as a repressor of SHMT and methylenetetrahydrofolate dehydrogenase activities, two reactions involved in the supply of one-carbon folate derivatives to the C1 metabolism (Scott et al., 2000). A dihydrofolate (DHF) reductase inhibitor methotrexate while causing overall folate depletion caused enhanced transcripts levels of cytosolic folylpolyglutamate synthase (*FPGS)* in Arabidopsis indicating that its expression is feedback regulated by polyglutamylation status of the cell. High concentration of polyglutamyl folate substrate exerts negative effect on (FPGS) in both mammalian and plant systems (Storozhenko et al., 2007b; de la Garza et al., 2007; Tomsho et al., 2008)) while the monoglutamyl THF and derivatives induce *FPGS* (Waller et al., 2010). An intriguing observation was made when *GCHI* and *ADCS* transgenes were co-expressed in tomato, the transgenic fruits accumulated high levels of folate and had increased expression of the downstream genes *dihydroneopterin aldolase, aminodeoxychorismate lyase*, and mitochondrial *FPGS* (*FPGSm*). In contrast the pABA overexpressing rice transgenic lines showed strong reduction in folate levels in the seeds (Waller et al., 2010; Storozhenko et al., 2007b).

It was speculated that the very key event of C1 metabolism, i.e. methionine biosynthesis which compulsorily requires THF derivatives must have some role in the regulation of its biosynthesis. This speculation was later substantiated by the finding that methionine and homocysteine had repressing and inducing effects on FPGS and DHFR activities (Lewandowska et al., 1996). Methionine is also known to repress the activities of SHMT and methylene tetrahydrofolate dehydrogenase. Considering that methionine is synthesized from homocysteine involving methyltetrahydrofolate (which is frequently the most predominant folate species detected in the plants) and is a precursor of S-adenosylmethionine (SAM, a key metabolite for most methyltransferase reactions), the possibility of association between methionine biosynthesis and methyltetrahydrofolate levels seems to be strong. Therefore, methionine seems to one of the key regulators governing the folate demand in a cell. Another factor presumed to regulate folate biosynthesis is light. Spronk and Cossins (1972) reported that the folate levels in green pea cotyledons were higher than etiolated tissue. They reasoned that this could be because of ongoing photorespiration in the cotyledons and associated glycine to serine conversion which is catalyzed by folate dependent SHMT (Okinaka and Iwai, 1970).

In essence, these reports indicate possible existence of a hitherto unknown, yet essentially diverse regulatory mechanisms of folate biosynthesis in different organisms

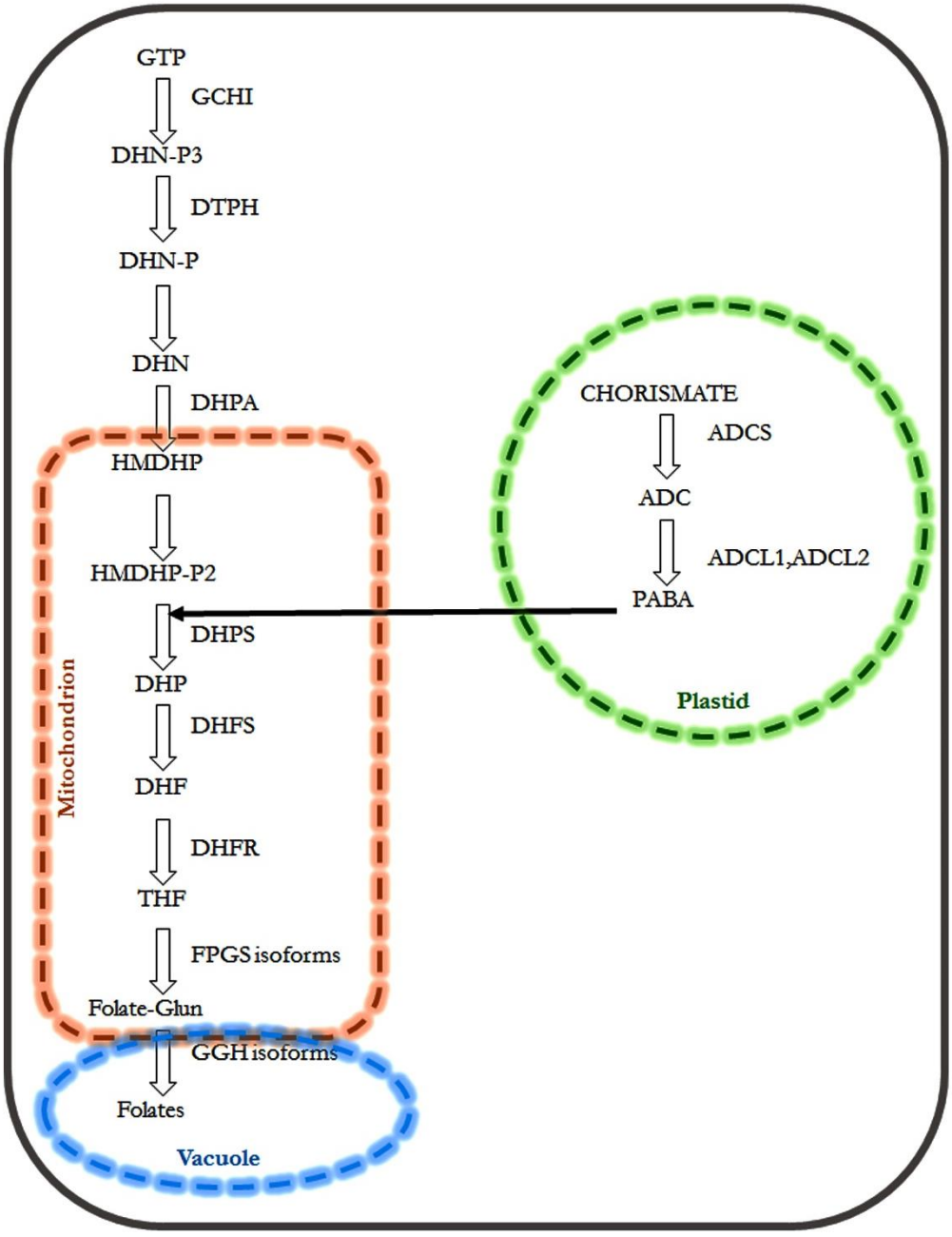


Figure 2.7. Folate biosynthesis pathway. Various cellular sites like cytosol, plastid and mitochondria are involved in pterin, pABA and THF assembly. Vacuoles sequester and store folates.

(Figure 2.8). It is also likely that different species may have different levels of folate regulation Hanson and Gregory III, 2011). It is therefore likely that in addition to few common regulatory steps each species may have independent levels of regulation to meet its folate requirement.

2.4.2 Folate breakdown

Folates are very labile molecules, susceptible to oxidation and photo oxidation. Among the folate vitamers THF and dihydrofolate (DHF) are the most labile molecules. Even at physiological pH, folates are not very stable when exposed to light or oxygenated environment (Hanson and Gregory III, 2011). Plants show elevated folate breakdown of approximately 10% per day after harvest. Since the oxidation products of folates are not detected in higher concentrations, therefore it is likely that breakdown products are efficiently salvaged or recycled (Orsomando et al., 2006). Unlike animal system folate degradation in plants is mainly by non-enzymatic cleavage and there is little evidence regarding existence of the plant folate catabolizing enzymes.

The first cleavage of a folate molecule occurs at the C_9-N_{10} bond resulting into pterin and p-aminobenzoylglutamate (pABA-Glu) or its polyglutamyl forms. Further oxidation of pterins ultimately generates pterin-6-aldehyde (Hanson and Roje, 2001). Later it was found that Arabidopsis and tomato tissues readily converted the breakdown products to folate synthesis precursors. pABA-Glu was hydrolyzed to p-aminobenzoate and glutamate, and dihydropterin 6-aldehyde was reduced to 6-hydroxymethyldihydropterin (HMDHP). Also, pABA-Glu hydrolase activity was reported in cytosol/vacuole and mitochondrial fractions of pea leaves (Bozzo et al., 2008). Pterin salvage reaction was NADPH dependent and later found to be catalyzed by an Arabidopsis gene (*At1g10310*) encoding a pterin aldehyde reductase (Orsomando et al., 2006; Noiriel et al., 2007).

2.5 Transport of folate precursors and derivatives

Folate transport mechanism is less clearly understood in plants. On the contrary, bacterial and mammalian folate transporters are well studied and characterized (Matherly and Goldman, 2003; Zhao et al., 2009). Since folate biosynthesis encompasses cytosol, plastid, mitochondria and folate sequestration occurs in vacuoles, hence, various precursor molecules like pABA, HMDHP, folate species like THF, THF-Glu(n) and salvaged molecules like pABA-Glu(n) etc. need to cross these compartments and thus, presence of transporters is obvious. Hanson and Gregory (2011) reviewed advances in the area of

plant-specific folate transporters and considering the folate biosynthesis, polyglutamylation and salvage reactions, concluded the possibility of minimum nine carriers in plants of which only three are characterized. pABA moves through simple diffusion across the organelles being weakly acidic in nature. Nevertheless, most of the free pABA is continuously converted into pABA glucose ester in the cytosol but stored in the vacuole. So far the transport molecules assisting this cytosol to vacuole pABA-glucose ester movement are not known (Quinlivan et al., 2003).

Pterin transport has been elucidated to a small extent with the help of homology searches. Based upon homology with the Leishmania folate biopterin transporter, nine Arabidopsis homologs were found, but only one was functional (Lemley et al., 1999; Bedhomme et al., 2005). Mitochondria being the sole site of folate biosynthesis, the presence of folates in other cellular compartments indicated that mitochondrial folate transporters (MFT) must be present in plants similar to the MFTs present in human and hamster cells (Perchiniak et al., 2007). So far from Arabidopsis three likely transporters AtFOLT1, At2g32040, and AtMRP1 have been cloned and characterized. Though, an Arabidopsis gene that (At5g66380, AtFOLT1) bears close homology to MFT was not functional *in planta* (Bedhomme et al., 2005). Another folate transporter, At2g32040 was shown to aid the uptake of mainly folates monoglutamate forms in plastids (Klaus et al., 2005). Mammalian multidrug resistance associated protein (MRP) subfamily which is an ATP binding cassette transporter family, use folate and its analogs as substrates (Zeng et al., 2001). The Arabidopsis MRP1 (At1g30400) showed the ability in transport of folic acid and methotrexate when expressed in yeast (Raichaudhuri et al., 2009).

Among all the classes of mammalian folate transporters viz. mitochondrial transporter, the reduced folate carrier, the intestinal proton–coupled folate transporter, various MRPs, and glycosylphosphatidylinositol linked folate receptors, only the mitochondrial and MRP transporters have reasonably close plant homologs (AtFOLT1 and AtMRP1, respectively) which are known to transport folates (Eudes et al., 2010). More insights into the transporters will help in elucidating the fluxes of precursors as well as folate derivatives in a cell that will help in designing better folate enhancement mechanisms.

2.6 Folate estimation in crops and vegetables

The bulk of food constituting daily diet in most of the modern day cultures is potato, rice, corn and wheat which are not folate rich and are inadequate to meet the folate

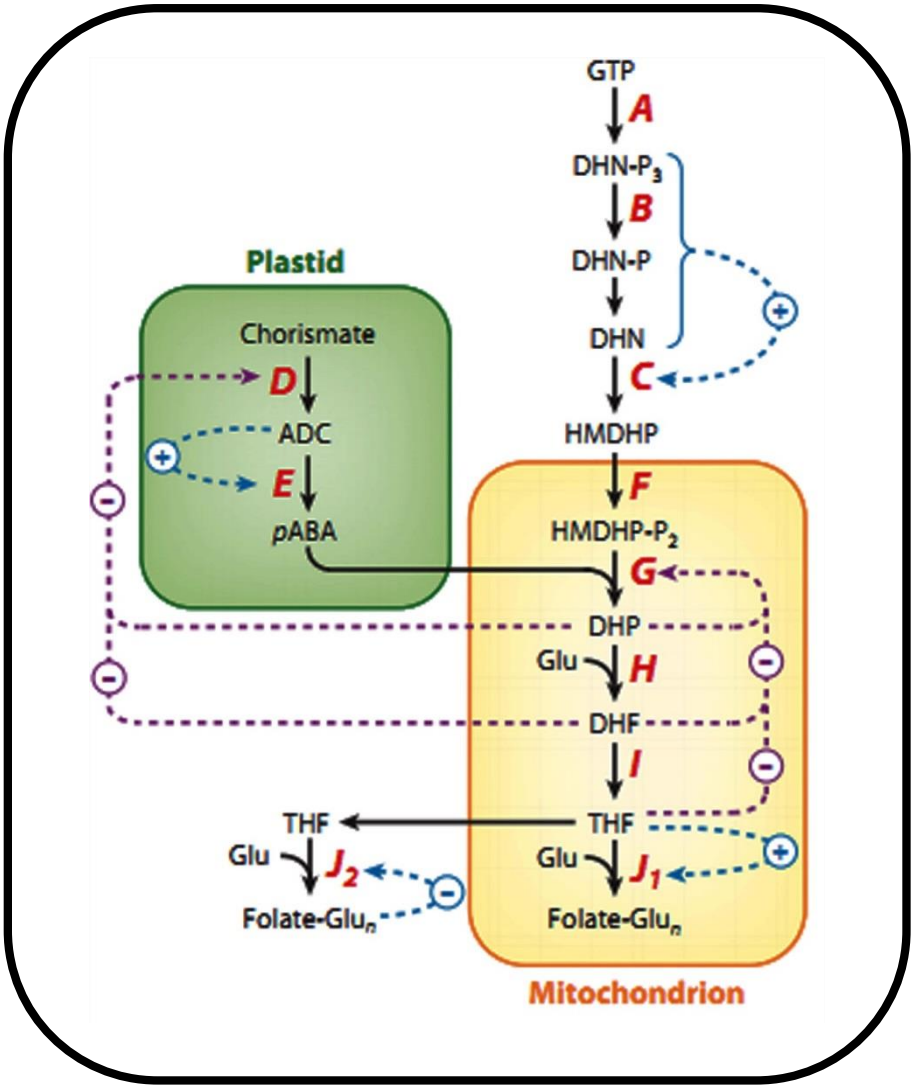


Figure 2.8. Probable regulatory points at the enzyme and gene levels in folate biosynthesis. Broken purple lines denote potential enzyme-level feedback inhibition mechanisms identified by in vitro studies. Broken blue lines denote gene-level feedback repression (minus signs) or feedforward activation (plus signs) mechanisms inferred from transcriptome data. Folate biosynthesis enzymes are shown by letters as in Figure 2. Dihydromonapterin has been omitted for simplicity. Abbreviations: ADC,aminodeoxychorismate; DHF, dihydrofolate;DHN, dihydroneopterin; DHP, dihydropteroate;Glu, glutamate; Glun, polyglutamate; HMDHP,6-hydroxymethyldihydropterin; P, phosphate;P2, diphosphate; P3, triphosphate; pABA,p-aminobenzoate; THF, tetrahydrofolate (Hanson and Gregory III, 2011).

RDA (http://ods.od.nih.gov/factsheets/Folate-HealthProfessional/). To ensure the proper folate intake, the understanding of its levels in commonly consumed foods is necessary. Many studies have been conducted with the objectives of finding fruits, vegetables and other edible foods rich in folate levels.

Several studies have examined the folate levels in commonly consumed foods within a nation or culture. The folate content of common breakfast items like bread, rice, pasta, spaghetti, cereals etc. was assessed by Pfeiffer et al. (1997), diet constituents of typical Italian food (Ruggeri et al., 1999), regular food items like frozen peas, frozen spinach, fruit juice, fillet, beef liver, apples, egg yolk were also analyzed for their folate content (Ndaw et al., 2001; Ringling and Rychlik, 2013).

Attempts were also focused towards understanding the natural diversity of folate level among the cultivars of a single or different crop species. Several potato and tomato germplasms (Goyer and Navarre, 2007; Iniesta et al., 2009), beans, pulses and lentils (Han and Tyler, 2003; Jha et al., 2015) have been analyzed for folate levels. Various strawberry and wheat varieties were also screened for folate content and considerable variations were found among them (Stralsjo et al., 2003; Piironen et al., 2008). The examination of spinach varieties for folate levels revealed that despite being a very rich source of folate spinach leaves showed 3-fold varietal differences (Shohag et al., 2011). In addition several studies also examined the folate losses due to culinary practices, industrial food preservation techniques and folate extraction protocol itself (Stea et al., 2007; Delchier et al., 2013 Lerma-Ortiz et al., 2016) (Table 2.4). Though folate estimation methods differ among various laboratories, yet spinach, beans/pea/lentils, citrus fruits, green leafy vegetables, liver and fish meat have been repeatedly recognized as folate rich foods (Table 2.1).

2.6.1 Folate fortification and Enhancement in Crop Plants

During last decade considerable attention has been given to folate deficiency symptoms and implementation of food fortification measures such as mandatory addition of folic acid in cereals (food fortification). In addition, distribution of folic acid supplements through primary healthcare centers and awareness programs to inculcate healthy eating habits has also been carried out to ensure adequate folate intake through dietary sources. However, there are many practical difficulties in eradicating folate deficiency particularly in underdeveloped countries as well as in the areas where a monotonous type of diet comprising starchy food items (like potato, rice, maize or cassava) is consumed. Therefore, during the last decade special efforts have been directed towards

enhancement of the folate levels using metabolic engineering in popularly consumed grain and fruit crops.

These efforts stem from the perspective that in addition to assessment of naturally folate rich food crops/items and artificial fortification of food it is equally important to enrich the staple diet by modifying the biosynthesis machinery to meet the ever increasing food demands of population. Food fortification (flour, rice, bread) with folic acid or intake of folic acid pills are not feasible approaches in developing countries due to infrastructural constraints and lack of proper public distribution systems. Moreover there are chances that over-optimal intake of folic acid may causes an altogether different set of complications (Lucock, 2000; Quinlivan and Gregory, 2003).

The improvement of folate content in crops via. biofortification is a more cost effective approach as the investment in its development is a one-time process and does not lead to recurring expenditure year after year as in case of food fortification. Metabolic engineering, molecular marker assisted breeding, and mapping of high folate QTLs are preferred approaches for nutritional enhancement. Many strategies have been tested to improve the natural folate levels of several crop species like rice, corn, tomato etc. Metabolic engineering aims to enhance folate by higher production of the intermediate metabolites, facilitating the end product stability, directing the fluxes or controlling the rate of catabolism are a few approaches that can be used.

2.6.1.1 Metabolic engineering of pterin and pABA levels

Two main genes of the biosynthesis pathway targeted for folate enrichment are *GCHI (GTP Cyclohydrolase I)* and *ADCS (Aminodeoxychorismate synthase)* which catalyzes synthesis of pterin and pABA (para-aminobenzoic acid) respectively, the primary precursors involved in folate biosynthesis. It was hypothesized that pterin biosynthesis governed by GCHI in plants is a rate limiting step; and therefore overexpressing *GCHI* by transgenic means could be a potential folate enhancement strategy (Hossain et al., 2004). This was achieved by overexpressing bacterial (*E. coli*) *GCHI=folE* (free from feedback inhibition) under the CaMV 35S promoter in Arabidopsis which though lead to a massive increase in pterin levels in the leaves of transgenic lines but only a modest enhancement of folate levels by 2-4 fold .

Another group used overexpression of mammalian *GCHI* in tomato using the fruit ripening specific tomato E8 promoter. It was assumed that mammalian *GCHI* is subjected

Table 2.4. Various studies showing conducted for folate estimation in various food items and plant food sources.
Note:- Trienzyme= use of protease, amylase and conjugase, Only conjugase=single-enzyme treatment.

Samples	Extraction	Detection	Study
White bread, wheat bread, rice, and spaghetti.	Trienzyme	HPLC-diode array detector	Pfeiffer et al. (1997)
Yeast, spinach, beef liver, beef fillet and peas, pig's liver (CRM 487) and milk powder (CRM 421).	No enzyme, only conjugase, trienzyme	HPLC-UV VIS	Ndaw et al. (2001)
Camembert, rye bread, brown bread, baguette, wheat germs, spaghetti, chickpeas, pea and pea soup, savoy cabbage, spinach, cauliflower and dry vegetable mix (BCR 485).	Chicken pancreas+rat serum conjugase	Stable isotope dilution assay (SIDA), HPLC-MS/MS	Ringling and Rychlik (2013
Cabbage, chard, amaranth, spinach, silverbeet, watercress, mustard green, water convolvulus, mungbean sprouts, coriander, broccoli, snow pea sprouts, boxthorn, garlic chives, alfalfa, basil, and spring onion.	Chicken pancreas conjugase	Microbiological assay	Iwatani et al. (2003
Potato cultivars	Trienzyme	Microbiological assay	Goyer and Navarre (2007)
Tomato cultivars	Trienzyme	HPLC-fluorescence detector	Iniesta et al. (2009
Beans, pea and lentils	Trienzyme	Microbiological assay	Han and Tyler (2003)
Strawberry cultivars	Chicken pancreas conjugase	Radio protein binding assay (RPBA)	Strålsjo et al. (2003)
Spinach cultivars	Rat serum conjugase	HPLC-UV	Shohag et al. (2011)
Beans, pea, lentil and chickpea cultivars	Trienzyme	UPLC-MS/MS	Jha et al. (2015)

to feedback regulation by an inhibitory protein which plants lack, and though its end product tetrahydrobiopterin is also inhibitory, it not reported in plants and therefore feedback inhibition was not expected in the overexpression lines. This strategy too, though yielded considerable increase in pterin pool (3-140 fold) but only 2 fold enhancement in the folate levels. However, when these transgenic fruits were supplied with pABA through stalk, there was 10 fold increase in folate content (Diaz de la Garza et al., 2004). This lead to the presumption that inadequate concentration of pABA may have hampered the increase in the levels of folate in transgenic lines over-expressing *GCHI*. Based on the outputs of these transgenic efforts; the possible lacunae in these efforts were considered and new alternatives were formulated for folate enhancement (Basset et al., 2005). Since *GCHI* overexpressing lines showed extremely depleted pABA pools, therefore Basset et al. (2004) suggested that either pABA feeding to these lines or supplying it by overexpression of *ADCS* gene may boost the folate levels.

Taking cue from these studies, the follow-up study used generation of two transgenic lines, one overexpressing mammalian *GCHI* (for pterin) and another Arabidopsis *ADCS* (for pABA) genes in tomato. The two traits were combined by crossing the above transgenic lines resulting in almost 25 fold higher folate levels in ripe tomato fruits (Diaz de la Garza et al., 2007) than the control fruits, yet a very high amount of pterins and pABA also accumulated in the fruit. Concurrently, similar success was achieved in rice when Arabidopsis *GCHI* (G lines) and *ADCS* (A lines) genes were expressed alone and together (GA lines) under rice globulin (Glb-1) and rice glutelin B1 (GluB1) promoters respectively to exploit the natural feedback regulation of these genes and avoid substantial accumulation of pABA and pterins. It was found that only GA lines showed high folate accumulation ranging from 15-100 times more folate than that of WT. Though these lines accumulated excess pABA but that was well within the safe daily intake limit (Storozhenko et al., 2007b). In another study the *GCHI* expression in lettuce also resulted in some increase in folate (Nunes et al., 2009).

2.6.1.2 The overexpression mediated enhancement is not universal in application

Similar usage of transgenic approach to increase the folate levels by overexpression of *GCHI* and *ADCS* genes failed to boost folate levels in some other crops (Blancquaert et al. (2013). This failure highlighted that the folate regulation in different crops appears to be different and needs more intense investigations (Storozhenko et al., 2007). Currently the regulatory points in folate biosynthesis are not fully known, which is evident by the gene overexpression studies in potato and Arabidopsis which did not yield the elevated

folate levels like in tomato and rice. In potato and Arabidopsis there appears to be an altogether different type of regulation which is yet be deciphered (Morandini, 2013).

It is now believed that mere overexpression of the genes of very initial and committed steps of pathway is not sufficient and more information about channeling these excess intermediates towards the site of folate synthesis (i.e. mitochondria) and of pterin and pABA transporters is required (Morandini, 2013). An alternative approach where the newly synthesized folates is directed to a metabolically inert compartment such as vacuole also can be used for enhancing its accumulation. Similarly, controlling the breakdown/catabolism or favoring the formation of more stable folate form are few other approaches which can also be used (Basset et al., 2005). Since only limited information about plant folate transporters is available there is a lack of efforts made to use transport channels for understanding the folate flux within a cell.

The role of vitamins is extremely important for plants too, and they synthesize them to support their own metabolism. Thus, vitamin biosynthesis is not a constitutive process but a regulated one. The biosynthesis of vitamins is regulated according to the developmental needs like embryo development, nitrogen use efficiency during seedling development, salt stress or light regulated in photosynthetic tissues. It must be noted that plants do not synthesize surplus vitamins by compromising/disturbing other metabolic channels, but only as and when required to meet environmental challenges (Smith et al., 2007).

Currently limited information is available regarding regulation of folate biosynthesis by using T-DNA insertion or RNAi silenced lines of the key pathway genes. Folate biosynthesis-metabolism and its regulation in plants is still an emerging field and it can be better understood by study of mutants and natural genetic variant in folate levels. The use of natural germplasm is very important as they can provide information which is hitherto unknown and folate levels can be correlated with natural genetic polymorphism.

2.7 Introgression lines (ILs) as useful resource for elucidating the trait associations

The potential of broad genetic variability remains underutilized due to obstacles like hybrid sterility and epistatic interactions in the segregating F2 population which make it very difficult to associate a trait with a given genic region. To overcome this limitation tomato introgression lines were created by Eshed and Zamir (1994a) by crossing *Solanum lycopersicum* cv. M82 with *Solanum pennellii* (LA0716) and marker assisted repeated backcrossing with M82 to ensure that only a small defined fragment of *S. pennellii* existed

in the M82 genome. This resulted in creation of 76 ILs representing the whole *S. pennellii* genome in the form of overlapping segments of its chromosomes. The ILs are near isogenic lines (NILs) to each other except the fragment from *S. pennellii* and therefore any trait variation between them and recurrent parent can be associated with genes underlying the introgressed fragment of the wild relative (Figure 2.9). The power of ILs in discovery and characterization of genes regulating the traits of agronomical importance lies in the fact that they contain exotic chromosomal fragment from wild relatives who have a rich genetic base and remain, to some extent underexploited (Zamir, 2001).

Since their creation, tomato ILs have been extensively used worldwide and mostly for QTL analysis through bin mapping (Figure 2.10). A few examples are QTLs for total soluble solid content in tomato, identified on chromosome 1, 5 and 7 (Eshed and Zamir, 1994b). Similarly, the BRIX QTL (BRIX9-2-5) was characterized by Fridman et al. (2000) which is known to increase glucose and fructose contents in cultivated tomato fruits in various genetic backgrounds and environments (Elliott et al., 1993). BRIX9-2-5 was delimited to a single nucleotide polymorphism-defined recombination hotspot of 484 bp spanning an exon and intron of a fruit-specific apoplastic invertase. They suggested that the differences between the Brix9-2-5 alleles of the two species were associated with a polymorphic intronic element. Ascorbic acid QTL was fine mapped to chromosome 9 and shown to be consisted of multiple regions (Stevens et al., 2008). It was associated with monodehydroascorbate reductase (MDHAR) activity and tolerance to chilling stress. Rousseaux et al. (2005) identified a total of 20 QTLs including five for antioxidant capacity of the water-soluble fraction (TACW) (ao), six for ascorbic acid (aa), and nine for total phenolics (phe). Some of these QTL (ao6-2, ao6-3, ao7-2, ao10-1, aa12-4, phe6-2, and phe7-4) increased levels as compared to the parental line *L. esculentum*. However, the traits largely fluctuated with the environmental conditions and their regulation appears to be complex with strong environmental effects. Table 2.5 shows a list of traits and identified determining QTLs.

Apart from identifying QTLs, IL population has undergone various modifications in order to increase the mapping resolution of bins (Alseekh et al., 2013; Ofner et al., 2016). 285 smaller sub-lines from original 76 ILs were developed by Alseekh et al. (2013). Further increasing the quest for resolution, a Backcrossed Inbred Line (BIL) population is recently developed by Fulop et al. (2016).

Similar populations have also been developed containing introgressions of other wild Solanum species (*S. lycopersicoides*, *S. habrochaites*, *S. neorickii* and *S. pimpinelifollium, S. sitiens, S. peruviavum, S. chmielewskii* and *S.cheesmaniae*) into modern cultivars. There are also exotic libraries of wild species introgressed into other crop species, including rice, barley, soybean and pepper. Moreover, analogous approaches have been used in mammalian systems; several different populations of congenic or chromosomal substitution strains of mice have been generated and genotyped (Fernie et al., 2006). The use of ILs continues and is a promising field for mapping of useful traits.

2.8 Mutants in the folate biosynthesis pathway

There is a substantial lack of folate pathway mutants. Till now only few mutants related to folate metabolism have been identified and these too were discovered, unexpectedly by observing the unusual phenotypes during screening of mutagenized seedlings in Arabidopsis. The available folate pathway mutants can be classified in two general categories-(1) Insertional mutants (T-DNA insertion lines) and (2) chemically mutagenized individuals with point mutations. Most of the folate pathway genes have been characterized using Arabidopsis insertional mutants and the latter category has not been explored well. There are no reports of TILLING (Targeting induced Local Lesion IN Genomes) in folate pathway genes in plants to the best of our knowledge. Though transgenic approaches have focused on overproduction of pterin and pABA however the, stabilization of the folate through polyglutamylation is equally essential. Therefore, existing information about the genes regulating folate polyglutamylation (*folylpolyglutamate synthase*) and folate deglutamylation (*Gamma glutamyl hydrolase*) genes is reviewed in following sections.

2.8.1 Folylpolyglutamate synthases and their roles

Polyglutamylation of THF and its C1 derivatives is achieved by folylpolyglutamate synthase (FPGS). Polyglutamylated folates are preferred by the interacting enzymes contributing to 1-carbon transfer reactions and in fact, this interaction is more essential than optional. For example, the cobalamin-independent methionine synthase in plants, yeast and *Euglena gracilis* can only utilize polyglutamylated forms of 5-$CH_3H_4PteGlu_n$ ($n \geq 3$) (Ravanel et al., 1998; Cherest et al., 2000). The presence of a glutamate tail impedes the movement of folate polyglutamates across hydrophobic barriers by increasing their anionic nature (Appling, 1991) which perhaps helps in retention of folates in a particular compartment.

Table 2.5. The prevalence and diversity of traits and associated QTL identified by the *S. pennellii* ILs

QTL number	Trait number	Trait measured	Reference
36	6	Carotenoids	McQuinn and Giovannoni, unpublished
18	2	Fruit phenotypes	White and Giovannoni, unpublished
6	1	Vitamin C	McQuinn and Giovannoni, unpublished
584	101	Primary metabolites	Y Semel, unpublished
889	74	Primary metabolites	Schauer et al., 2006
841	35	Morphology and yield	Semel et al., 2006
88	23	Volatile compounds	Tieman et al., 2006
20	3	Antioxidants	Rousseaux et al., 2005
82	9	Sugars and acids	Baxter et al., 2005
81	9	Metabolites, brix and fruit weight	Causse et al., 2004
30	8	Leaf morphology	Holtan and Hake, 2003
16	1	Intensity of red color in ripe fruit	Liu et al., 2003
104	6	Yield related traits	Eshed and Zamir, 1995

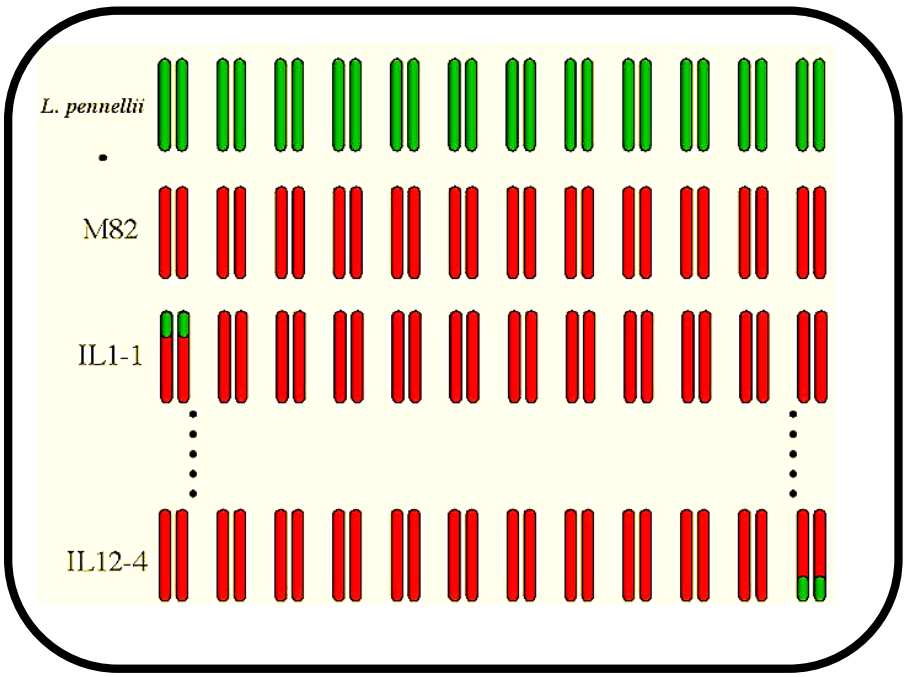


Figure 2.9. Schematic representation of introgression of tomato wild relative *S. pennellii* chromosomal segments (green) into the *S. lycopersicum* cv. M82 (red). Each introgression line (IL) contains a different length of segment of *S. pennellii*. Through repeated backcrossing, the 76 ILs are near isogenic to each other and M82. Picture adapted from The IL story (**http://zamir.sgn.cornell.edu/Qtl/il_story.htm**).

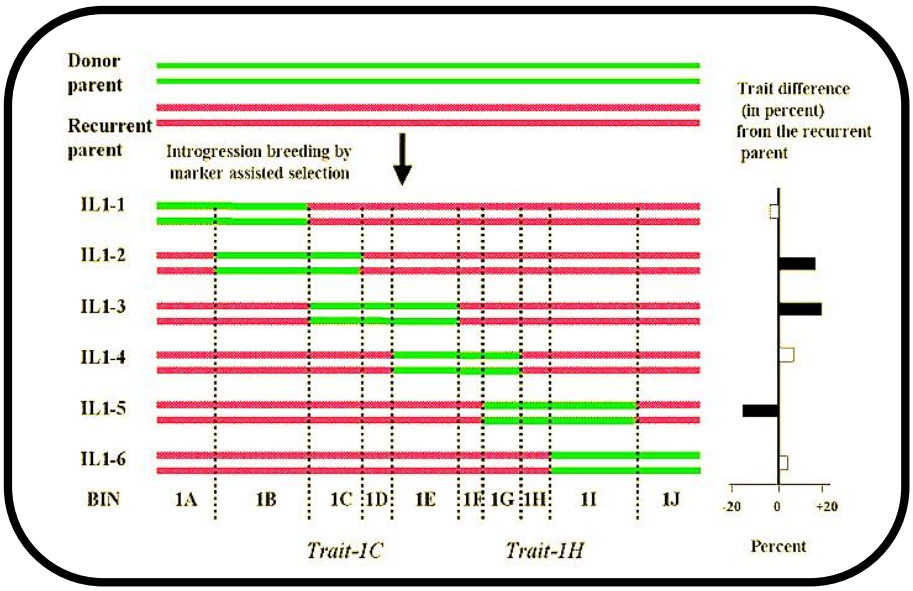


Figure 2.10. Schematic representation of bin mapping. Each IL contains a different length of donor chromosome and two ILs can have overlapping segments. The non-overlapping portion constitutes a bin with a unique donor parent composition. Bin mapping helps in identifying trait QTL. A trait QTL is mapped by phenotyping all lines in randomized replicated trials and presenting the results as percent difference from the recurrent parent M82 (0 value; black bars indicate a significant difference from the common control; empty bars indicate non-significant differences). Combined analysis of the data for all lines defines a QTL that increases the phenotypic value to bin 1C and a reducing QTL to bin 1H. The mapping scheme presented here was adopted from actual tomato data (Fridman et al. 2002). Picture adapted from The IL story (http://zamir.sgn.cornell.edu/Qtl/il_story.htm).

Firstly, rat liver FPGS were isolated and characterized by Mcguire et al. (1980) showing that enzyme concentration was inversely related to the glutamate chain length and postulated a self-regulatory significance of this association. Li et al. (1992) studied higher plant polyglutamates and showed that broccoli leaf formyl and methyl folates were mainly (80%) diglutamylated while methylene and unsubstituted leaf folates were mainly (78%) hexaglutamylated which again indicates towards the presence of need based self-regulatory cues for FPGS action in cells. The obvious preference of Serine hydroxymethyltransferase (SHMT) (which participates in glycine to serine interconversion during photorespiratory buildup of glycine) for 5,10-methylene THF polyglutamates also suggests that polyglutamylation in plants is need based and depends on preferences of the interacting enzyme (Ravanel et al., 2001).

2.8.1.1 Characterization of wild type FPGS and *FPGS* mutants

Imeson and Cossins (1991) characterized the FPGS isolated from 14 days old pea seedlings and showed that higher plant FPGS were similar to bacterial and mammalian FPGS with a molecular weight 68000 Da, and ATP and THF dependency (Cossins, 2000). Three isoforms of FPGS were characterized in Arabidopsis by Ravanel et al. (2001) encoded by three genes namely plastidial, mitochondrial and cytosolic forms (*FPGS1, FPGS2* and *FPGS3*). Their functions are redundant, and mutant phenotypes can be observed only in double mutants except *fpgs1* (*FPGSp*) mutant that has short root (Mehrshahi et al., 2010; Srivastava et al., 2011; Jiang et al., 2013). Recently, FPGS has been shown to affect lignin quantity and composition in maize brown mid rib (*bm4*) (Li et al., 2015; Srivastava et al., 2015) by usage of mutator transposon (*Mu*) insertion mutant and T-DNA insertion mutants respectively.

Point mutations in folate pathway genes and their characterization is very limited and serendipitous. For example, the mutation in *fpgs1* gene that is located at splice junction releases chromatin silencing on a genome-wide scale as it is indirectly essential for DNA and histone methylation (Zhou et al., 2013). The mutation in another *fpgs1* mutant (*moots koom 2=mko2*) which harbours a C to T transition in the 14th exon leading to R445Q change in the C-terminal domain of protein caused exhaustion of root apical meristem soon after germination. This indicates that folate is needed for stem cell specification and continuation of indeterminacy of root apical meristem (Reyes-Hernández et al., 2013). These mutations were isolated through forward genetic screening of mutagenized population bearing some unusual phenotype, and the underlying gene was identified only after fine mapping/ deep sequencing of the individuals. Currently there is also no

information on naturally occurring variations in the folate pathway genes among accessions of various crop plants.

2.8.2 *Gamma glutamyl hydrolases* (γ- glutamyl hydrolase)

Gamma glutamyl hydrolases (γ- glutamyl hydrolases or GGH) were isolated as early as 1940s from chicken pancreas (Kazenko and Laskowski, 1948; Dabrowska et al., 1949). Subsequently GGH were purified from many sources. While bovine hepatic GGH was characterized by Silink et al. (1975), GGH from rat intestinal mucosa were isolated by Elsenhans et al. (1984) and further identified and characterized in mammalian systems by Halsted (1989). GGH are peptide bond cleaving enzymes that play a very important role in folate metabolism. After the dietary folate is ingested by animals, it is converted from polyglutamyl to monoglutamyl forms before absorption. This conversion takes place in jejunal brush borders and animal GGH are lysosomal in location. Since its discovery, GGH has been extensively studied in relation to various types of cancers and association of GGH activity with the disease condition and treatment (Schneider and Ryan, 2006; Kawakami et al., 2008).

2.8.2.1 Plant γ- glutamyl hydrolases

The role of GGH in plants remained relatively obscure till late 1990s. In pea cotyledons the cellular localization of this enzyme was suggested to be cytosolic (Lin et al., 1993) while Orsomando et al. (2005) showed that the GGH activity was confined in the vacuoles. Most of the eudicots have 2 or 3 GGH (tomato has three GGH isoforms) while monocots have one (Akhtar et al., 2008). The common characteristics between mammalian and plant GGH are conserved catalytic residue (Cys-110 and His-220 except in soybean and tomato GGH3), conserved N-glycosylation motif near the N terminus; a glycoprotein characteristic feature and vacuolar (lysosomal in mammals) localization (Galivan et al., 2000; Orsomando et al., 2005; Akhtar et al., 2008). Akhtar et al. (2010) showed that overexpression of *GGH* in vacuole caused excessive deglutamylation and approximately 40% reduction in total folate content. They also observed that RNAi mediated silencing of either *GGH1* or *GGH2* alone did not cause any effect on total folate content. T-DNA disruption of one *GGH* and RNAi silencing of the other *GGH* together resulted in 34% increase in total folate content. In fact, their functions *in planta* appear to be complementary to each other. Presently, there is no mutant information available for *GGH* gene except for the RNAi or T-DNA insertion lines used by various groups while pursuing specific research objectives (Akhtar et al., 2010). Variations are crucial units for adaptation and also, essential tool for TILLING which identifies the genic variations. The rate of naturally

occurring variations is very slow and mutagenesis of plants followed by TILLING scales the identification of artificially created variations/ mutations in a population for functional genomics.

2.8.3 TILLING/EcoTILLING as a tool to isolate folate pathway gene variants

Targeting Induced Local Lesions IN Genomes (TILLING) is a reverse genetics technique, originally developed in plants for identification of point mutations in genes for which sequences were available (McCallum et al., 2000). Since the first report of the TILLING, there have been many changes to make it more sensitive, high-throughput, user friendly and low cost except for the generation of large number of mutagenized lines which is still a bottleneck. TILLING is a better technique than RNAi, T-DNA insertion or overexpression methods since it yields both loss and gain of function mutants and after thorough characterization useful traits can be readily transferred to the cultivars of choice. Mutant lines that result from chemical/physical mutagenesis are not transgenic which ensures immediate release and wide acceptability.

TILLING approach has been successfully applied to many organisms including both, animal and plant systems. For example TILLING was used to investigate different targets 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) and lotus (Perry et al., 2003) etc. Recently, TILLING in tomato has generated much interest and 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., 2010).

Basically, TILLING involves- (1) mutagenesis treatment of population using chemical or physical mutagens, (2) DNA isolation from the mutagenized population, (3) knowledge of genomic region to be TILLed and (4) detection of mutation and study of its effects.

2.8.3.1 Preferred modes of creating a mutagenized population

Mutagenesis using physical mutagens like fast neutrons and gamma irradiation can result in big indels (Li et al., 2001; Sato et al., 2006) and therefore is not suitable for any PCR based detection approach. These methods have been successfully used in crop breeding for forward screening of beneficial characters and direct varietal release. On the other hand, point mutations can result in many alleles of the same gene instead of a knockout which is the basic idea of functional genomics (Henikoff et al., 2004). Alkylating chemicals like ethyl methane sulphonate (EMS) or N-ethyl-N-nitrosourea (ENU) are

better suited for TILLING and create characteristic GC-AT changes in the genome (McCallum et al., 2000; Greene et al., 2003). There are only ca. 5% truncations reported for EMS mutagenized population (Greene et al., 2003).

2.8.3.2 DNA isolation and pooling

Mutagenesis step is followed by isolation of genomic DNA from the mutagenized population (mostly in M_2 generation to ascertain the presence of heritable changes and thus indicative of germ cell mutagenesis). To reduce the overall cost, number of PCR reactions and subsequent steps, the DNA is pooled, amplified using the primer of interest and mutations are detected. DNA pooling has been subject to many changes depending upon the sensitivity of the detection method. It has progressed from 5 fold (McCallum et al., 2000), 8 fold (Colbert et al., 2001; Perry et al., 2003 and Cooper et al., 2008), to 96 fold based on the limits of detection (Tsai et al., 2011; Guo et al., 2015) of a particular method.

2.8.3.3 Mutation detection methods

Sensitivity of detection methods varies greatly and has a profound effect on detection and thereby a major deciding factor of pooling depth. The common detection methods are high resolution denaturing PAGE (LI-COR platforms), denaturing HPLC (dHPLC), capillary electrophoresis (CE), high resolution melt (HRM) analysis and recently, next generation sequencing (NGS). These techniques are different from each other with respect to skill, cost, throughput and detection limits.

The LI-COR platforms of polymorphism detection utilize high resolution denaturing PAGE separating the IR (Infra-Red) dye end labeled and CEL1 digested PCR products. CEL1 is an endonuclease isolated from celery which is specific for DNA mismatches and efficiently cuts the 3′ side of the mismatch site in a heteroduplex (Oleykowski et al., 1998). CEL1 mismatch digestion can detect deletions, insertions and base substitutions. The fragments resulting from CEL1 digestion are electrophoresed at a very high voltage. The fluorescence of migrating fragments is captured by the LI-COR DNA analyzer and the saved gel image is then visualized in Adobe Photoshop software.

In capillary electrophoresis (CE), the DNA fragments migrate in a strong electric field (~15 KV) generated within a capillary. Here, the migration not only depends on the mass and net charge since the conducting solution also moves along, a phenomenon known as electro-osmotic effect which significantly increases the resolution of DNA fragments (Le et al., 1997).

Denaturing HPLC (dHPLC) method of mutation detection involves ion-pair reverse phase liquid chromatography of PCR products or DNA fragment up to 1.5 Kb

long. The alkylated non-porous particle material and partial heat denaturation condition under linear acetonitrile gradient create heteroduplexes among PCR products which show differential retention on column than the homoduplex species and are easily distinguished through their elution profile (Liu et al., 1998).

In HRM (high resolution melt) analysis, fluorescently labeled primers are used. The PCR products are directly used for monitoring their melting transition. Differences in melting temperatures (T_m) between the homozygotes and heterozygotes are used for variant detection (Gundry et al., 2003; Lochlainn et al., 2011).

NGS (next generation sequencing) is the latest technique of mutation detection which was discovered in mid 90s and since then has revolutionized the entire scenario of techniques depending upon sequencing. Briefly, DNA sequencing libraries are generated which are then sequenced through synthesis. The sequencing process takes place in a massive parallel manner thus generating megabases to gigabases of sequence reads and therefore NGS is also known as massive parallel sequencing (Reis-Filho, 2009).

Among the aforementioned methods, except NGS; all other methods have mostly opted for 4, 8 or 12 fold pooling strategy going beyond which results in missing the mutations. While LI-COR can screen up to 1000 bp length effectively, it involves many steps after PCR (with IR dye end-labeled primers), CEL1 or ENDO1 aided mismatch specific digestion, precipitation, gel casting, loading and analysis of gel which makes it very prone to errors and mutations can be missed. CE and HRM are faster and only include a PCR step after which the process of mutation detection is relatively easier (Chan et al., 1996; Garritano et al., 2009). CE is very sensitive and mostly can screen 130-400 bp length of DNA (Kuypers et al., 1993; Kozlowski and Krzyzosiak, 2001). The effective amplicon length for HRM is also similar (<450 bp) (Chen et al., 2014) but both these techniques are relatively hassle free and faster than LI-COR and especially useful for genes with multiple short exons interrupted by longer introns.

2.8.3.4 Next generation sequencing (NGS)

NGS systems are typically represented by SOLiD/Ion Torrent PGM from Life Sciences, Genome Analyzer/HiSeq 2000/MiSeq from Illumina, and GS FLX Titanium/GS Junior from Roche (Liu et al., 2012). It multiplexes the gene targets and both pooling depth and sensitivity are very high. It has been successfully used in plant species like tomato (Rigola et al, 2009), wheat and rice (Tsai et al., 2011) and peanut (Guo et al., 2015). The pooling depth up to 64-96 has been effectively used for mutation detection using NGS technology. Also, exact location and nature of mutation can be ascertained

unlike the previously described technique where the identification of mutant line follows Sanger sequencing of the individual mutant DNA. However, the detection of mutation through NGS requires extensive deconvolution by softwares. For variant calling, Kharabian-Masouleh et al. (2011) used CLCbio, while Marroni et al. (2011) employed VarScan (Koboldt et al., 2009). Another popular pipeline for mutation calling in complex pooled samples is Coverage Aware Mutation calling using Bayesian analysis (CAMBa) (Missirian et al., 2011). CAMBa has been successfully used by Tsai et al. (2011). The efforts for easier analysis of NGS output and following bioinformatics are still progressing (Nielsen et al., 2011; Wang et al., 2015).

2.8.4 TILLING populations in tomato

Tomato being the model crop for fleshy fruit related physiological studies, has been a target for TILLING since a long time, even before the completion of its genome sequencing (Emmanuel and Levy, 2002). Mutant populations in tomato have been developed by many groups using EMS treatment, gamma-ray irradiation, and fast-neutron irradiation. Following are the progresses made in Tomato TILLING population development, reviewed by Okabe and Ariizumi (2016).

Menda et al. (2004) produced ~6,000 EMS mutant lines and 7,000 fast-neutron mutant lines in the **M82** background. The details and seeds of this population can be found through an associated database- 'The Genes That Make Tomatoes' (http://zamir.sgn.cornell.edu/mutants/). Also, 5,508 EMS mutant lines were developed in **Red Setter**, as detailed in the LycoTILL database (http://www.agrobios.it/tilling/; Minoia et al. 2010), 8,225 EMS mutant lines were produced using tomato cultivar **TPPADASU** (Gady et al. 2009), and more than 5,000 EMS mutant lines have been generated in **Heinz 1706** (http://tilling.ucdavis.edu/index.php/Tomato_Tilling). **Micro-Tom** was used for mutagenesis by Saito et al. (2011) who developed 8,598 EMS and 6,422 gamma-ray irradiation mutant lines, which are accessible through the 'TOMATOMA' resource. Micro-Tom was also used by INRA-Bordeaux (France), where ~8,000 EMS mutant lines have been developed so far and 30,000 annotations are available for more than 150 phenotypic categories (Just et al. 2013) in the associated Micro-Tom Mutant Database (MMDB; not yet publicly available).

While TILLING utilizes the induced mutations and analyzes their effects which are created through mutagenesis and did not exist naturally, EcoTILLING helps in

identifying the natural alleles of a gene which might have evolved during adaptation and affected the gene function.

2.9 EcoTILLING:- Evaluation of natural polymorphism

Thorough assessment of natural variations in a population helps in understanding the degree of polymorphisms that a gene can exhibit and still remain functional or adapted to its environment. As a tool to catalog the natural variations, EcoTILLING was soon adapted after TILLING for understanding the population polymorphism and trait correlations (Comai et al., 2004). EcoTILLING utilizes basically the same procedures as those for TILLING except the mutagenesis step. Here the comparison is made between the various germplasms and a reference/control cultivar of choice. Since domesticated/cultivated varieties usually have a narrow genetic base, a number of polymorphisms like SNPs, microsatellite repeats and short indels are observed abundantly in naturally growing germplasms or those adapted to exotic environments but not in the reference cultivar.

EcoTILLING was first applied to Arabidopsis for various DNA methyl transferases, putative GATA type transcription factor and phytochrome interacting factor (PIF) and 55 haplotypes for 5 loci were scored. However, haplotype variations were mostly restricted to intronic regions. Subsequently EcoTIILING was carried out in Barley for powdery mildew resistance related genes (Mejlhede et al., 2006), for melon necrotic spot virus (MNSV) resulting in the discovery of a new allele of translation initiation factor elF4E. elFAE was associated with MNSV resistant phenotype (Nieto et al., 2007). Poplar EcoTILLING was carried out for genotyping purpose of the existing varieties (Gilchrist et al., 2006). EcoTILLING in rice has been extensively used for drought tolerance due to the high water demand of this crop. Leung et al. (2004) identified putatively causal SNPs in Pp2a4 (protein phosphatase gene) and DREB1 (drought responsive element binding protein1). McIntyre et al. (2006) explored the variations in sucrose phosphate synthase (SPS) gene in sugarcane for its possible role in sugar accumulation. Recently, EcoTILLING has been used for wheat kernel hardness (Wang et al., 2008), abiotic stress tolerance in barley (Xia et al., 2013), salt related genes in rice (Negrao et al., 2013) and winter hardiness related gene BvFL1 in beetroot (Frerichmann et al., 2013).

Since the beginning of EcoTILLING, efforts were made to further cut down the steps and cost involved in it because EcoTILLING does not require as big a population as in case of TILLING. Kadaru et al., 2006 and Raghavan et al., 2007 used agarose gel based EcoTILLING in rice accessions without using labeled primers and showed it to be

equally effective. Agarose gel based detection significantly reduces the cost, time and effort involved in LI-COR based approaches.

NGS technology is also increasingly being applied for EcoTILLING, population genetics and plant breeding. NGS based EcoTILLING can help in determining the genetic basis of agriculturally important traits, development of molecular markers for linkage mapping, association mapping, wide crosses, epigenetic modifications, transcript profiling, mapping of useful alleles identified through forward genetics etc. (Varshney et al., 2009; Schneeberger and Weigel, 2011; Varshney et al., 2014). The possibilities are endless with the new algorithms for NGS data analysis being developed fast (Nielsen et al., 2011).

Chapter 3

Materials and Methods

3.0 Plant material and growth conditions

3.1 Plant material

The natural accessions of tomato (*Solanum lycopersicum)* were obtained from NBPGR (National Bureau of Plant Genetic resources (http://www.nbpgr.ernet.in/), IIVR (Indian Institute of Vegetable Research, Varanasi, India (http://www.iivr.org.in/) Bejo Sheetal, Jalna, India and TGRC (Tomato Genetic Resource Center, University of California, Davis (http://tgrc.ucdavis.edu/). The wild relatives of tomato namely *S. pimpinellifolium* (LA1589), *S. cheesmaniae* (LA0483), and *S. habrochaites* (LA1777) were obtained from Tomato Genetics Resource Centre (University of California, Davis). Seeds of tomato cultivar Arka Vikas, a local South Indian variety were obtained from IIHR Bangalore and tomato cultivar M82 X *S. pennellii* (LA0716) introgression lines (ILs) were kindly provided by Dani Zamir, Hebrew University, Israel.

3.1.1 Growth conditions and tissue collection

The seeds from tomato accessions were surface sterilized with 4% (w/v) sodium hypochlorite for 10 min, which also softened the seed coat, followed by washing under running tap water till the traces of hypochlorite were removed. Thereafter the seeds were spread on moist germination paper placed in plastic germination box (9.5 cm length, 9.5 cm breadth, 5 cm height) and incubated in darkness at 25±3°C for germination. Emergence of radicle was scored as the first sign of germination. The germinated seeds were then transferred to coconut peat mixture (Sri Balaji Agroservices, Madanapalle, Andhra Pradesh, India) filled in the wells of plastic seedlings trays. The seedlings were grown under 16/8 hours light/dark conditions (100 µmol m-2s-1) in growth room at 25±3°C for 10-15 days till the establishment of seedlings. The seedlings were then transferred to the open field. Unless specified all the plants were grown in an open field at University of Hyderabad under drip irrigation. The *S. pennellii* ILs were grown in greenhouse.

For leaf tissue, fully expanded leaves from 4th node were harvested. In some cases, due to leaf miner (*Tuta absoluta*) outbreaks, infested plant parts had to be removed and the leaves were taken either from 4th, 5th or 6th node. This pattern was maintained throughout the study. Uniformity in sample collection was maintained as much as possible.

Plants started flowering after 15-20 days of transfer to the soil (most of the accessions usually flower after the appearance of 8th to 12th node). Flowers of the first truss were tagged at the time of anthesis to monitor the age of developing fruit thereafter. Fruits

at the 28-35 days after anthesis were harvested for MG stage. The transition from mature-green to red-ripe stages also varied among the accessions, requiring 8-15 days duration to attain the red-ripe stage. Harvesting of fruits was carried between 8.30-10.30 am. Immediately after harvesting, the fruits were kept on ice and taken to the laboratory. Fruits were cut into two halves for deseeding and pieces of fruit pericarp were snap frozen in liquid nitrogen, homogenized using hand held homogenizer (IKA A11, Germany) and stored at -80°C till further use. Leaflets of fully expanded leaves were collected from the fourth node of approximately 45 days old plants, frozen in liquid nitrogen, hand homogenized using mortar and pestle and stored at -80°C till further use.

3.2 TILLING (Targeting Induced Local Lesions In Genomes) 120 mM remutagenized population development

A large 120 mM ethyl methane sulphonate (EMS) mutagenized population of a high yielding variety of tomato (*Solanum lycopersicum* cv. Arka Vikas) was used as the starting material for re-mutagenesis. In view of the low mutation frequency detected in M_2 DNA of aforementioned population, (Chaitanya Charakana, Ph.D. thesis, 2015) and to increase the frequency of mutations, M_2 (deemed as M_0) seeds were re-mutagenized as per following protocol. Mutagenized (120 mM EMS) M_2 seeds (deemed as M_0) were soaked in 100 mL of double distilled water with ethyl methane-sulfonate (EMS) at a concentration of 120 mM in conical flasks. Flasks were gently shaken for 24 h at room temperature. Finally, seeds were placed in muslin cloth and tied with a thread. These muslin cloth bags were placed in a 5 liter beaker filled with water and washed extensively under running tap water for 8 hours. After the EMS treatment, seeds were germinated on germination paper and seedlings were transferred to 50-well egg trays containing decomposed coconut peat. These M_2-M_1 seedlings were transplanted at three weeks old stage at a spacing of 1.3 m x 40 cm in open field with drip irrigation facility.

The efficacy of mutagenesis was evaluated by noting the events of chlorotic/chimeric leaves (M_2-M_1) in seedling or later stage in the population (Figure 3.1). These M_2-M_1 plants were allowed to self-pollinate and produce M_3-M_2 seeds. Approximately 8-10 M_3-M_2 seeds were sown in decomposed coconut peat in greenhouse to grow the M_3M_2 generation. Nearly 15 days after sowing, juvenile leaf samples (approximately 100 mg tissue weight) were collected from these seedlings, placed in 1.5 mL Eppendorf tubes and stored at -80 °C. The leaf tissue was stored till DNA isolation for mutation screening by TILLING. After three weeks from sowing, four individual plants (A, B, C and D) from each M_3M_2 family were transplanted in open field with drip

Figure 3.1. Albino/yellow seedlings observed in M_2-M_1 generation after EMS mutagenesis treatment.
Note:- Red arrows indicate albino /yellow cotyledon and leaf, blue arrow indicates normal green cotyledon and leaf. The albino seedlings do not survive and often die within a week of cotyledonary stage.

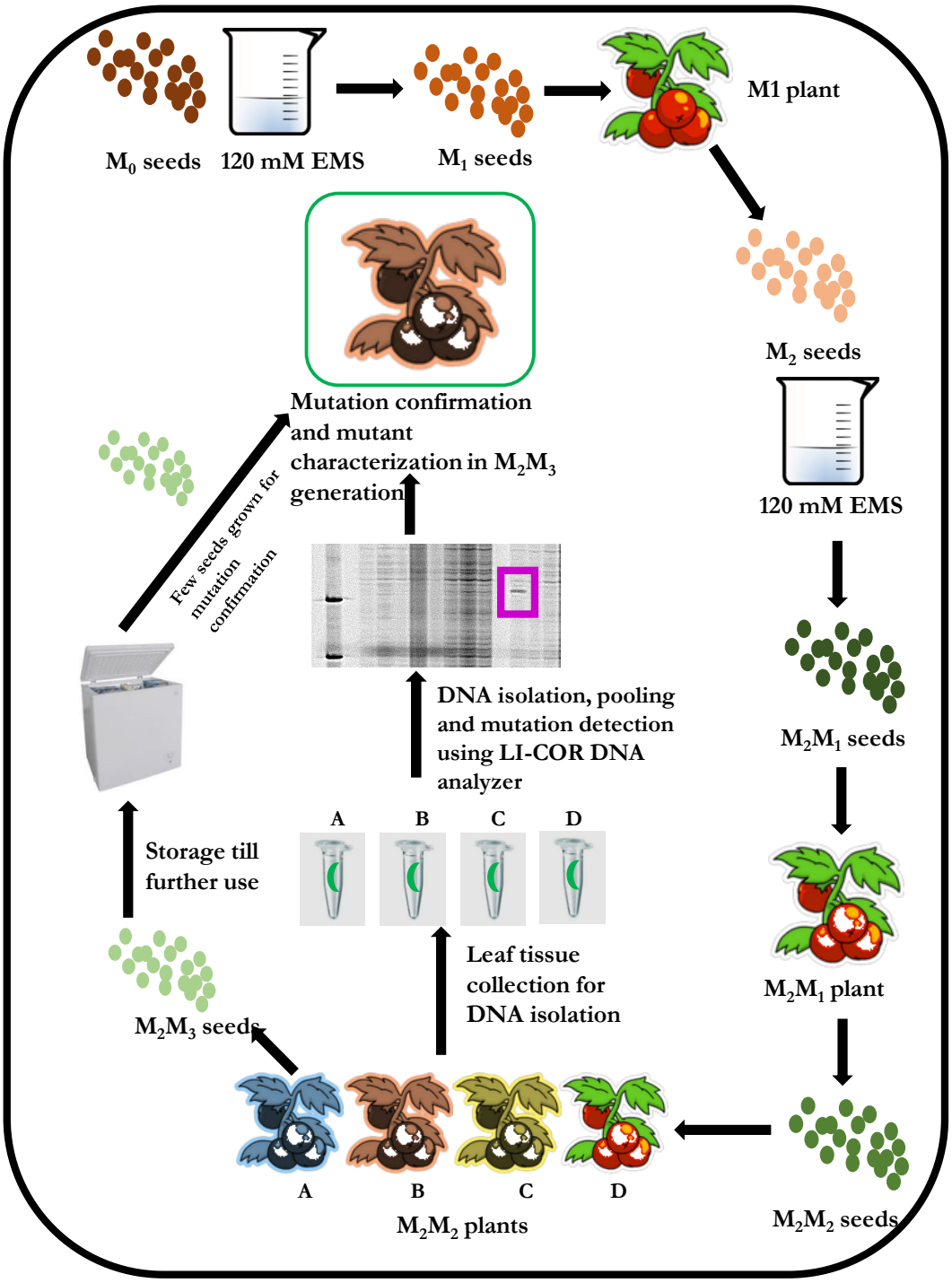


Figure 3.2. Pictorial representation of EMS double mutagenesis of tomato (cv. Arka Vikas) population. Clockwise- Mutagenesis, population development, seed collection, development of M_2M_2 family structure tissue collection for DNA isolation(A,B,C,D) and seed harvest.

irrigation facility and allowed to self-pollinate. After fruit set red ripe (RR), fruits were collected individually from each plant (A, B, C, D), squeezed and seeds were harvested. Thus, four different series of seeds were obtained. Each fertile M_3M_2 plant was treated as an independent line and fruits were collected from individual plants to obtain M_4M_3 seeds. The lines that did not yield seeds were excluded from TILLING and respective stored leaf tissue of those lines was discarded (Figure 3.2).

3.2.1 DNA isolation

DNA isolation was carried out according to Sreelakshmi et al. (2010) with some modifications. Briefly, ~100 mg leaf tissue was homogenized with 750 µL of preheated (at 65°C) DNA 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) in 1.5 mL Eppendorf tubes. 2 µL proteinase K (20 mg/mL stock) was added to each sample and incubated at 65°C for 30 minutes, temperature was lowered to RT and 400 µL ammonium acetate (6 M) was added to the samples and after gentle mixing, centrifuged (Sigma 1-15PK, Germany) at 18894 *g* for 30 min at RT. The supernatant (~990 µL) was transferred to a fresh tube and equal volume of ice cold isopropanol was added. After gentle mixing, it was incubated at -20°C for 1 hour to precipitate the DNA, followed by pelleting by centrifugation at 13000 rpm for 30 min at 4°C.

The pellet was washed with 70% (v/v) ethanol to remove traces of salts from the pellet. After drying, the pellet was dissolved in 175 µL Milli-Q water and 4 µL of RNase (10 mg/mL) was added to each sample and incubated at 37°C for 30 min. Thereafter, 125 µL of ammonium acetate (6 M) was added, tubes were gently inverted 4-5 times followed by centrifugation at 18894 *g* for 30 min at room temperature (RT). Approximately 300 µL supernatant was collected in a fresh tube and mixed with 2.5 volumes of absolute ethanol. Tubes were incubated at 4°C for 20 min and then centrifuged at 18894 *g* for 30 min at room temperature. Pellet was washed with 70% ethanol (10 min, 18894 *g*) and air dried. Finally the pellet was dissolved in 100 µL of 10 mM Tris, pH 8.0 and incubated at 4°C for overnight to dissolve the DNA pellet. The quality and quantity of DNA was checked through agarose gel electrophoresis and Picogreen dye based quantification (Singer et al., 1997) respectively. Genomic DNA was stored in -80°C till further use.

3.2.2 Bi-dimensional (2-D) pooling of DNA for mutation detection using LI-COR platform

2-D pooling of DNA samples was done according to Tsai et al. (2011) with slight modification. After quantification, DNA was equalized to a common concentration with 10 mM TRIS and was arrayed into 8*8 master plate format before pooling. 8 fold pooling of DNA was performed; pooled DNA was distributed in 96 well PCR plates at a final concentration of 10 ng (per well) using Apricot 96 well pipettor; plates were dried, sealed and stored at -20°C or used directly for TILLING PCR. Two dimensional pooling yielded a 'row' plate and its complimentary 'column' plate (96 well format) with 768 DNA samples pooled in each plate. In two dimensional pooling each DNA sample is present in two different plates in a format that after mutation detection in a particular well of row and column plate, the plant line bearing mutation can be identified to a single plant level (Figure 3.3). In this manner DNA was pooled from 2,304 samples and three sets of plates (R1, R2, and R3) and complimentary (C1, C2, and C3) plates were made for TILLING. Our pooling strategy is similar to Sreelakshmi et al. (2010) whereas, Tsai et al. (2001) used 12*8 format master plates to prepare row and column pools. While our pooling strategy (8*8 format master plates) can directly indicate the mutant line number after row and column plate confirmation, bi-dimensional pooling method of Tsai et al. (2011) narrows down to 8 individual line numbers which need to be sequenced further to identify the exact mutant line. This considerably hinders the high-throughput aspect of entire pooling strategy.

3.2.3 Tri-dimensional (3-D) pooling of DNA for mutation detection using NGS platform

After two dimensional pooling where the pooling depth was only 8 fold, suitable for LI-COR based mutation detection system, we further adopted the three dimensional pooling (Tsai et al., 2011) using DNA of the same remutagenized population. This pooling was used for TILLING using Next Generation Sequencing (NGS) technology which is more sensitive than the LI-COR based mutation detection systems. Briefly, equalized DNA was arrayed into 12 master plates in 8*8 format. All 64 DNA samples of one plate were pooled together, thus comprising one 'D' pool and similarly 12 'D' pools were made. When all 12 plates were hypothetically stacked upon and horizontally cut into 6, 6 plates, and DNA of each column of all 6 plates was pooled together, that formed 'column pools' which were 8+8=16. Again, hypothetical stack of plates were cut vertically into 4 columns in each stack and DNA was taken row wise for the stack having a height of 12 cells which gave rise to 16 'row pools'. While the extent of pooling in D pools was 64 fold, it was 48 fold in both, column and row pools (Figure 3.4).

3.3 PCR and mutation detection:-LI-COR

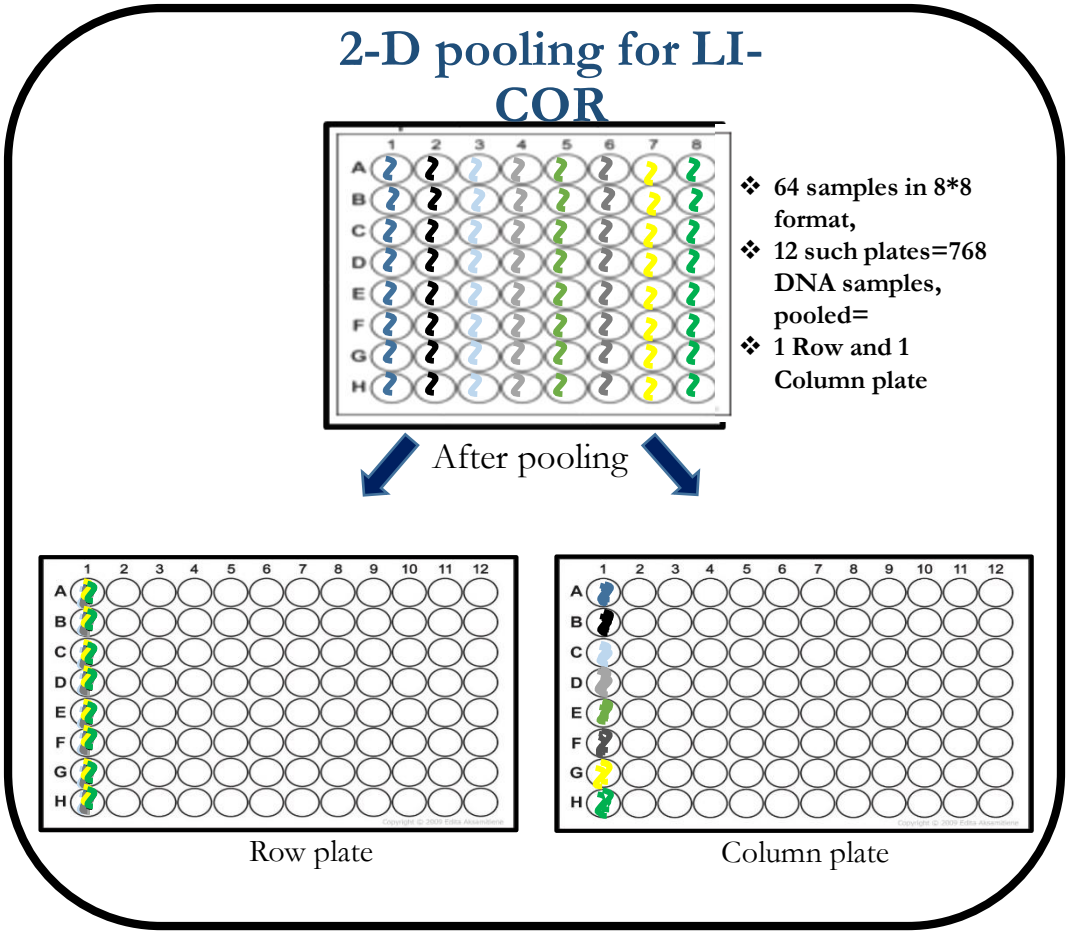


Figure 3.3. Graphic representation bi-dimensional pooling strategy. Samples from A1 to A8 are pooled in A1, B1 to B8 in B1 of row plate and so on. Samples from 1A to 1H are pooled in in A1, 2A to 2H in B1 of column plate and so on. 768 DNA samples arrayed into 12 plates make 1 row 1 column plate. 3 sets of plates were prepared in a similar manner containing the DNA of 2304 mutagenized tomato lines for Li-COR based TILLING.

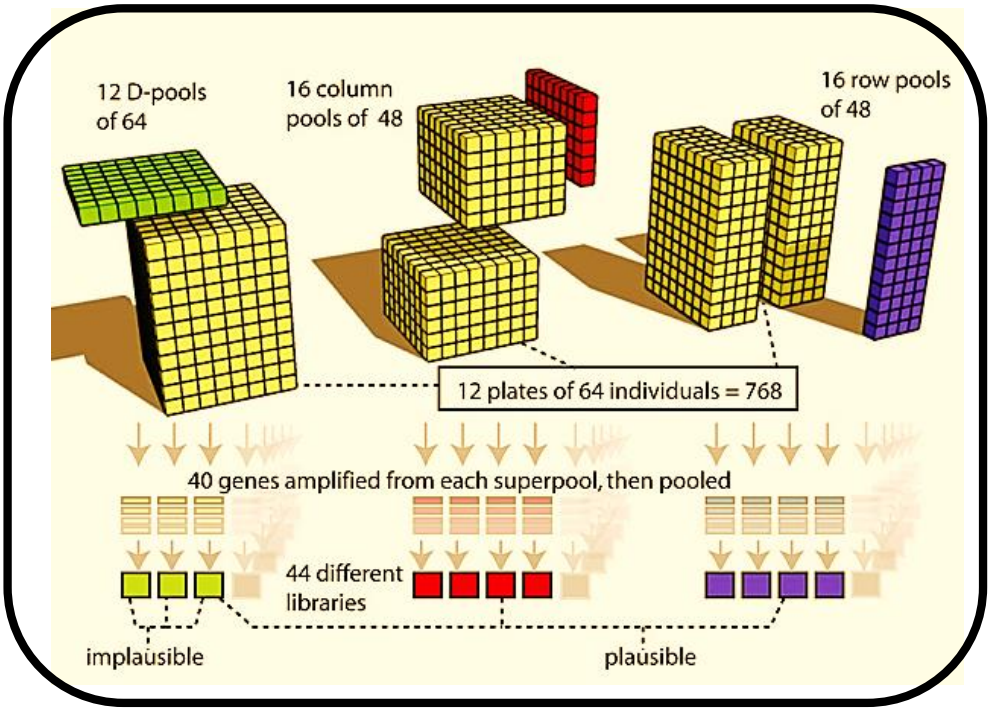


Figure 3.4. Tridimensional pooling and sampling strategies for mutation discovery. Individual genomic DNAs represented by the small cubes were arrayed on plates of 8X8 wells. Pools were formed by combining individual DNA along the three-dimensional axis, here called arbitrarily D-pools, column pools, and row pools. Each of these three pooling arrangements is illustrated by the color arrays (green, red, or blue). The pooling resulted in 44 mixed templates (Tsai et al., 2011).

Two genes namely *folylpolyglutamate synthase* (*FPGS*) and *γ-glutamyl hydrolase* (*GGH*) genes were selected for TILLING in present study. In tomato *FPGS* is reported to have two isoforms, one plastidial (*FPGSp*) and one mitochondrial (*FPGSm*) and they are present on 5th and 4th chromosome respectively (Waller et al., 2010). *GGH* exists in 3 isoforms, namely *GGH1, GGH2* and *GGH3* which are supposed to be vacuole localized (Akhtar et al., 2008 and Orsomando et al., 2005). Gene sequences of *FPGS* and *GGH* isoforms (*FPGSm*), (SGN ID Solyc04g016550.2), plastidial (*FPGSp*), (SGN ID Solyc05g052920.2) and *GGH1*, (SGN ID Solyc07g062270.2), *GGH2*, (SGN ID Solyc07g062270.2), and *GGH3* (SGN ID Solyc07g062280.2) were retrieved from Sol Genomics Network (SGN, http://solgenomics.net/).

The CODDLE program (Codons Optimized to Discover Deleterious Lesions) combined with the PRIMER3 tool was used to define the gene region which would yield maximum missense changes, if mutated and primers were designed accordingly (Figure 3.5). CODDLE software compares the protein block models in the entire database and according to the position specific scoring matrix (PSSM) scores, predicts the sites in the gene of interest which if mutated would result in most missense changes and truncations. A complete primer list used in TILLING and EcoTILLING with SOL IDs, amplicon size and genomic region amplified are shown in Table 3.1.

For LI-COR based TILLING, a nested PCR strategy was used (Figure 3.6) where first set of primers amplify a CODDLE predicted region of gene from pooled DNA. PCR cycling conditions for first step were- 94°C- 4 min, 94°C-20 sec, 55°C-45 sec with extension at 72°C–2 min, 34 more cycles to step 2, 72°C-10 min and incubation at 12°C.

Subsequently PCR products of first step were used as a template for second step amplification where second set of unlabeled primers bearing universal M13 sequence at 5' end covering a shorter region within the 1st step amplicon and amplified a product of roughly 1000 bp size. In addition, mixture also contained 700 nm and 800 nm IR dye labeled M13 forward and reverse primers. The combination of 1 nM concentration of unlabeled: labeled primers was as follows- F:FL, R:RL=0.7:1, 0.5:1.2 (F=forward unlabeled, FL= forward labeled, R=reverse unlabeled, and RL=reverse labeled). Touchdown PCR cycling conditions for second step were- 94°C- 4 min, 4 cycles of 94°C-20 sec, 60°C-45 sec with a decrement of -2.0°C per cycle, 72°C-1 min 30 sec followed by 30 cycles of 94°C-20sec, 52°C-45 sec, 72°C- 1 min 30 sec, 72°C-10 min. Additional heteroduplexing steps were added to the program as following- 98°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.

The presence of heteroduplex was detected by using a mismatch specific endonuclease CEL1 that cleaves the heteroduplex DNA at the site of mismatch resulting in fragmented DNA. Briefly, 25 µL of CEL1 digestion master mix (1X CEL1 digestion buffer (10 mM HEPES buffer pH 7.0, 10 mM KCl, 10 mM MgCl2, 0.002% (v/v) Triton X100 and 10 µg/ml BSA) and CEL1 enzyme at 1:300 dilution (1 µL/300 µL CEL1/ digestion buffer) and 20 µL heteroduplexed PCR product were mixed. The mixture was incubated at 45°C for 15 min and then the reaction was stopped by adding 10 µL stop solution (2.5 M NaCl, 75 mM EDTA, pH 8.0 and 0.5 mg/ml blue dextran).

The DNA was precipitated by the addition of 125 µL of cold absolute ethanol and incubated at -80° C for 15- 30 min. followed by centrifugation at 4500 rpm for 30 min in a Sorvall RC6 centrifuge, SH3000 plate rotor. The pellet was washed with 70% (v/v) ethanol, dried in a dry bath at 80°C and then suspended in 8 µL of formamide loading buffer (37% (v/v) de-ionized formamide, 1 mM EDTA and 0.02% (w/v) bromophenol blue). The PCR products were denatured by heating the plates/tubes to 94°C for 2 min and then placed on ice. These fragmented products were resolved on high resolution denaturing PAGE (polyacrylamide gel electrophoresis). Since primers used for second step PCR was M13 tailed (and tagged by IR dye), the fragments can be visualized and located easily by high resolution PAGE setup (LI-COR 4300 DNA analyzer, LI-COR Biosciences, Nebraska, USA) as reaction mixture also contained IR dye labeled primer bearing M13 sequence.

After detection of mutation in a line, genomic DNA from that line was re-amplified and subjected to agarose gel based mismatch detection assay (Sharma et al., 2011) to confirm the mutation. This assay utilizes the property of CEl1 enzyme for cleaving both the DNA strands at the site of mismatch. When a polymorphic site in the amplified product pairs with the wild type copy of the PCR product during heteroduplexing reaction, CEL1 identifies and cleaves the DNA at that place. The CEL1 digested products can be seen on agarose or polyacrylamide gel. Moreover, to ascertain whether the mutation is heterozygous or homozygous, the genomic DNA was digested either alone or after mixing with the WT-AV DNA. The presence of fragments in unmixed genomic DNA indicated heterozygous nature of mutation while fragment detection only after mixing with WT-AV DNA indicated mutation to be in homozygous state. After agarose gel based confirmation, almost 10 M_4M_3 seeds from respective line were grown, DNA was isolated, and mutation was reconfirmed at the plant level.

3.3.1 PCR and mutation detection:-Next Generation Sequencing (NGS)

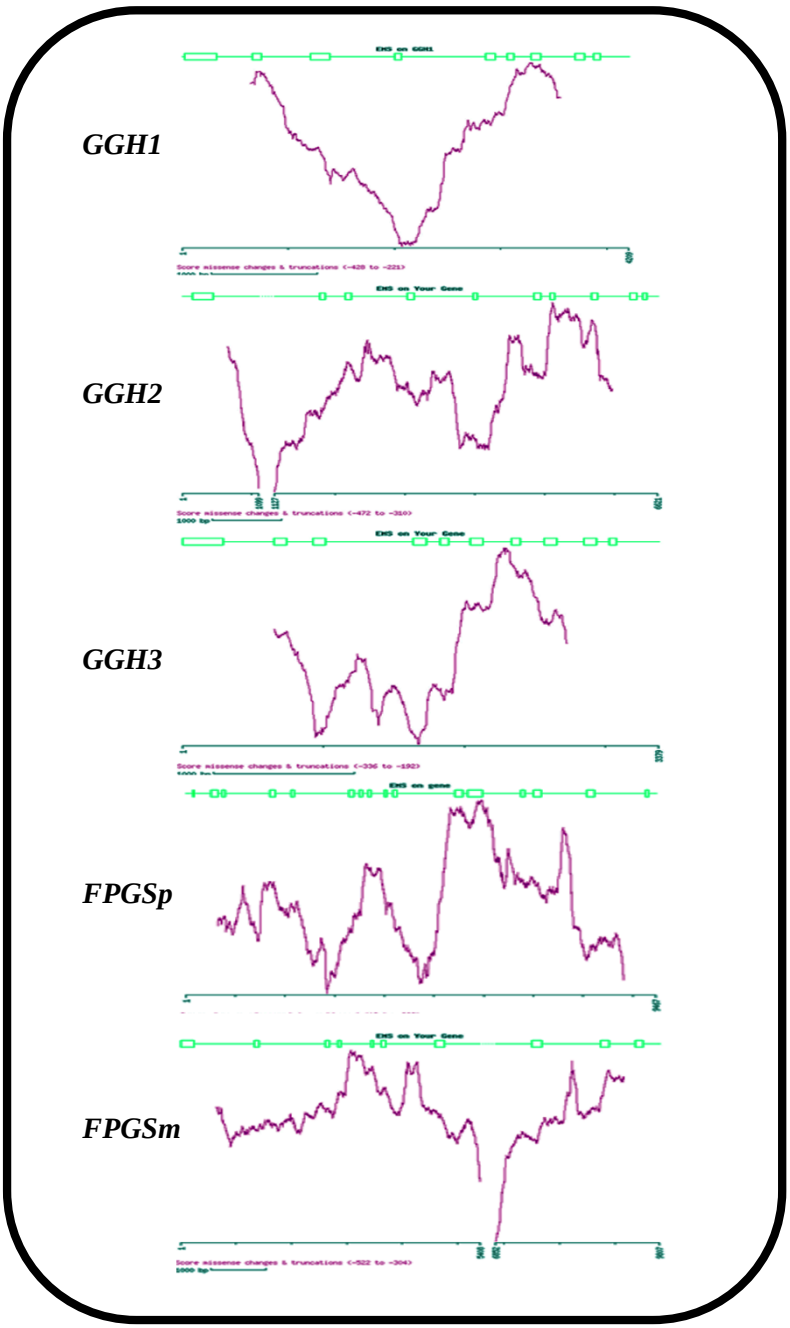


Figure 3.5. Regions of *GGH1, GGH2, GGH3, FPGSp* and *FPGSm* genes where occurrence of polymorphism is most likely to cause missense and truncation changes as per the prediction of CODDLE software. (http://blocks.fhcrc.org/proweb/coddle/)
Note:-Green boxes show the exons and intervening line shows introns. The purple peaks represent regions of high probability for a deleterious change.

Table 3.1. List of genes, their SOL ID, primers sequence, chromosomal position and amplicon size used for both EcoTILLING and TILLING in folate pathway genes.
The sequences of M13 universal primer were 5′TGTAAAACGACGGCCAGT3′- (IRD700- labelled) and 5′AGGAAACAGCTATGACCAT3′ (IRD800- labelled), flanking 5 end of second step forward and reverse primers respectively.

Gene	SOL ID	PCR step	Primer direction	Primer sequence 5′-3′	Chromosomal position screened for SNPs	Amplicon size
FPGSm	**Solyc04g016550.2**	1st step	Forward	AGGGGTTTTTTGGATTAGCTTTTA	-	1238 bp
			Reverse	TTGTGTGTGCTTATCCATGAAA		
		2nd step	Forward	TAGTCTTTCCCAAACACCTCCA	**SL2.50ch04 7377603-7378603**	1000 bp
			Reverse	TAGCATATCTAAGTTGAAGCAGC		
FPGSp	**Solyc05g052920.2**	1st step	Forward	GGTCACTGCAAAGCTGATGA	-	1627 bp
			Reverse	AACATCAACGCTCCCAAGTC		
		2nd step	Forward	CAGGTCTTGCATAGCATGTAACTCT	**SL2.50ch05 63100090-63098864**	1226 bp
			Reverse	AAGTCGCTCAAAACAAAAGGTTC		
GGH1	**Solyc07g062270.2**	1st step	Forward	TGTGTAGCACTTCTGTTTTCTAGC	-	1936 bp
			Reverse	CGGTTTACATCACACACTAACTTC		
		2nd step	Forward	GCATTGTTCATAAACTGTCAACTAGAG	**SL2.50ch07 65053192-65054515**	1323 bp
			Reverse	CACAAGACTAGAAGCATAACAGTG		
GGH2	**Solyc10g007410.2**	1st step	Forward	TCTGTTTCATTGCCTGCCGAA	-	1263 bp
			Reverse	AAATGTCGTCAAAGGGCTTG		
		2nd step	Forward	TTCCAGCGATTTGTGATTA	**SL2.50ch10 1774277-1775113**	836 bp
			Reverse	CATGGGAGGAACAGAATACA		
GGH3	**Solyc07g062280.2**	1st step	Forward	GAGAAATTCAGTGTATCAAATCAGGCA	-	1300bp
			Reverse	TGCAGAATTACCTACCCGATCT		
		2nd step	Forward	TTCCCACCTACATTAGTAAAGAAGT	**SL2.50ch07 65057527-65058497**	970 bp
			Reverse	ATCGTCGAATGAATGAAAACAC		

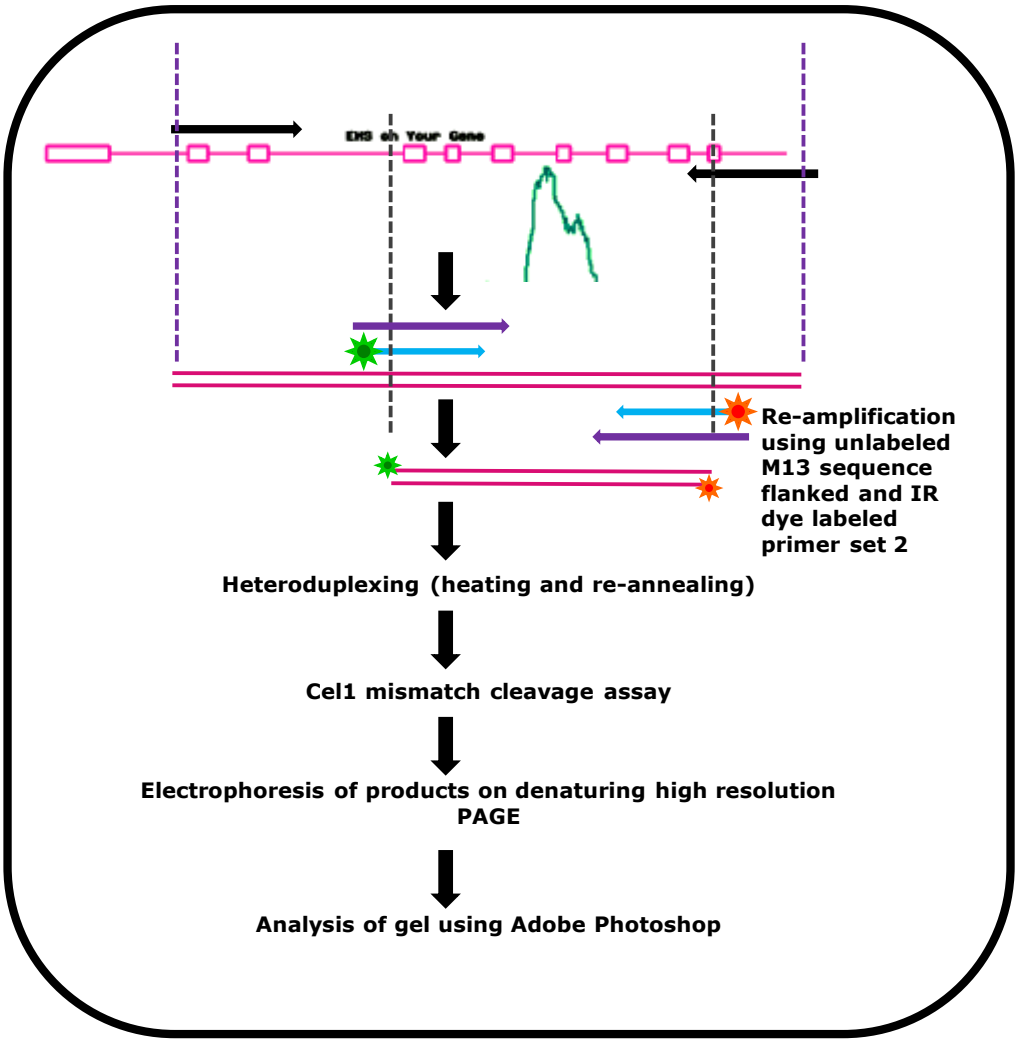


Figure 3.6. Schematic representation of nested PCR strategy used for EcoTILLING and TILLING. The CODDLE predicted region (green jagged peaks) of gene of interest are amplified by primer set 1 flanking that region (purple dash lines) by a margin of 200-300 bp. Resulting PCR product is re-amplified using IR700 and IR800 labeled smaller (grey dash lines) primer set 2. Second step PCR uses two primer sets. Purple arrows represent primer set 2 with M13 sequence overhangs and blue arrows with green and orange star represent M13 primer 5′ end labeled with IR700 and IR800 dye. Thus the PCR product gets IR dye labeled on both the strands at respective 5′ ends. After heteroduplexing reaction followed by CEL1 mismatch assay, cleaved products have either of the ends labeled with the fluorescent dye and this fluorescence is recorded in LICOR DNA analyzer

For mutation detection using NGS, all 44 superpools (16 row pools, 16 column pools and 12 D-pools) were amplified with the gene specific primers encompassing the region of interest, chosen as per CODDLE (Codons Optimized to Discover Deleterious LEsions) prediction. The amplicon size was 600±10 bp for each gene. The primer sequences and amplicon sizes used for NGS are shown in Table 3.2. Amplicons were checked on agarose gel to ensure proper amplification, vacuum dried and sent to Dr. Bradley J. Till, Plant Breeding and Genetics Laboratory, IAEA (International Atomic Energy Agency), Austria for sequencing. The pools were indexed using unique Illumina barcodes for library construction. Libraries were prepared using Illumina reagents, according to the manufacturer's specifications (Illumina, http://www.illumina.com) and sequenced on a MiSeq platform followed by SNP calling. The mutation effect was analyzed by SIFT (Sorting Intolerant from Tolerant, http://sift.jcvi.org/) with default parameters. After putative mutations were predicted, individual lines were identified through deconvolution of the pooling strategy. DNA from those lines was amplified and mutations were confirmed on agarose gel. Mutant lines with SIFT score <0.05, indicating the likely deleterious effects were further grown for characterization.

3.4 EcoTILLING

A population of natural accessions of tomato obtained (annexure 1) comprising 391 accessions was grown. 10 seeds from each accession were germinated. Leaf tissue from each accession was collected from almost 20 days old seedlings for DNA isolation followed by shifting of seedlings to open field after 21 days from germination. DNA isolation was done as per Sreelakshmi et al. (2010) (section 3.2.1). DNA was heated at 65°C for 10 minutes in a water bath and its quality and quantity was checked on agarose gel and using nanodrop. Equalized DNA was aliquoted into 96 well plates (5 ng per well), vacuum dried and stored in -20°C till further use. This DNA was amplified using respective gene specific primers and screened for Single Nucleotide Polymorphism (SNP) on LI-COR DNA analyzer. All the steps involved in EcoTILLING were essentially the same as described in TILLING sections (3.2 and 3.3) except following modifications. The DNA was not pooled in case of EcoTILLING and instead, represented individual accession's DNA. Also, to make heteroduplexes for CEL1 mismatch assay, wild type (Arka Vikas) DNA was mixed with the natural accession's DNA in a ratio of 1:1 during the first step PCR. SNPs appeared on LI-COR gel in the form of bands and their respective cuts were confirmed in the complimentary channel. The accessions showing identical fragment size on LI-COR gels

were grouped as a single haplotype. Further, DNA from each haplotype was sequenced to determine the exact position and type of polymorphism.

3.4.1 *In silico* survey of polymorphism in 444 tomato accessions

To assess the naturally occurring variations in folate metabolism genes selected in present study, SNP data of 84 accessions of *S. lycopersicum* and wild relatives of tomato (Aflitos et al., 2014) and 360 tomato accessions including many wild species of tomato were sequenced (Lin et al., 2014) were analyzed. All vcf files for this data are accessible to the public with the SNPs scored for each accession (with reference to Heinz 1706) in downloadable form at ftp://ftp.solgenomics.net/genomes/tomato_150/ for 84 accessions and at ftp://ftp.solgenomics.net/genomes/tomato_360/ for 360 accessions. From the above data set, we exclusively examined exonic SNPs and their potential consequences on the protein function in 6 genes of folate metabolism viz. *dihydrofolate synthase (DHFS), Folylpolyglutamate synthase-mitochondrial (FPGSm), Folylpolyglutamate synthase-plastidial (FPGSp), γ-glutamyl hydrolase 1, 2 and 3 (GGH1, GGH2 and GGH3).* The altered protein sequences were grouped on the basis of exact similarity. For each gene, different number of groups were formed. The phylogenetic analysis was carried out with the help of ensemble large workflow in the ETE toolkit (http://etetoolkit.org/). The amino acid lengths varied from 337 residues (GGH3) to 554 residues (DHFS) and the visualization of alignment was not feasible. Therefore, only those places/residues are shown in the alignment which are different in any of the groups created during the initial alignment of altered protein sequences. Name of only one randomly chosen member of each group was shown in the alignment for clarity and ease of visualization (remaining members and their names are presented in annexures 2, 2.1, 3 and 3.1. Finally, each change was evaluated using Variant Effect predictor (VEP) tool from Ensembl (http://www.ensembl.org/Tools/VEP) and the frequencies of all types of changes were compared.

3.5 Folate extraction and estimation

3.5.1 Chemicals and folate standards

5-Methyltetrahydrofolate (5-CH_3-THF), tetrahydrofolate, (THF), 5,10 methenyltetrahydrofolate (5,10-CH^+THF), 5-formyltetrahydrofolate (5-CHO-THF) and 5,10-methylenetetrahydrofolate (5,10-CH_2THF) were purchased from Schirck's Laboratory (Jona, Switzerland). Folic acid (FA), methotrexate (MTX), ascorbic acid, β-mercaptoethanol, LC-MS grade acetonitrile and α-amylase (from *Bacillus* sp., A6814) were obtained from Sigma Aldrich Co. (St. Louis, USA). Milli-Q water (18.2 Ω at 25°C) was

Table 3.2. List of genes, their SOL ID, primers sequence, chromosomal position and amplicon size used for TILLING in folate pathway genes using next generation sequencing.

Gene	SOL ID	PCR step	Primer direction	Primer sequence 5′-3′	Chromosomal position screened for SNPs	Amplicon size
FPGSp	**Solyc05g052920.2**	1st set	Forward	GTGATTTTCGGTATTTCTTGTGCAG	**SL2.50ch05 63097069-63097652**	583 bp
			Reverse	AGCAGAAAAAGCTTTGGTGTCATTT		
		2nd set	Forward	GTTGATGTGAACATGTATTGTTGTGC	**SL2.50ch05 63099186-63099776**	590 bp
			Reverse	AAGCTTACCTGGGTTGAACTTTTCT		
GGH1	**Solyc07g062270.2**	1st set	Forward	TGCAAATTATGCTGGAAGTATG	**SL2.50ch07 65051240-65051822**	582 bp
			Reverse	CTTCATTTTCTAGACTGTTTTCGAC		
		2nd set	Forward	TTGAGTAGCTTTTTCAGGGTATTGAC	**SL2.50ch07 65053738-65054348**	610 bp
			Reverse	CGGAATCACTCAATCTGAAATTACC		

obtained from a purified water system (Millipore, Bradford, USA). LC-MS grade formic acid (HCOOH) was obtained from Fisher Scientific (Loughborough, UK). Potassium dihydrogen phosphate, dipotassium hydrogen phosphate, folic acid casei medium (M-543-100G), Protease (from *Streptomyces griseus*, RM6186) and activated charcoal were obtained from HiMedia (Mumbai, India). For microbiological assay, *Lactobacillus rhamnosus* (ATCC 7469) was obtained from MTCC (Microbial Type Culture Collection) Chandigarh (mtcc.imtech.res.in/) (ATCC 7469=MTCC 1408). Protease (P5147) and sodium ascorbate were obtained from Sigma Aldrich Co. (St. Louis, USA). Rat plasma was obtained from National Institute of Nutrition (NIN), Hyderabad, India.

3.5.2 Spectrophotometric purity correction standard stock preparation

Stock solutions of folate standards (1 mg/mL) were prepared in 50 mM potassium phosphate solution, pH 4.5 containing 1% (w/v) of ascorbic acid and 0.5% (v/v) of β-mercaptoethanol except folic acid (FA) and methotrexate (MTX), which were dissolved in neutral or basic pH potassium phosphate buffer. The standard stock solutions were freshly diluted in the extraction solution to prepare working solutions. The remaining stock solutions were flushed with nitrogen gas, and small aliquots were stored at -80°C. MTX was used as an internal standard.

Before preparing the stock solutions, purity of the standards was calculated according to their molar absorption coefficients. For spectrophotometric measurements, standards were dissolved in 0.01 M phosphate buffer (100 ng/µL), 0.01 N HCl was used to dissolve 5,10-Methenyl THF and 1 N NaOH was used to dissolve Folic acid. Molar absorption coefficients were obtained from Zhang, et al. (2003), except for 5,10-Methenyl THF which was taken from Moldt et al. (2009). The absorbance readings were taken immediately using UV-Visible spectrophotometer (UV-1800 Shimadzu corporation, Kyoto, Japan). Afterwards, the concentration was calculated using the Beer-Lambert law (**$A=\varepsilon cl$**, where A=absorbance, ε=molar extinction coefficient [L mol^{-1} cm^{-1}], c=concentration [mol L^{-1}] and l=path length [cm]). The values obtained spectrophotometrically (estimated) and the known amount of standard (actual) were compared and the percentage purity was calculated for each standard. Also, standards (which were conjugated with sodium, calcium or chloride for better solubility) were corrected for their actual contribution to the powder weight.

The respective purities of 5-Methyl THF, 5-Formyl THF, THF, 5,10-Methenyl THF, 5,10-Methylene THF and folic acid were 90, 65, 90, 80, 77, 78% respectively (Table 3.3). All standard curves were plotted after correcting the purity factor of each standard.

3.5.3 Enzyme preparation for folate extraction

Protease (2 mg/mL) and α-amylase (20 mg/mL) were dissolved in Milli-Q water, and aliquots were stored at -20°C. To remove endogenous folate from rat plasma and α-amylase, 100 mL of rat plasma and α-amylase were mixed with 5 g of activated charcoal (http://www.koko.gov.my/CocoaBioTech/Protein%20Purification6.html). This mixture was incubated on ice for 1 hour with intermittent stirring followed by centrifugation at 5,000 *g* (Sorvall Lynx 6000, Thermo Scientific, USA) for 10 min at room temperature. The supernatant was filtered through a 0.22 μm filter, divided into 1 mL aliquots, and stored at -20°C. Protease was used without pre-treatment and was stored in -20°C.

3.6 Folate extraction and estimation:- Microbiological Assay (MA)

Folate was estimated in a population of 113 accessions of tomato including Arka Vikas that was considered as wild type for comparison. The stock culture of *L. casei* was made in Lactobacillus broth and maintained on agar medium (M369-100G, **M366-100G**, HiMedia). Many parameters like aeration, incubation time, inoculum dose etc. were optimized to achieve steady bacterial growth. A modified method of Wilson and Horne (1982) was used to prepare cryoprotected cells for using as inoculum. For preparing cryoprotected cells, 10 mL of Lactobacilli broth was inoculated from an agar culture and incubated for 24 hours at 37°C (inoculum culture). Meanwhile, Folic Acid casei Medium (**M543-100G, HiMedia)** was prepared as follows- 9.4 g of powder medium and 50 mg of sodium ascorbate were dissolved in 200 mL of distilled water, folic acid (0.3 μg/L) was added to it, and the medium was autoclaved at 121°C for 15 minutes. The medium was cooled and 0.5 mL of inoculum culture was added to it and incubated for 17-18 at 37°C, then cooled in an ice bath. This chilled culture was mixed with an equal volume of cold, sterile glycerol (80% v/v) and stored in 2 mL aliquots in microcentrifuge tubes in -20°C.

Standard curve for bacterial growth was made against increasing folic acid concentrations. All parameters related to the bacterial growth were standardized first in 1.5 mL Eppendorf tubes and later, for 200 μL volume of microtiter plate wells.

3.6.1 Extraction method

Folate was extracted from tomato fruit tissue using trienzyme extraction described by Goyer and Navarre (2007) with some modifications. Extraction was carried out in 2 ml wells of a 96 well plate. Briefly, 100 mg fresh or stored homogenized fruit tissue was suspended in 1 mL extraction buffer (0.1 M potassium phosphate pH 7.0, 1% (w/v) ascorbic acid and 0.1% (v/v) β-mercaptoethanol, flushed with liquid nitrogen). Plates were

Table 3.3. Percentage purity calculation of Folate Standards.

Standard	pH	ε# (L mol^{-1} cm^{-1})	λ (nm)	Purity % (Mean±S.E.)
5-Methyl THF	7.0	31700	290	89.8±2.9
5-Formyl THF	7.0	37200	285	64.8±3.0
THF	7.0	29100	297	89.6±2.4
5,10-Methenyl THF*	7.0	24900	354	79.3±3.1
5,10-Methylene THF	7.0	32000	294	77.2±4.3
Folic acid**	7.0	27600	282	77.5±1.8

Note:-Standards were dissolved in 0.01 M phosphate buffer (100 ng/μL).
*0.01 N HCl was used to dissolve 5,10-Methenyl THF.
**1 N NaOH was used to dissolve Folic acid
#Molar absorption coefficients were obtained from Zhang et al. (2003), except for 5,10-Methenyl THF which was taken from Moldt et al. (2009).

transferred to boiling water bath for 10 min and immediately cooled on ice. After addition of 10 μL protease (10 mg/mL), sample was incubated at 37°C for 2 h followed by transferring the plates to boiling water bath for 10 min and immediately cooling on ice. Thereafter, 25 μL α-amylase (20 mg/mL) and 25 μL rat plasma conjugase was added to the sample and incubated at 37°C for 3 h after which samples were transferred to boiling water bath for 10 min, immediately cooled on ice and centrifuged for 10 min at 3000 *g*. The clear supernatant was transferred to fresh plates. All extraction steps after first boiling step were carried out in a laminar airflow bench to avoid the need of filtration of extracts. No significant difference in folate content was observed between samples processed with or without filtration. Since folates are light sensitive the extracts were protected from light to prevent the photooxidation of folates during extraction and storage.

3.6.2 Inoculation and incubation of extracts with bacterial culture

To the sample wells of a microtiter plate containing 100 μL assay medium, 50 μL buffer, 40 μL cryoprotected cells (25 times diluted in 0.9% (w/v) NaCl), 10 μL of plant extract (32 times diluted) was added. The wells with buffer blank contained 100 μL assay medium and 100 μL buffer (50 mM potassium phosphate buffer, 0.15% sodium ascorbate (w/v), pH 6.1), while wells with inoculum blank contained 100 μL assay medium, 60 μL buffer and 40 μL cryoprotected cells (25 times diluted in 0.9% (w/v) NaCl). Plates were incubated static at 37°C for 18 h and thereafter absorbance was recorded at 540 nm in a microplate reader.

3.6.3 Calculation of total folate

Good correlation was obtained between bacterial growth response and increasing doses of folic acid with R^2 value being 0.9834. The standard curve was subjected to regression analysis and a linear trend line equation was obtained. The absorbance values obtained with the fruit extracts were put into the equation and total folate was back calculated as μg/100 g fresh weight (FW).

3.7 Folate extraction and estimation:-LC-MS

3.7.1 Sample extraction procedure for LC-MS

Total folate was extracted following the procedures of Tyagi et al., (2015) from 52 accessions. Briefly, 100 mg homogenized tissue was suspended in 650 μL of extraction solution (50 mM potassium phosphate, 1% (w/v) ascorbic acid, 0.5% (v/v) β-mercaptoethanol, 1 mM calcium chloride, pH 4.5, flushed with nitrogen) in a 2 mL Eppendorf tube. The homogenate was boiled for 10 min, and then cooled on ice. Thereafter, 10 μL of α-amylase (20 mg/mL) was added and tubes were incubated at room

temperature for 10 min. Following that 2.5 μL protease (2 mg/mL) was added and incubation was carried out at 37°C for 1 h. The protease activity was terminated by transferring the tubes to boiling water bath for 5 min and cooling on ice. For deconjugation of folate polyglutamates to monoglutamates, 100 μL of rat plasma was added to each sample and tubes were incubated at 37°C for 2 h. Enzymatic activity was stopped by transferring the tubes to boiling water bath for 5 min and cooling on ice followed by centrifugation for 30 min (14,000 *g*, 4°C). The supernatant was filtered through the 0.22 μm filter (MDI Advanced Micro-devices) and the filtrate was ultra-filtered at 12,000 *g* at 4°C for 12 min using 10 kDa molecular weight cut off membrane filter (Pall Corporation, USA) for sample cleanup prior to LC-MS analysis. The resulting filtrate was transferred to an autosampler vial from which 7.5 μL was injected on column.

3.7.2 Liquid chromatography condition and mass spectrometry settings

For LC-MS, all the parameters used were essentially the same as described earlier by Tyagi et al. (2015). The folate derivatives were separated on a reversed phase Luna C18 column (5 μm particle size, 250 mm × 4.60 mm ID) (Phenomenex, USA) using Waters Acquity^{TM}UPLC system (Milford, USA) running in HPLC mode, coupled to a binary pump, an autosampler, and controlled by Xcalibur 3.0 software (Thermo Fisher Scientific, San Jose, USA). For mass spectrometry, Exactive^{TM}Plus Orbitrap mass spectrometer (Thermo Fisher Scientific, USA) was operated in alternating full scan and all ion fragmentation (AIF) mode equipped with positive heated electrospray ionization (ESI).

3.7.3 Folate quantification

External standards were used for folate quantification. Sensitivity was confirmed by evaluating the limit of detection (LOD; calculated as: $3.3\sigma/S$, where σ is the standard deviation and S is the slope of calibration curve) and limit of quantification (LOQ; calculated as: $10\sigma/S$). Least-square regression analysis was used for data fitting. After confirmation of individual peak identity on the basis of m/z and their fragmentation products, quantification was done according to the response of mass detector to the folate standard. Linearity of each folate standard was evaluated by plotting the peak area at different concentrations and sample concentrations were calculated from the equation $y=mx+c$. Endogenous residual folate of trienzyme (rat plasma+α-amylase+protease) was corrected by running blank samples and subtracting the values from the sample extracts. Sum of all the folate vitamers was expressed as microgram per 100 gram of fresh weight.

3.8 Extraction and estimation of pABA

For pABA analysis, method described by Navarrete et al. (2012) was used. Briefly, 100 mg homogenized tissue was suspended in 1000 µL cold methanol, mixed and centrifuged at 2600 *g* for 15 minutes at 4°C. The supernatant was transferred to a fresh tube and pellet was re-suspended in 1000 µL cold methanol for re-extraction. The sample was again centrifuged at 2600 *g* for 15 minutes and supernatant was collected. Both the supernatants were pooled and dried *in vacuo* at room temperature. 1000 µL of Milli-Q water was added to the dried sample and sonicated for 5 minutes. From this, 400 µL of the solvated portion was mixed with 50 µL of 2N HCl and incubated at 80°C for 2 hours in a thermomixer. The sample was cooled and 50 µL of 2M NaOH was added for neutralization of the reaction. Finally, the sample was passed through ultra-filter at 12851 *g* for 15 min and injected for HPLC-MS analysis. The LC-MS instrumentation and conditions for pABA analysis were- HPLC column C-18 phenomenex, 150 mm, particle size 5 µm, internal diameter 4.6 mm. Mobile Phase A (0.1% FA in H_2O), mobile phase B (0.1 % FA in methanol). pABA m/Z 138.0545 (retention time 5.03 min) was detected in positive mode.

3.9 Extraction and estimation of pterins

Pterin analysis was carried out according to Díaz de la Garza et al. (2004). Briefly, 100 mg tissue was suspended in 0.6 mL of methanol and mixed. To this, 250 µL of chloroform and 50 µL of water were added and shaken for 40 min at room temperature in a thermomixer. After this, another 225 µL of $CHCl_3$ and 340 µL of water was added to it and kept in shaking for another 20 minutes followed by centrifugation at 5000 *g* for 15 minutes to separate the layers. The upper, aqueous phase was transferred to fresh tubes, dried *in vacuo*, and dissolved in 100 µL of Milli-Q water. Samples were oxidized by adding 0.1 volume of a solution of 1% I_2 and 2% KI (w/v) in 1N HCl and incubated in darkness for 1 hour. Excess I_2 was removed by adding 10 µL of 5% (w/v) sodium ascorbate. Finally, the sample was ultra-filtered at 12851 *g* for 15 min and injected for LC-MS analysis. LC-MS instrumentation and conditions were- UPLC column C-18 hypersil gold 50 mm, 1.9 µm particle size, and internal diameter 4.6mm. Mobile Phase A (0.1% FA in H_2O), mobile phase B (0.1 % FA in acetonitrile). Pterins were detected in positive mode. The respective m/z and retention time for neopterin, monapterin, 6-hydroxymethylpterin and 6-carboxypterin were 254.0882, 254.0882, 194.0672, 208.0462 and 2.63, 3.33, 5.57, 5.68 minutes. Neopterin and monapterin were identified on the basis of retention time.

3.10 Total RNA extraction, cDNA synthesis and real time PCR

Total RNA was extracted using hot phenol method (Verwoerd et al., 1989). Powdered leaf and fruit pericarp tissue from three biological replicates at different developmental stages were taken. Approximately, 500 mg of powdered tissue was transferred to a falcon tube containing 1 mL of buffer A (0.1M LiCl, 0.01 M EDTA, 1% SDS, 0.1 M Tris [pH 9.0])/phenol mixture (1:1) preheated at 80°C and vortexed. To the tube 500 µL chloroform was added. After vortexing, tubes were kept horizontally on a rocker for 30 min at room temperature. After this incubation, samples were transferred to 2 mL microcentrifuge tubes and then were centrifuged at 11180 *g* for 10 min at 4°C. The upper phase was transferred to a new 1.5 mL microcentrifuge tubes and 1/3 volume of 8 M LiCl was added to it. After mixing gently samples were incubated overnight at 4°C. Next day, samples were centrifuged at 10,000 rpm for 30 min at 4°C. Supernatant was discarded and pellet was dissolved in 2 M LiCl followed by a short spin at 11180 *g* for 5 min. Supernatant was discarded and pellet was washed two times with 70% ethanol and finally with absolute ethanol by centrifuging at 11180 *g*, 10 min each wash. The pellet was air dried and dissolved in 20 µL of diethyl pyrocarbonate (DEPC) treated MQ and stored at -80°C till further use.

3.10.1 Determination of RNA quality

RNA concentration and quality was determined using an ND0100 Nanodrop UV-Vis spectrophotometer. To check the integrity of isolated RNA, 1-2 µg of RNA sample was mixed with 5 µL MiiliQ water and 2 µL 6X loading dye containing EtBr and loaded on agarose gel (1.2%, w/v) containing 4% (v/v) formaldehyde and 1X MOPS. All the solvents and the gel apparatus used were pretreated with 0.01% (v/v) diethyl pyrocarbonate (DEPC). For FA gel buffer, 10X MOPS buffer (200 mM MOPS, 50 mM sodium acetate, 10 mM EDTA, pH adjusted to 7 using 2N NaOH) was prepared beforehand. For 1 L of FA gel running buffer, 100 mL of 10X MOPS buffer, 20 mL of formaldehyde (37%, v/v) and 880 mL DEPC treated MilliQ water were mixed together and used fresh. Gel preparation was carried out inside a fume hood. Electrophoresis of RNA was performed in 1X FA (formaldehyde) buffer at constant voltage (50 V) for 1-2 hr. Thereafter, gel was visualized using UV filter in a gel documentation unit (Alpha Imager, India) and integrity was determined by checking the presence of two intact bands of 18S and 28S rRNA as a marker for RNA quality (Sambrook et al., 1989). The RNA was incubated with RNAse free DNAse (Promega) as per manufacturer's protocol to eliminate the genomic DNA contamination. The concentration of DNase was 1 unit for 1 µg of RNA.

3.10.2 cDNA synthesis and Quantitative Real-time PCR

Table 3.4. List of genes and primers sequence used for gene expression analysis of folate biosynthesis pathway genes.

Gene symbol	Primer sequence 5′ to 3′ (Forward & Reverse)	Primer length (bp)
GCHI	GTTATTCAGGATGCTGTTAGAGTCC & TCTTGTTCCTTGTCTTAGAGCCTTA	25, 25
DTPH	ATGAGGAAAATCTTGCATCACACT & CATACCATTCCCATCCATCACATT	24, 24
DHPA	CAAGTATCCAGAGGTATCTGCTGTT & GTATCTAATGATCTCGACACCCAAG	25, 25
DHPS	CACTTCGAGTACCCTATGTAGCAAT & CGTAAAGTTCAGATGCTACATCCTT	25, 25
ADCS	TCAATAGAGAAGGAAAGGGCTAGAT & CCTCCTTTACACATTCTGTCACTTC	25, 25
ADCL1	CAAGCTCTGCTAAATCTTATCTTGG & ATACTGTTCGAGCAAAAGCTACTGA	25, 25
ADCL2	GCACACTTCCTTTGTTACCTATCAT & TCTTCCCAAAGTAAATCTGAGAGTG	25, 25
DHFS	ACTTTATGCAACTTAGGATCAGTGG & GTTGACGTTAGCTTTAGACTTCTGC	25, 25
DHFR	GTTCAGGAAGTTTTGACATTGCTAC & AATAGAGAGACAATAAGGCGAGGAT	25, 25
FPGSp	TACTTCTCTGCTCTACGCTTCAAAT & AGTTGGCTTTATCTCATCCTTACCT	25, 25
FPGSm	GAGCTTGGACAAACGGTAGTATTTA & TTTCCAGGAGTGTATGAGGGTATAG	25, 25
GGH1	GAAGATTTCCCCCCGTGTTGCTAAAGA & CAATACCCTGAAAAAGCTACTCAAA	27, 25
GGH2	CCCGATTATATATACAGAGCCTCCT & CTTCAAAGTAGAGACCCTTCTTGCT	25, 25
GGH3	ACCAGAAGCTCAAAAAGTACTTGAC & GACTGTTGAAAACGTAAACCTCATC	25, 25
β actin	GTCCCTATTTACGAGGGTTATGC & CAGTTAAATCACGACCAGCAAGATT	23, 25
Ubiquitin 3	GCCGACTACAACATCCAGAAGG & TGCAACACAGCGAGCTTAACC	22, 21

cDNA was prepared from approximately 2 µg of RNA using High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, 4374967) following the manufacturer's instructions. Quantitative Real time PCR (qRT PCR) was carried out in AriaMx Real-Time PCR System (Agilent Technologies). The transcript abundance was measured in 10 µL volume of the SYBR Green PCR Master Mix (Takara, Japan) containing cDNA corresponding to 5 ng of total RNA with gene specific primers (Kilambi et al., 2013). The ΔCt value was calculated by normalizing Ct values of each gene to the mean expression of the two internal control genes (*β-actin* and *ubiquitin3*). Primers used in this study were designed from Primer 3. Primer sequences are given in Table 3.4.

Chapter 4

Folate estimation in tomato natural accessions using Microbiological Assay (MA) and Liquid chromatography-Mass Spectrometry (LC-MS)

4.1 INTRODUCTION

4.1.1 Microbiological assay with *Lactobacillus casei*

Lactobacillus casei subspecies *rhamnosus* (ATCC 7469) is used in the folate estimation as it is an auxotroph for folate and shows a directly proportional growth with increasing concentration of folate present in the growth medium. Since the advent of this technique, a number of microbes have been used for vitamin estimations. Initially, different bacteria viz. *L. casei*, *Streptococcus faecalis* and *Pediococcus cerevisiae* were used to determine various derivatives of folic acid and their relative abundance in the body (Grossowicz et al., 1962a, b). As the use of microorganism for folate estimation gathered pace, it was observed that results were often not reproducible among laboratories and there were issues with culture maintenance as well. The assay microorganisms were regularly maintained through serially diluted cultures which was both, cumbersome and produced varying results. However, Howard (1956) showed that direct low-temperature storage of test organisms in an appropriate suspension medium was possible without significant loss of virulence or the specific properties of the organism. This technique facilitated the storing of cultures for quick use.

Henderson and Huennekens (1974) showed that folate intake into *L. casei* is governed by many factors including pH, temperature, dependency on glucose, and occurs against the concentration gradient. Folate transport inside *L. casei* is greatly inhibited when there is excess amount of folate in the medium as showed by saturation kinetics experiments. Later on, a folate binding protein was isolated and characterized by Henderson et al. (1976) providing insights about the uptake and transport mechanism employed by *L. casei*. Shane and Stokstad (1975) reported that among *L. casei*, *S. faecalis* and *P. cerevisiae,* only *L. casei* was able to import and metabolize polyglutamates and various other folate forms. This study further established the use of *L. casei* as a choice organism for conducting folate estimations routinely.

Microbiological assay (MA) of vitamins is based on understanding of auxotrophic nature of certain microorganisms and this insightful discovery goes back to early 1940s (Arcot and Shreshtha, 2005). The microbiological assay of folate was started after Stokstad (1943) reported that liver and yeast extracts which are rich in folate, had positive effect on the growth of lactic acid bacteria such as *Lactobacillus rhamnosus* and *Streptococcus lactis*. This method soon gained wide acceptance and popularity and in fact, microbiological method of folate estimation was accorded Official Status by AOAC (Method 992.05 (2002)) and AACC (AACC Method 86-47). Other than *L. rhamnosus*, *Streptococcus faecalis* (ATCC 8043),

Table 4.1. Relative response of microorganisms to unhydrolyzed folate derivatives

	Lactobacillus casei	*Streptococcus faecium*	*Pediococcus cerevisiae*
Folic acid (PteGlu)	96	100	0
10-Formyl-PteGlu	100	100	0
Dihydrofolate	100	100	0
THF	100	100	100
10-Formyl-THF	99	100	100
5-Formyl-THF	100	100	100
5-Methyl-THF	100	0	0
$PteGlu_2$	100	100	0
$PteGlu_3$	100	0	0
$PteGlu_4$	66	—	—
$PteGlu_5$	20	—	—
$PteGlu_6$	4	—	—
$PteGlu_7$	2	—	—
10-Formyl-$H_4PteGlu_3$	82	—	—
10-Formyl-$H_4PteGlu_5$	3	—	—
5-Methyl-$H_4PteGlu_3$	90	0	0
5-Methyl-$H_4PteGlu_5$	2	0	0
5-Methyl-$H_4PteGlu_7$	1	—	—
Pteroate	0	100	0
10-Formyl-pteroate	0	100	0
5-Formyl-H_4 pteroate	0	+	0

Note:- the relative response is a comparison of folate concentration required to support half maximum growth rates of the bacterium.
Data adapted from Shane (2011).

Pediococcus cerevisiae (ATCC 8081), *Tetrahymena pyriformis* (geleii) (ATCC 30008), and *Bacillus coagulans* (Keagy, 1985; Hawkes and Villota, 1989) were also used for folate assay. However, later it was found that *S. faecalis* and *P. cerevisae* do not respond to methyl folates (Keagy, 1985). Gradually, *L. rhamnosus* was mainly used for microbiological assay of folate because of its almost equal response to all folate monoglutamate forms (Table 4.1) (Horne and Patterson, 1988 and Shane, 2011). A major improvement in aseptic addition methods was achieved when chloramphenicol resistant strain of *Lactobacillus casei* was used which significantly reduced the chances of contamination (Cooper and Jonas, 1973).

Since the MA of folate estimations was adopted for routine estimations in blood serum and food items, maintenance of the assay organism proved to be challenging. Preservation of bacterium through regular batch cultures was cumbersome and did not provide identical inoculum every time. Therefore, preparation and use of cryoprotected cells was started (Grossowicz et al., 1981 and Wilson and Horne, 1982) which could be stored for 6-8 months without loss of viability or assay response and gave better reproducibility. Another modification introduced in MA was its automation and carrying out the assay in 96 well microtiter plates (Newman and Tsai, 1986 and Horne and Patterson, 1988). The 96 well plates could be read in a microplate reader. Also, the assay sensitivity was higher (Horne and Patterson, 1988). Different optimizations in MA have been reviewed by Hyun and Tamura (2005) and Pandrangi and LaBorde (2004). The latter emphasized on the necessity for 'microaerophilic environment' for optimal growth of cryoprotected cells i.e. low oxygen and high carbon dioxide (5–10% CO2).

Subsequently, after the introduction of aforementioned changes, MA has been continuously used for folate estimations in various matrices for example, human serum (Spray, 1964), serum and red blood cells (O'Broin and Kelleher, 1992), cereal grain products (Rader et al., 1998), commonly consumed Australian vegetables (Iwatani et al., 2003), in regular Korean diet items (Yon and Hyun, 2003), potato (Goyer and Navarre, 2007), cut-packed fresh vegetables (Fajardo et al., 2015), Bangladeshi food items (Rahman et al., 2015) etc.

Currently Microbiological assay (MA) for folate estimation is the most popular method and has incorporated many modifications like use of chloramphenicol resistant strain of *Lactobacillus casei*, preparation and use of cryoprotected cells instead of batch culture and scaling to high-throughput analysis by using microtiter plates instead of tubes/flasks. It is

considered as an appropriate method of folate estimation which is high throughput, cost effective and easier to establish in any laboratory with minimal microbial expertise.

Considering above advantages MA was used to estimate total folate content in tomato natural accessions, for screening large as a high throughput method.

4.2 RESULTS

4.2.1 Response of *L. casei* for folic acid in the medium

Tomato accessions were grown in the year 2011-12 (October-March) in an open field at University of Hyderabad campus. The prevailing weather conditions for the growing season I (tissue collection for microbiological assay=MA) and season II (tissue collection for LC-MS) are shown in Table 4.2. The population was initiated with 391 accessions including the reference cultivar Arka Vikas but due to losses by various reasons, red ripe (RR) fruits with 3 or more replicates for only 160 accessions could be procured. From this, due to handling purposes, MA was carried out for 113 accessions with three or more replicates of RR fruit samples. The MA indirectly measures total folate present in a sample instead of various folate vitamers. The standardization of microbiological assay for folate estimation comprised two main steps- (1) standardization of suitable growth conditions for *Lactobacillus casei* with respect to space, aeration, inoculum, and amount of folate present in the medium, and (2) extraction of folate from tomato fruit tissue and appropriate inoculation dose during the assay.

4.2.2 Standardization of growth conditions

At first cryoprotected *Lactobacillus casei* cells were prepared according to Grossowicz et al. (1981) and several parameters like inoculation time, amount of inoculum, and concertation of folic acid in the growth medium were evaluated. These standardizations were done in 2 mL tubes and we observed a linear increase in bacterial growth till 0.4 ng folic acid (FA) (Figure 4.1). However, for high throughput folate estimations, the assay conditions were scaled down for microtiter plates, where again, inoculum size and incubation time was re-standardized. In microtiter plates the linear growth could be observed only till 0.1-0.15 ng after which it became irregular. These aberrations were most pronounced in 40 hours incubation and reproducible and consistent results were not obtained (Figure 4.2). Despite using the fresh solution of standard for every assay, sometimes unexpected loss of linearity was observed (Figure 4.2A). It was decided to try another method of preparation of cryoprotected cells.

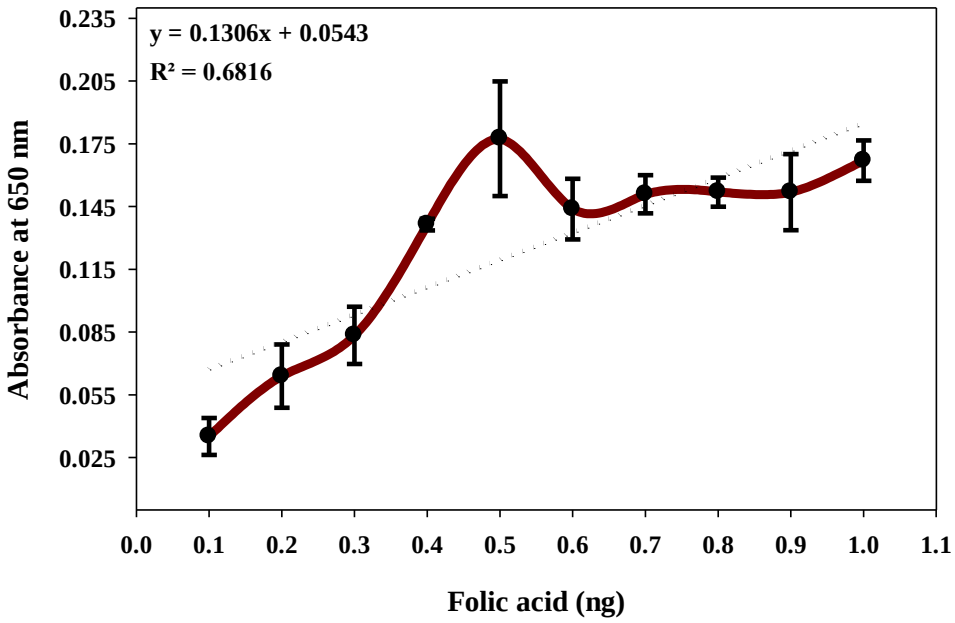


Figure 4.1. Standard curve for microbiological folate assay using glycerol cryoprotected *L. casei* cells prepared according to Grossowicz et al. (1981). **Note**:- The assay was performed in 2 mL microcentrifuge tubes containing 1 mL assay mixture.

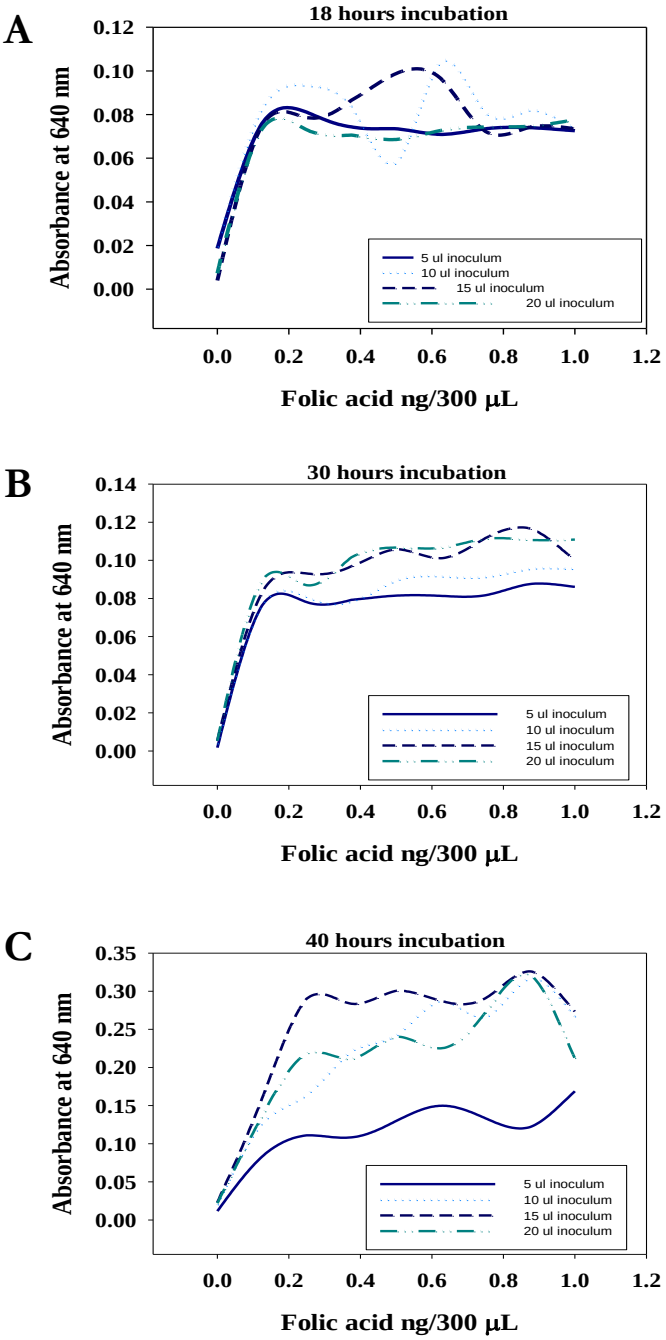


Figure 4.2. Standard curve for microbiological folate assay using glycerol cryoprotected *L. casei* cells prepared according to Grossowicz et al. (1981).
Note:- The assay was performed in microtiter plates containing 300 µL of assay mixture. Variations in inoculum size (box) and incubation time (A=18 hours, B=30 hours, C=40 hours) were analyzed simultaneously.

Table 4.2. Average temperature and humidity for the months during which tomato accessions were grown in the year 2011-12 and 2012-13 at University of Hyderabad, Hyderabad.
Note:- The data for Hyderabad weather were retrieved from 'Weather Underground' (**https://www.wunderground.com/**).

Season I- October 2011 to February 2012			
Month	**Temperature (Max.) °C**	**Temperature (Min.) °C**	**Relative Humidity (%)**
October	30	21	49
November	29	17	45
December	29	16	51
January	31	17	51
February	33	19	41
Season II- October 2012 to February 2013			
Month	**Temperature (Max.) °C**	**Temperature (Min.) °C**	**Relative Humidity (%)**
October	31	20	56
November	30	18	55
December	30	17	45
January	31	18	45
February	31	19	43

Therefore, we changed the protocol and prepared a fresh batch of cryoprotected cells using another protocol (Wilson and Horne, 1982). The method of Wilson and Horne (1982) was both, brief and simpler. In this protocol *L. casei* culture was not grown first on folate supplied and then on folate efficient medium (to deplete the intracellular folate), with pelleting and washing steps. Wilson and Horne (1982) mixed the active culture with the ice cold 80% (v/v) glycerol and stored the aliquots. A major improvement was significant reduction in assay response time (30-40 hours in Grossowicz et al., 1981 and 17-18 hours in Wilson and Horne, 1982). The efficiency of cryoprotected cells made according to Wilson and Horne (1982) for folate estimation was examined in microtiter plates. Therefore the data from this method cannot directly compared with that of Grossowicz et al. (1981) where we used 2 mL tubes. Nonetheless Wilson and Horne (1982) not only the reduced the assay duration but, also improved the sensitivity of the assay by 10-20 folds. Similar results were obtained by Horne and Patterson (1988) where 10 fold more sensitivity was observed between assays conducted in 3 mL tubes and 96 well microtiter plates.

Cells prepared using Wilson and Horne (1982) protocol gave better response and a *L. casei* growth curve was plotted against increasing folic acid concentrations. The growth of cells was linear with increasing amount of folate and growth curve showed a good correlation value (R^2=0.98). The equation obtained through this curve was used for all subsequent estimations of folate in tomato samples (Figure 4.3). Another advantage of this method was that a single batch of cryoprotected cells were sufficient for assaying large number of samples, as the cells are diluted by 25 times before inoculation. Major differences between Wilson and Horne (1982) and Grossowicz et al., (1981) are summarized in Table 4.3.

4.2.3 Folate extraction from tomato (*Solanum lycopersicum*) accessions

Our next objective was to extract folate from tomato fruits from a large population that needed a high throughput method. The protocol also required to be applicable for a variety of tomato germplasm with obviously different composition of red ripe (RR) fruits. Goyer and Navarre (2007) reported 'trienzyme extraction and microbiological assay of folate in a number of potato germplasm. This protocol seemed suitable for application to tomato considering that potato is a Solanaceae member. Tri-enzyme treatment is routinely used for folate estimation and rat plasma is used as a source of conjugase enzyme which produces monoglutamates. The Goyer and Navarre (2007) method was adopted for MA estimation of folate albeit that 100 mM potassium phosphate buffer, pH 7 instead of 50

mM HEPES/50 mM CHES, pH 7.85 was used for extraction. Though, HEPES and CHES qualify as buffers of choice in biological systems but due to high cost of these buffers, inexpensive potassium phosphate buffer was opted. Comparison of extraction efficiencies of both the buffers showed similar efficiency for folate extraction for tomato and other vegetable samples (spinach, potato, green chili and capsicum) (Figure 4.4).

For subsequent experiments potassium phosphate extraction buffer (0.1 M potassium phosphate pH 7.0, was used with 1% (w/v) ascorbic acid and 0.1% (v/v) β-mercaptoethanol, flushed with liquid nitrogen). Table 4.4 enlists the number of accessions for which folate was estimated from NBPGR, TGRC, IIVR, Bejo Sheetal Seed Company, and reference cultivar Arka Vikas from IIHR. Total folate content between the accessions ranged from 5-60 μg/100 g FW. Arka Vikas showed folate content 40 μg/100 g FW. Majority of the accessions showed folate content ranging from 20-40 μg/100 g FW and this group consisted of s 51% of all the accessions while 36% accessions had total folate ranging from 5-20 μg/100 g FW (Table 4.5, Figure 4.5). Only 13% accessions exhibited more than 40 μg/100 g FW of total folate. Though, 3 or more replicates were used for each accession, yet variations in folate levels due to individual differences among the biological replicates grown in open field which resulted in higher values of standard error.

4.3 DISCUSSION

4.3.1 Folate extraction and estimation

Folate extraction is a complex procedure considering the labile nature of natural folates. The method of folate estimation involves thermal (boiling) and enzymatic release of all matrix bound folate, followed by deconjugation of polyglutamylated folate forms into monoglutamylated state, sample cleanup and estimation. The extraction process involved three enzymatic steps along with intermittent boiling to denature the enzyme (protease). Selection of conjugase source largely depends upon the desired end product. Rat plasma conjugase efficiently de-glutamylates folate to monoglutamate form, has a pH optima between 6.2-7.5, is fairly stable during long term storage and its endogenous folate can be easily removed through charcoal treatment (Horne et al., 1981). These properties make it an ideal source of conjugase for folate extraction. Rat plasma conjugase exhibits its optimum activity in 6.2-7.5 pH range (Horne et al., 1981), while the optimum pH range of α-amylase and protease used in this study is 6.0-7.0 and 7.0-8.0 respectively. Considering that most of the folates are fairly stable at 7 or above pH, therefore the extraction buffer pH of 7 was selected.

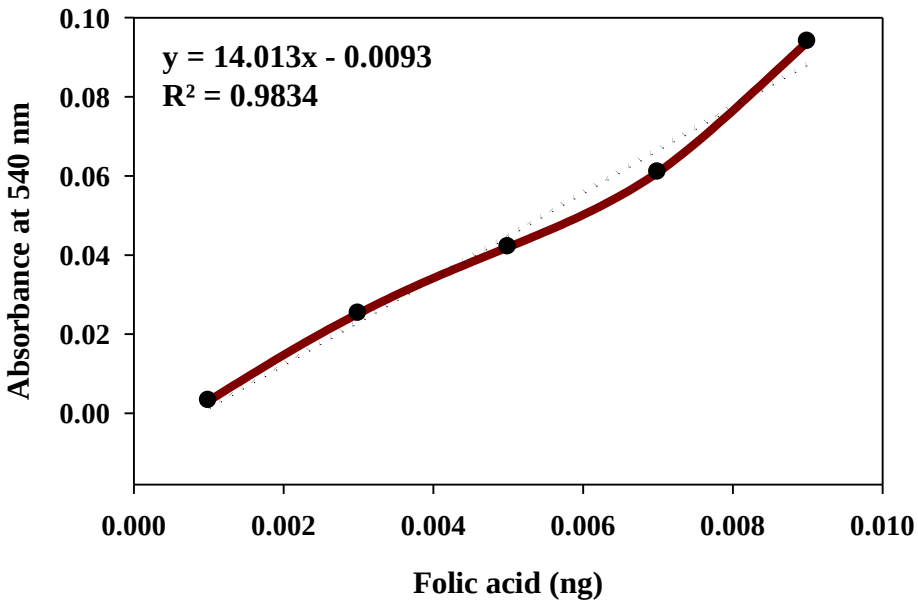


Figure 4.3. Standard curve for microbiological folate assay using glycerol cryoprotected *L. casei* cells prepared according to Wilson and Horne (1982).

Note:- The assay was performed in 200 µL of assay mixture and the doses of folic acid were used from 1 pg to 8 pg.

Table 4.3. Comparison between the two methods of cryoprotected cells preparation.

Parameters/steps	**Grossowicz et al. (1981)**	**Wilson and Horne (1982)**
Preparation steps	Multiple	Few
Absorbance at	650 nm	540 nm
Assay time	30-48 hours	18 hours
Effect of heavy inoculum	Growth inhibitory	Growth increased
Use of phosphate buffer (pH 6.1)	Not used in the assay mixture	Used in the assay mixture

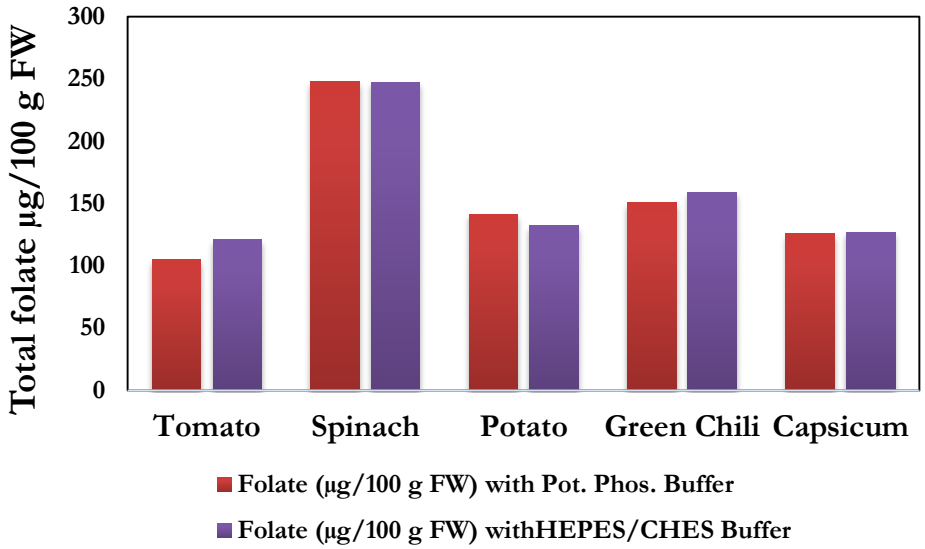


Figure 4.4. Comparison of folate extraction efficiency of potassium phosphate and HEPES/CHES buffer for different samples

Table 4.4. Source and number of accessions used for folate estimation using microbiological assay.

Source	Number of accessions
NBPGR	68
TGRC	20
IIVR	21
Bejo Sheetal	4
IIHR	1
Total=113 accessions	

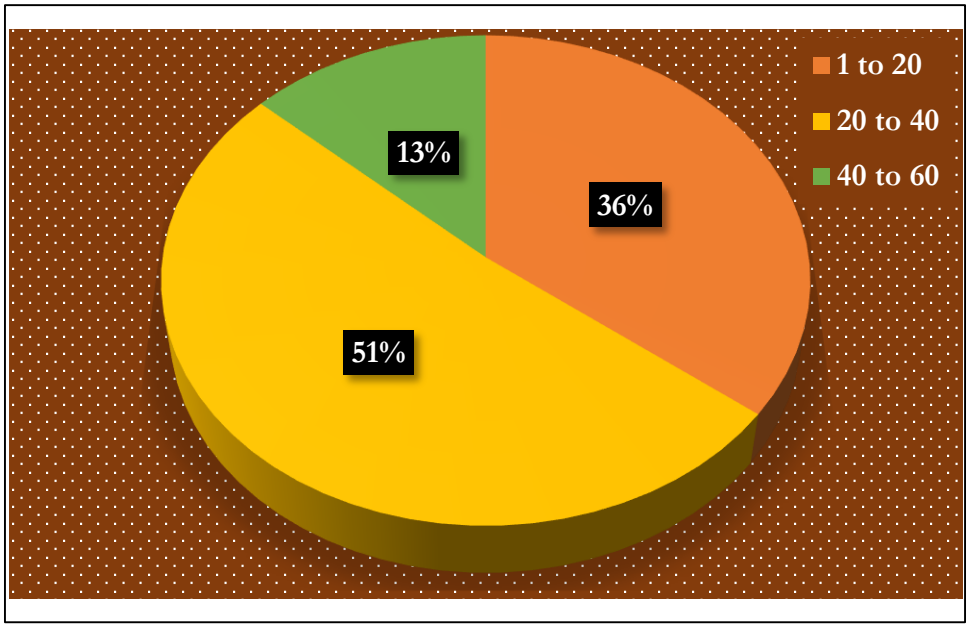


Figure 4.5. Distribution of folate content among accessions. Ranges represent Folate (µg/100 g FW) for each group of accessions, number of accessions falling in a particular folate range are shown as % value of all the accessions. Folate estimated through MA

Table 4.5. Grouping of 113 tomato accessions with respect to total folate level in red ripe fruits. The folate level was estimated using microbiological assay.

Group	Total folate content (µg/100 gm FW)	No. of accessions	% of total accessions
1	<20	40	35.4
2	20-30	27	23.9
3	30-40	31	27.4
4	40-50	13	11.5
5	50-60	2	1.8

The folates exist in various forms such as Tetrahydrofolate (THF), 5-Methyltetrahydrofolate (5-CH3-THF), 5,10-Methenyltetrahydrofolate (5,10-CH^+-THF), 5-formyltetrahydrofolate (5-CHO-THF) and 5,10-methylenetetrahydrofolate (5,10-CH_2-THF) and these forms readily interconvert among each other with the change in pH. However, MA estimates only total folate content in a sample therefore these interconversions do not affect the final values. For complete extraction, pH of the buffer is of lesser importance (for MA) than the enzymes which are most critical for release of bound folate (Hyun and Tamura, 2005). However, buffer pH is of importance in case of chromatographic estimation methods (Discussed in a latter section).

Heat and enzymatic treatments in various studies are non-complementary in nature and in addition inter and intra assay variations can exist for folate estimation both within the laboratory as well as between laboratories depending on the protocols and durations of treatments. Martin et al. (1989) first reported that use of trienzyme (protease+amylase+conjugase) treatment results in total higher folate value than those obtained with conjugase treatment alone. But, later many studies showed that conjugase treatment alone is equal to the trienzyme method (Konings, 1998). Despite the disparities among the extraction method and folate values, it is now established that pH of extraction buffer should be optimal for all enzymes used in a particular protocol, and heat treatments should be kept to a minimum if it does not lead to an appreciable increase in folate. Protease enzyme eventually releases all the matrix bound folates. Therefore, it should not be added along with amylase and conjugase as it might degrade the latter enzymes. The longer incubation times should also be avoided, instead enzyme units can be increased to complete the folate extraction in shorter time. The centrifugation is used as a last step to remove debris, if performed immediately after first boiling, it results in losses of any folate still bound to the matrix. Apart from this, if the manipulations (addition of enzymes, mixing, and supernatant transfer) are performed under a laminar air flow bench then the use of 0.22 μm membrane filters to eliminate microbial contamination can be avoided.

Our study shows that phosphate buffer can be used in place of HEPES/CHES which substantially reduces the cost of extraction for high-throughput analysis.

4.3.2 Folate content in tomato accessions

The folate levels among red ripe fruits of 113 tomato accessions varied from 4.5 to 60 μg/100 g FW among the accessions while reference cultivar Arka Vikas showed 39 μg/100 g FW folate (Figure 4.6). Median folate value was 25.6 μg/100 g FW. So far, total folate in tomato is reported to be 9-29 μg/100 g FW folate (Bekaert et al., 2007) while

another study showed 4.1-35.3 μg/100 g FW folate in few tomato cultivars (Iniesta et al., 2009). These studies indicate that tomato is not a rich source of folate, but number of samples used were too few to make a general statement. In the present study screening of 113 tomato accessions showed much wider folate range than the previous reports, which might be due to the larger sample size used in this study. Though, out of 391 natural accessions grown for this purpose, we could procure more than 3 replicates for only 113 accessions. However if more accessions are screened it would broaden the range of folate levels in tomato and may identify few accessions enriched in the folate levels.

4.3.3 Accuracy of the microbiological assay

Though, MA is an easy and cost effective set up for folate estimation, it leads to imprecise results. MA estimates the folate content by measuring the proportional increase in turbidity of culture in relation to exogenous folate levels. Though MA cannot distinguish between different forms of folate, it does provide an approximate estimation of total folate levels. A major limitation of this assay is that *L. casei* growth response differs to different forms of folate leading to imprecise results (Freisleben et al., 2002). This assay is also susceptible to extraneous folate and contamination by other microbes (Quinlivan et al., 2006). Notwithstanding above limitations, due to its inexpensive nature, MA is routinely used for estimation of total folate in food samples. Our results indicate that microbiological assay was inaccurate, and it either overestimated the folate level or underestimated the folate levels compared to LC-MS assay (See later section).

Comparison of relative efficiency of MA and LC-MS for estimation in human serum showed good correspondence between both assays (Fazili et al., 2007). On the other hand, in ready-to-eat breakfast cereals, this assay overestimated folic acid levels by 10–67% than the LC-MS based estimation (Phillips et al., 2010). Given the wide range of differences observed between MA and LC-MS assay, the latter assay more precisely estimates folate levels. Moreover, LC-MS method also distinguishes different folate species and is less susceptible to interference from other metabolites and/or inhibitors. Though MA gives an imprecise estimation, being cost-effective and easy to setup, it is more widely used for folate estimations. It is desirable that MA estimation of folate levels is also validated by LC-MS method for precise estimations. Therefore, after screening total folate in year 2011-12 tomato population, accessions were grown again in 2012-13 for estimation of different natural forms of folate through LC-MS in both mature green (MG) and red ripe (RR) fruits.

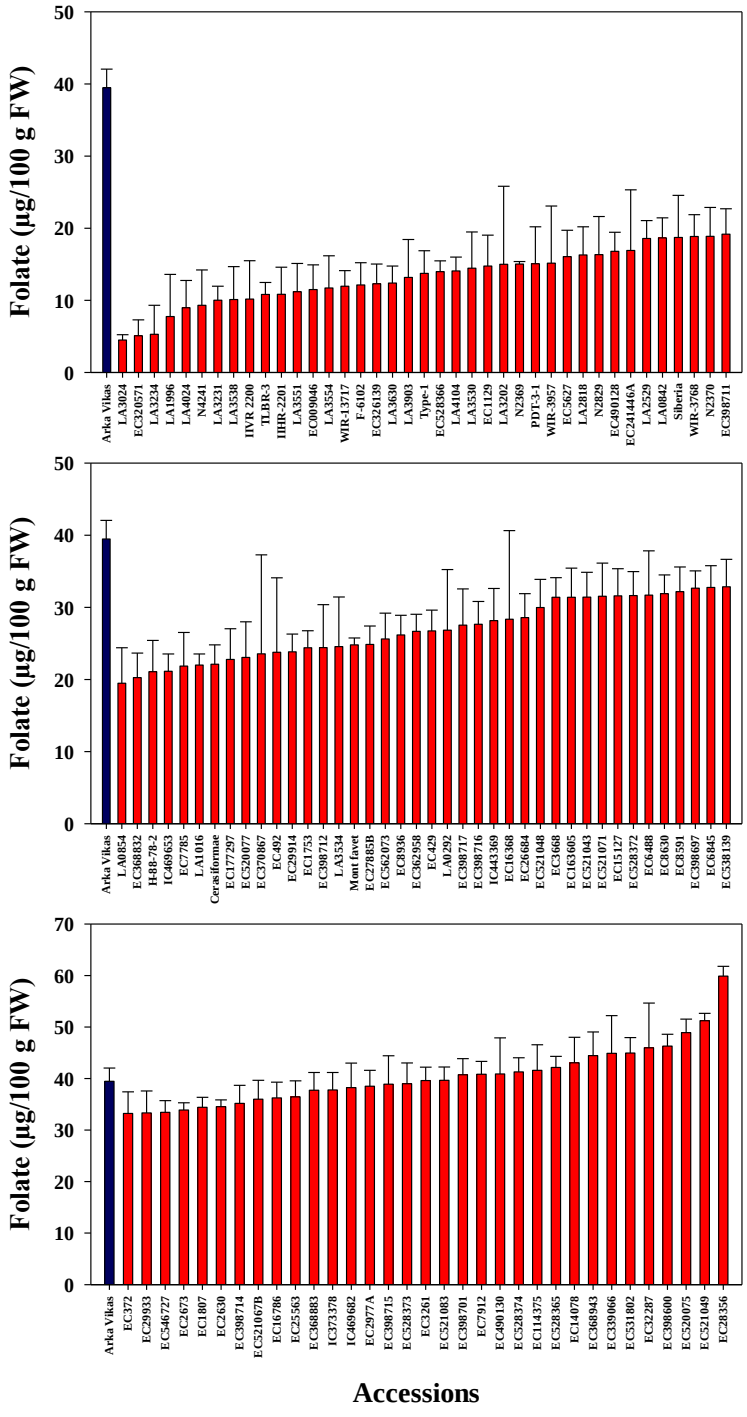


Figure 4.6. Total folate content among 113 tomato accessions in red ripe fruit. Folate was estimated using microbiological assay. The blue bar in every graph shows folate level of reference cultivar Arka Vikas, other accessions have been arranged in increasing order of folate content.

In conclusion, the preliminary estimation of folate among tomato accessions provided valuable information about its variations in tomato germplasm. Though, folate content screening has been endeavored in crops such as potato, wheat, spinach, etc., similar studies are lacking for a number of tomato germplasm. This study highlights the potential that tomatoes as the second most produced and consumed vegetable crop of India can offer in fighting the folate deficiency. However, the results can be validated using latest and high precision methods for accurate selection of the high folate varieties. Therefore, the folate estimations were carried out in the tomato accessions using the LC-MS method of folate estimation as described in next section 4.4.

Folate estimation in tomato natural accessions using Liquid Chromatography-Mass Spectrometry (LC-MS)

4.4 INTRODUCTION

. The advent of new technologies replaced microbial assay and allowed more precise quantifications of different folate vitamers. The microbial assay was followed by folate detection and quantification by chromatographic methods combined with UV, fluorescent or electrochemical detectors. The new detection methods not only increased the precision and the accuracy of the results, but also allowed detection and quantification of various folate derivatives and also their relative abundance. The combination of above methods with mass spectrometers (MS) further improved the qualitative and quantitative estimation of folates. The MS detection systems coupled with HPLC since 1990s allowed the separation and detection of four monoglutamates viz. THF, 5-Methyl THF, folic acid and 5 or 10-formyl THF in variety of the products such as multivitamin tablet, breakfast cereal, and beef and vegetable extract (Stokes and Webb, 1999).

The two main aspects of folate estimation are extraction and detection. Though the usage of trienzyme remains the main procedure to release folates from a variety of tissue and matrices (described in earlier sections), chromatographic methods have additional requirement of sample cleanup before injection (Arcot and Shreshtha, 2005). Sample cleanup is essential in removing the co-eluting compounds that can significantly hinder the detection performance. Main cleanup methods for folates are solid phase extraction (SPE) using strong anion exchange resin or affinity chromatography using folate binding protein (FBP) column. Among these two, SPE is more often used though the process of SPE requires analytical expertise. The commercial unavailability of FBP columns and long steps in preparing them in house makes them less popular and low throughput. Additionally FBP shows lesser affinity towards 5-formyl-tetrahydrofolate and therefore folate sample loading exceeding 25% of the capacity causes almost 10% loss of 5-formyl-tetrahydrofolate (Selhub, 1989; Pfeiffer et al., 1997; Konings, 1998). This loss occurs due to competition among folate species for the folate binding sites. A major improvement in sample cleanup was reported by Zhang et al. (2005) showing that ultrafiltration of the extracts with 5 kDa weight cut off membrane could effectively remove most of the polymers present in the plant matrices without affecting the recovery of folates. The ultrafiltration of the extracts simplified the entire process of folate extraction from homogenization to injection into HPLC.

In addition to modification of extraction process, detection of folate derivatives in various tissues was also improved. Pawlosky et al. (2003) compared HPLC-fluorescence detection and LC-MS system using a number of food reference materials and found highly comparable values between the two methods. This established HPLC-fluorescence detection as a reliable method. Nelson et al. (2004) compared solid phase affinity extraction (SPAE) using FBP and solid phase extraction (SPE) using C_{18} columns and MS detection and showed that SPAE was 10 times more sensitive than the SPE method. Above modifications significantly increased the precision and ease of folate detection and quantification. A method of folate extraction and estimation in tomato was reported which equally applied to tomato and other vegetables (Tyagi et al., 2015).

The LC-MS based detection for folate estimation has been used for several species such as spinach (Zhang et al., 2005), rice (De Brouwer et al., 2008), and various vegetables (Wang et al., 2010). With each study, either more number of folate species were successfully detected or their relative proportions in a sample were reported. Currently LC-MS based folate detection has been increasingly used for estimation of folate in the food matrices, different plant species and also for estimation of folate and related metabolites in plasma and other tissues in pregnancy and other medical conditions.

4.5 RESULTS

4.5.1 Folate extraction and estimation through LC-MS:-Optimization of parameters

Folate estimation using LC-MS method is different from microbiological assay (MA) in both extraction of folate from tissue as well as detection and quantification. Folate was estimated in 100 mg mature green and red ripe tomato fruit tissue using the protocol of Tyagi et al. (2015). This protocol is high throughput, cost effective and easier than many previous protocols with respect to various parameters described here. The molarity of potassium phosphate extraction buffer was 50 mM instead of 100 mM as used commonly in microbiological assay related protocols. The objective was to maintain the buffering capacity but keeping the salt content in the final extract to a minimum as salt seriously hinders with the ionization process.

A major difference was in sample cleanup which was adopted from Zhang et al., (2005). Folate extract cleanup and concentrations usually accomplished using solid phase extraction (SPE) and more recently with the help of folate binding protein (FBP) assisted affinity chromatography. SPE involves various strong anion exchange resins and elution of folate extract from these columns. It results in- (1) excessive dilution of extract during the elution and (2) higher concentration of salts in the extract used in the elution buffer.

Less abundant folate forms might not be detected due to excess dilution of the extract. There is also a chance that due to prolonged procedure of SPE, labile folates may degrade resulting in under-estimation of total folate. On the other hand, affinity chromatography which relies on FBP is a better approach of folate extract cleanup and concentration. However, various folate derivatives exhibit different binding property towards FBP therefore, this purification approach is better for tissue/samples which have predominantly single type of folate. Moreover, FBP columns are not commercially available, which is a major limitation in using this type of purification method. Contrary to this, cleanup of folate extracts using molecular weight cut off membrane filters was found to be a much better option with high precision mass spectrometry. Ultrafiltration removed polymers from the extract which are commonly associated with the plant sample matrix.

4.5.2 Folate content in tomato accessions

A total of 391 accessions were grown for folate estimation in tomato. However, after sorting of accession on the basis of availability of 3 or more replicates and proper handling of samples, folate levels in 52 accessions in red ripe and 27 accessions (overlapping with RR) in mature green fruits was analyzed using LC-MS. The folate values ranged from 18-58 μg/100 g FW and 9-42 μg/100 g FW in MG and RR stage respectively (Figure 4.7 and 4.8, Table 4.6). In earlier studies, the tomato folate content was reported to be 9-29 μg/100 g FW and 4.1-35.3 μg/100 g FW (Bekaert et al., 2007; Iniesta et al., 2009) and values reported in present study were similar to this range. However, the reported folate values in this study showed a broader spectrum of folate content in tomato which might be due to more number of accessions used in this study.

4.5.3 Relative proportions of folate vitamers among accessions

We detected four folate derivatives viz., THF, 5-Methyl THF, 5-Formyl THF and 5,10-Methenyl THF. However, THF was detected only in MG fruits of few accessions. The most abundant folate derivative was 5-Methyl THF which constituted 74-94% and 59-88% of total folate in MG and RR tomato fruits. The above abundance and range of folate values were in congruence with the previously reported distribution of folate forms (Rébeillé et al., 2006) (Figure 4.9 A, B). 5-Formyl THF was the second most abundant folate detected in both MG and RR tomato fruits. The predominance of 5-CH_3-THF form a useful trait since it is reported to be the most bioavailable form in animals (Scott et al., 2000) making it the preferred form for the fortification of food items (Scott et al., 2000). It is also one of the stable folate forms at different pH (2 to 10) (De Brouwer et al., 2007). Therefore, accessions with higher levels of 5-CH_3-THF can be ideal parental lines for

Table 4.6. Grouping of tomato accessions with respect to total folate content in fruits. The folate levels were estimated using LC-MS. Fruits of 27 accessions were analyzed at the mature green stage, and 52 accessions were analyzed at the red ripe stage.

Group	Total folate content (µg/100 gm FW)	Mature green stage		Red ripe stage	
		No. of accessions	%	No. of accessions	%
1	<20	2	7.4	9	17.3
2	20-30	7	25.9	31	59.6
3	30-40	9	33.3	10	19.2
4	40-50	5	18.5	2	3.8
5	50-60	4	14.8	-	-

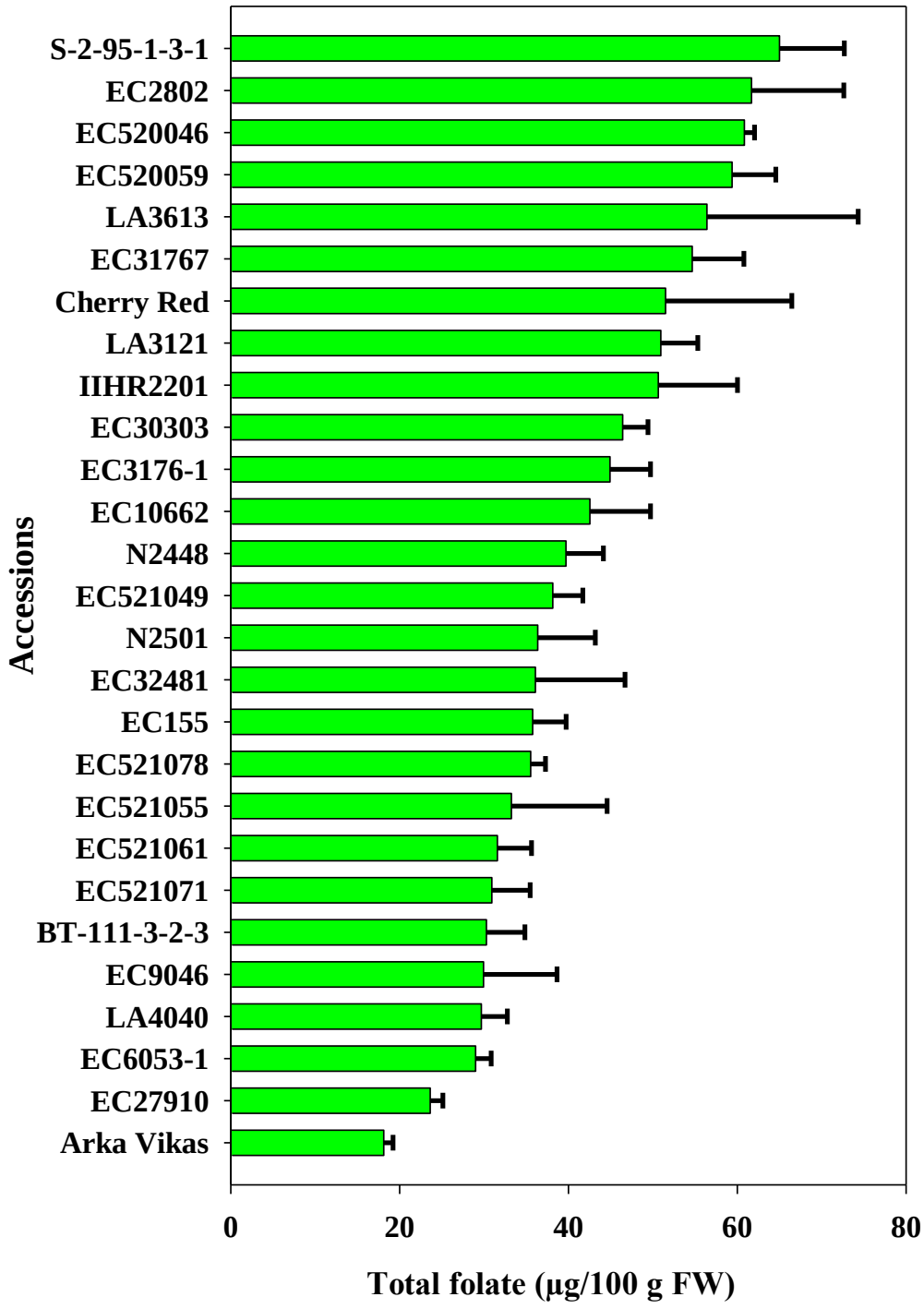


Figure 4.7. Total folate content in mature green fruits of 27 tomato accessions. Folate was estimated through LC-MS.

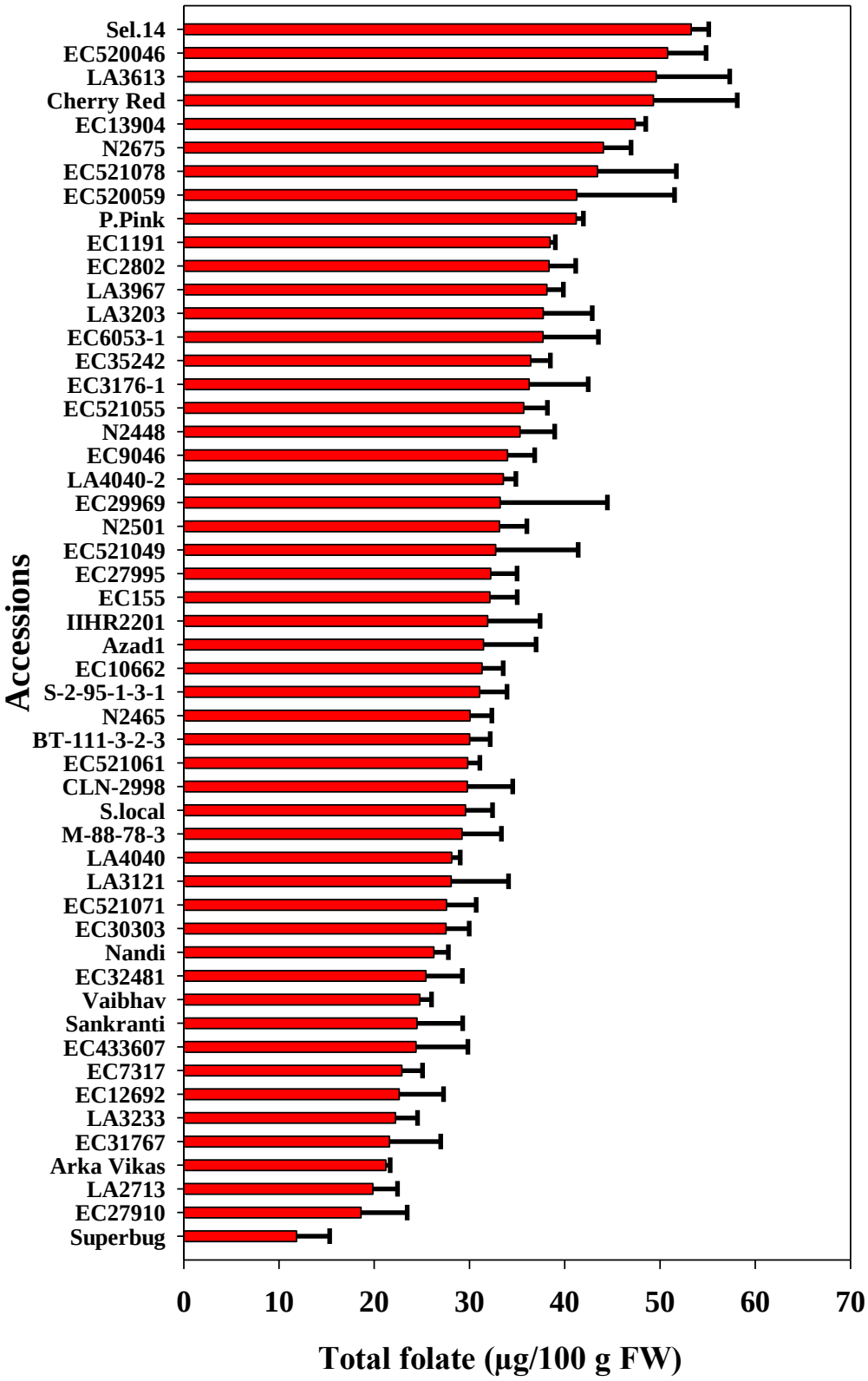


Figure 4.8. Total folate content in red ripe fruits of 52 tomato accessions. Folate was estimated through LC-MS.

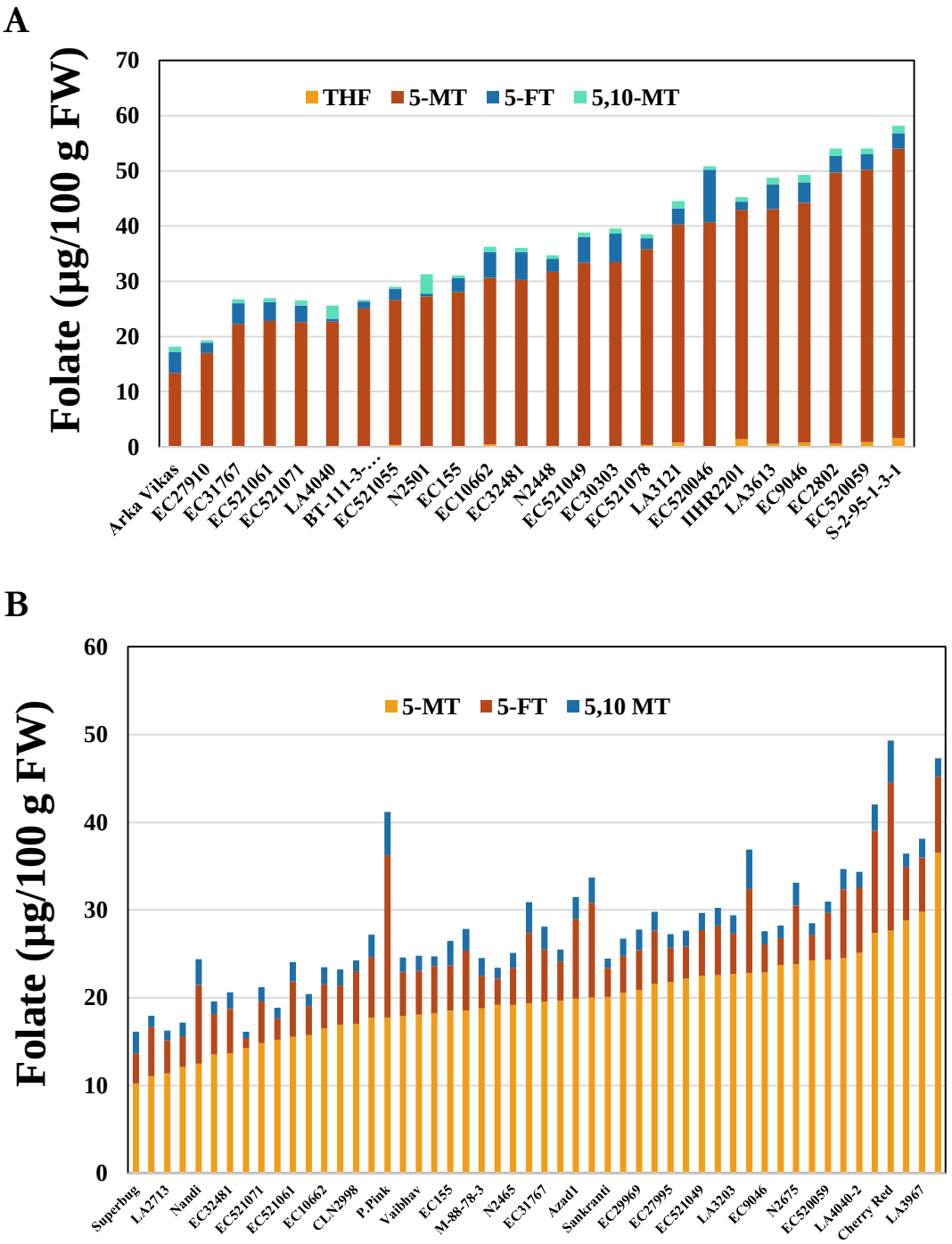


Figure 4.9. Different folate forms among accessions estimated through LC-MS Accessions arranged in increasing order of 5-Methyl THF. A=Mature green, B=Red ripe

breeding based biofortification approach. Tomato fruits are enriched in 5-CH_3-THF with 69.0% contribution to total folate in red ripe fruits.

4.5.4 Tomato fruit ripening and folate content

The folate content in tomato fruits either remained unaltered or decreased during the ripening process (Figure 4.10). Among the 27 accessions for which MG and RR data were compared, 11 accessions showed a decline in folate content ranging from 30 to 56% as the fruits ripened. Eight accessions showed less than 10% decrease in folate content in transition from MG to RR; while remaining 8 accessions showed little decline (11-23%) with the progression of ripening. Similar patterns were also observed in other studies (Iniesta et al, 2009; Strålsjo et al., 2003) and no specific pattern was discernible between folate content and ripening.

4.5.5 Folate content and fruit weight of wild species of tomato

The total folate content was measured in three wild relatives of tomato namely *S. cheesmaniae*, *S. pimpinellifoium* and *S. habrochaites* along with the *S. lycopersicum* cv. Ailsa Craig using LC-MS. These species were grown in open field and the values were compared with Ailsa Craig grown simultaneously. *S. cheesmaniae*, *S. pimpinellifoium* and *S. habrochaites* show orange, red and green color in their ripe stage respectively (Figure 4.11). Folate content was measured in the 39 DPA (days post anthesis) fruit and compared with the 39 DPA folate value of Ailsa Craig. The folate content in *S. cheesmaniae*, *S. pimpinellifoium* and *S. habrochaites* was 64, 73 and 77 μg/100 g FW while it was 47 μg/100 g FW in the Ailsa Craig. Folate content in the three wild species was 1.4 to 1.6 times higher as compared to Ailsa Craig (Figure 4.12). The fruit weight of *S. cheesmaniae*, *S. pimpinellifoium, S. habrochaites* and *S. lycopersicum* cv. Ailsa Craig was 0.45, 0.62, 1.72 and 28.33 g respectively.

4.5.6 Correlation between folate content and fruit weight

The correlation between 39 DPA fruit folate content (μg/100 g FW) and fruit weight (g) was calculated for the accessions where fruit weight data was available. The correlation between folate content measured by MA and by LC-MS and fruit weight was calculated separately and was -0.395 and -0.052 respectively (Table 4.7). The correlation between the folate content and fruit weight in wild relative was -0.896 suggesting a very strong negative correlation between the two traits (Figure 4.13, Table 4.7).

4.6 DISCUSSION

4.6.1 Variations in folate content and distribution of different forms among accessions

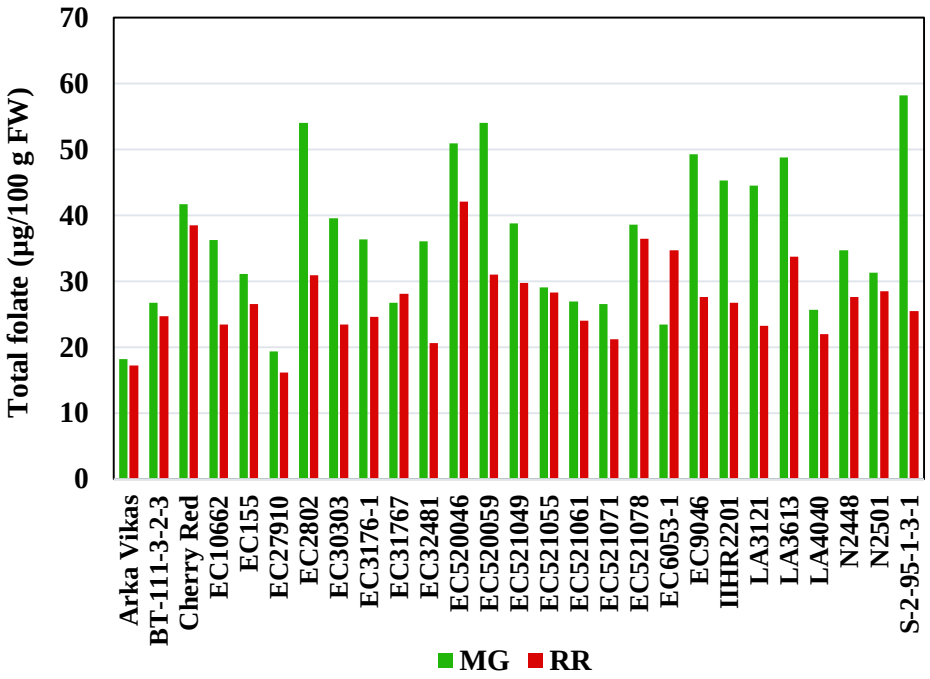


Figure 4.10. Comparison of MG and RR total folate in 27 accessions to tomato. Most accessions showed decreased folate content as the fruit attained ripening.

Screening of 52 tomato accessions helped in identification of high and low folate varieties which can be further investigated to understand governing factors causing higher and lower folate accumulation in the fruits of these accessions. The reference cultivar Arka Vikas (AV) showed 18 and 17 μg/100 g FW folate in MG and RR stages of fruit development and was lowest and second lowest with respect to total folate content in both MG and RR stages among other accessions. Similar studies of screening of cultivars of a particular crop show that folate levels show a wide variations based on genotype, location and number of cultivars used in the study. In a study showing folate content of 10 lentil cultivars showed 216-290 μg/100 g dry weight folate while another study including 4 cultivars showed 136-182 μg/100 g. Also, folate content varied among chickpea (42-125 and 351-589 μg/100 g) and pea cultivars (Sen Gupta et al., 2013; Jha et al., 2015). Studies encompassing several cultivars showed equally variable ranges of folate content as in 175 wheat genotypes (364-774 ng/g of dry matter in winter wheat and from 323-741 ng/g of dry matter in spring wheat) (Piironen et al., 2008), 67 spinach accessions (54-173 μg/100 g FW) (Shohag et al., 2011), 13 strawberry cultivars (335-644 μg/100 g of dry matter) (Strålsjö et al., 2003) and 67 potato cultivars (521-1373 ng/g of dry weight) (Goyer and Navarre, 2007). There are many factors which might govern total folate content. Since, all the accessions were grown in similar conditions, variable folate level indicated towards possible genotypic differences.

The relative distribution of folate vitamers among accessions in both, MG and RR fruits showed an overall prevalence of 5-methylTHF followed by 5-formylTHF. 5-MethylTHF is a stable folate derivative as well as the most bioavailable form for animal systems. Similar dominance of 5-methylTHF has been reported by Vahteristo et al. (1997) in a number of fruits, vegetables and canned food items. This is also the major circulatory form of folate detected in blood. Therefore, despite the genotypic differences, consistently higher contribution of this form in tomato folate is a beneficial trait for fortification approaches.

4.6.2 Variation of folate content with the ripening

In the absence of a clear pattern of decreasing folate levels with the progression of ripening, no conclusion could be drawn as to what is regulating the folate accumulation. Our results and outcomes of these studies indicate that though, a greater natural potential remains to be tapped and more genotypes should be screened, still focus should be shifted towards deciphering the factors which cause such differences.

4.6.3 Association between total folate and fruit weight

Our objective of screening for folate enriched accessions was to find varieties that can provide a substantial amount of folate recommended dietary allowance (RDA) in their standard serving size. However, the high folate content of wild relatives and their small fruit size prompted us to check the correlation between the folate content and fruit weight. Interestingly, we observed a strong negative correlation of -0.896 between these two traits among *S. cheesmaniae*, *S. pimpinellifoium* and *S. habrochaites* compared to the *S. lycopersicum* cv. Ailsa Craig. However, a very weak negative correlation (-0.395, -.051) was observed between total folate content and fruit weight among tomato accessions used in this study (Table 4.7). Similar results were also obtained by Ward et al. (2008) in wheat genotypes and Shohag et al. (2011) in spinach genotypes where folate content either showed no correlation or weak negative correlation with the biomass.

These results provide interesting insights towards the possible effects of domestication on tomato. So far, tomatoes have been selected for color, sugar, acidity, flavor, aroma, size and yield related traits and folate content was not the criteria for selection. Therefore, this trait might have been randomly distributed among the accessions irrespective of the fruit size. Perhaps, that is why no strong correlation was observed between the fruit weight and folate content among the natural accessions. At the same time, wild relatives of tomato have not been subjected to domestication and still preserve the fine balance between the biomass and metabolite buildup but this needs further studies and validation.

The objective of this study was to identify folate enriched tomato accessions that can be either readily used for commercial purpose or for breeding and study purpose. We identified few high and low folate lines that were taken forward for characterization. Our reference cultivar Arka Vikas also showed low folate content, in the range of other low folate accessions. Characterization of a high folate accession can offer better understanding of mechanisms/factors causing high accumulation and retention of folate in red ripe fruits. These accessions can also be used to introduce this character in local cultivars appropriate for a particular environment through introgression. The characterization results of selected accessions are discussed in the next chapter.

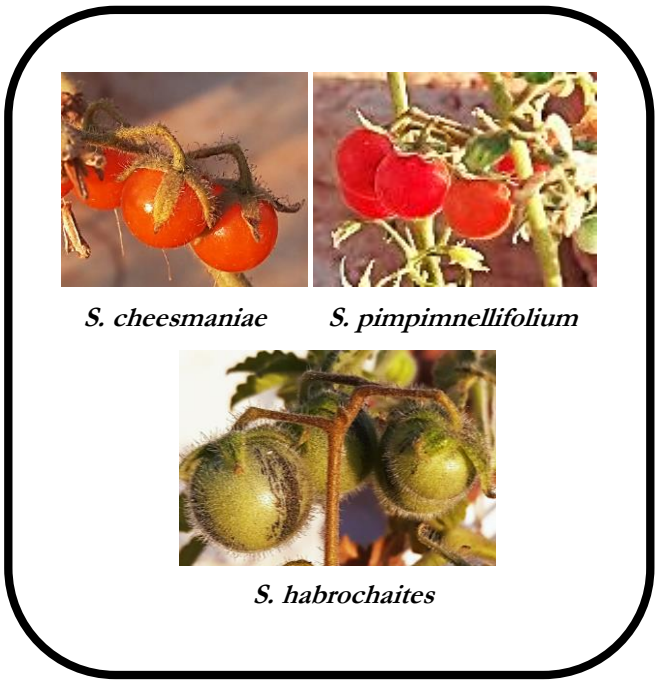


Figure 4.11. Tomato wild relative fruits in their ripe stage (39 DPA=days post anthesis). The accessions used are LA0483 (*S. cheesmaniae*), LA1589 (*S. pimpinellifolium*) and LA1777 (*S. habrochaites*).

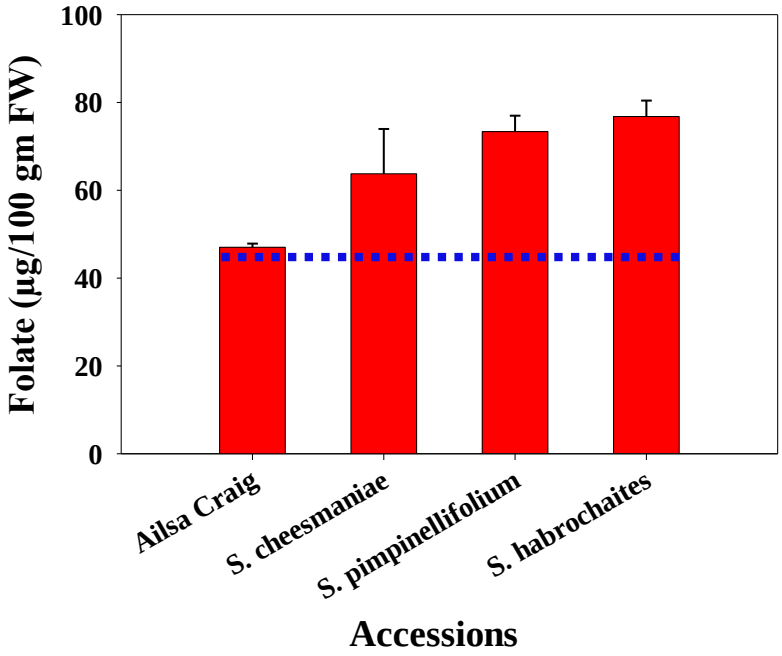


Figure 4.12. Total folate content in ripe stage of three wild species of tomato as compared to S. lycopersicum cv. Ailsa Craig. All the fruits were collected from open field grown plants at 39 days after anthesis.

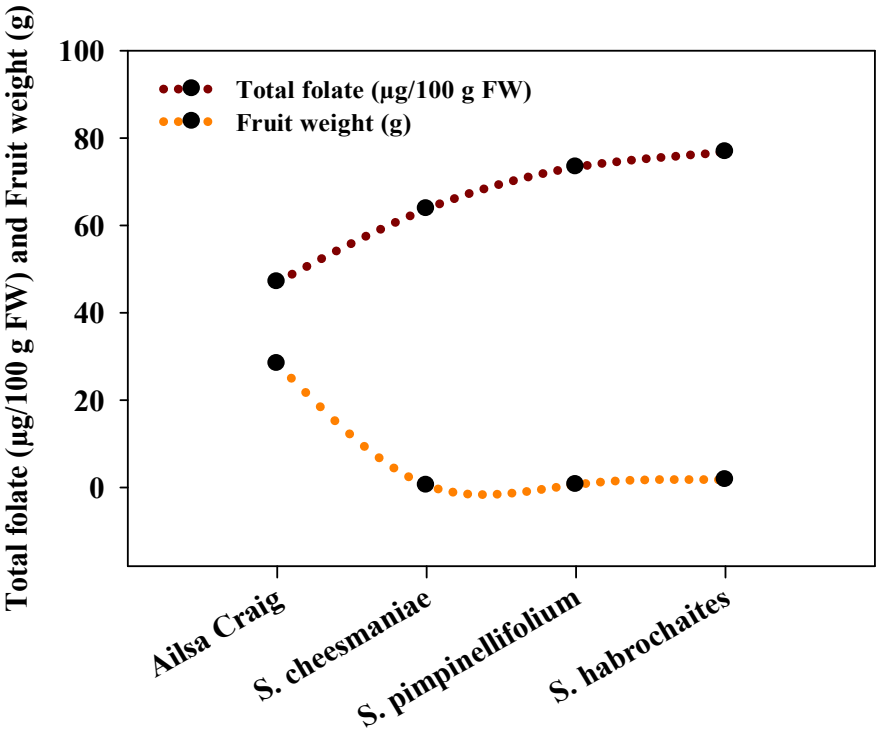


Figure 4.13. Correlation between total folate content (µg/100g FW) and fruit weight (g) in RR stage.

Table 4.7. Correlation between total folate content and fruit weight in red ripe fruit. MA, LC-MS= correlation between folate content and fruits weight of natural accessions, with MA and LC-MS method of folate estimation respectively. Wild relatives= correlation between folate content and fruits weight of wild relatives with LC-MS method of folate estimation. The negative values indicate negative correlation.

Correlation value	**MA**	Total folate (µg/100 g FW)	Fruit weight (g)
	Total folate (µg/100 g FW)	1	
	Fruit weight (g)	**-0.395**	1
Correlation value	**LC-MS**	Total folate (µg/100 g FW)	Fruit weight (g)
	Total folate (µg/100 g FW)	1	
	Fruit weight (g)	**-0.051**	1
Correlation value	**Wild relatives**	Total folate (µg/100 g FW)	Fruit weight (g)
	Total folate (µg/100 g FW)	1	
	Fruit weight (g)	**-0.896**	1

Chapter 5

Characterization of High and Low Folate Accessions)

5.1 INTRODUCTION

The main objective of screening of germplasms of any crop is finding accessions with traits that can improve the crop. The trait of interest can range from grain size, yield or level of beneficial metabolite(s), disease resistance, and drought tolerance to name a few. Our aim of screening the folate levels in tomato germplasms was to exploit the natural variations in folate content among tomato accessions and find a high folate variety, as well as a low folate variety. Identification of a high folate variety is a better approach in tackling the folate deficiency than enhancing its levels through genetic engineering. Such varieties can provide a good starting material for breeding practices and for extensive genetic and molecular assessment to understand the mechanisms for high/low folate accumulation.

A number of studies have reported folate levels among germplasms of different crops, for example tomato, wheat, legumes, lentils and strawberry (Iniesta et al., 2009; Piironen et al., 2008; Rychlik et al., 2007; Sen Gupta et al., 2013 and Strålsjo et al., 2003) yet the factors governing folate content differences among the same species are not known. Most of the studies have focused upon the effect of processing, harvest or the climate conditions in which the crop was grown. We still do not know the molecular mechanisms regulating the folate levels in different crop species.

Based on the results of folate screening (chapter 4), we selected Arka Vikas, EC27910, EC520046 and Sel.14. The folate levels in Arka Vikas and EC27910 were low as compared to EC520046 and Sel.14. The characterization involved reconfirmation of folate content in selected accessions, proportion of different folate forms, estimation of folate biosynthesis precursor molecules (pterin and pABA) and transcript abundance of biosynthesis pathway genes.

5.2 RESULTS

5.2.1 High folate accessions EC520046 and Sel.14

Two accessions, EC520046 and Sel.14 were initially selected for their high folate content. However, phenotype of EC520046 was inconsistent and showed segregation of different characters in the progeny in year 2014-15 (Figure 5.1). The four phenotypes were observed in mature green fruit stage and the plants were named as EC520046-1, 2, 3 and 4. The mature green fruit phenotype of four types varied in color, shininess of the fruit surface, green shoulder and appearance of notches on the fruits. 3 or more replicates of each type were used for folate estimation. While the screening of natural accessions grown during 2012-13 EC520046 had shown 2.8 and 2.4 fold more folate in than AV at MG and

RR stages respectively. However, the folate content of all (four) of them were varied from 18.6-30.4 in MG and 17.8-24.4 μg/100 g FW in RR stage (Figure 5.2). This segregation might have resulted due to various reasons such as- (1) mixing of seeds while harvesting or (2) cross pollination with some other accession in field. Importantly reconfirmation study did not show high levels of folate in 2014-15 growing season for which this accession was selected. Therefore, the detailed characterization of this accession was not followed. Another accession, Sel.14 showed consistent phenotype and was stable and therefore it was selected for studies pertaining to high folate accession.

Sel.14 (=selection 14 or Parbhani Yeshshree) was developed at Marathwada Agricultural University, Parbhani, India. We procured this variety from IIVR. During folate screening of 2012-13 total folate of Sel.14 was 42 μg/100 g FW in red ripe stage (MG data not available) which was 2.4 fold more than reference cultivar Arka Vikas and approx. 2.5 fold more than the low folate accession i.e. EC27910. Therefore, this accession was grown again in green house for characterization. Average fruit weight data are shown in Table 5.1 for AV, Sel.14 and EC27910. The phenotypic differences between Arka Vikas (AV) and Sel.14 were slight paleness of leaf color of Sel. than AV and mature green fruits were 3-notched but this feature was not conspicuous after ripening (Figure 5.3B and 5.5).

5.3 Total folate content and various forms of folate in leaves and fruits of Sel.14

We estimated folate content in both, leaves and fruits of tomato. Tomato leaves contain significantly higher amount of folate (Tyagi et al., 2015). Though, tomato leaves are not consumed due to high concentration of α-tomatine and solanine like glycoalkaloids (Rick et al., 1994), yet its folate content was measured to understand the mechanism governing high accumulation of folate in leaves; whereas the folate content decline in the fruits. Total folate in leaves of Arka Vikas and Sel.14 was 72 and 52 μg/100 g FW respectively. Folate content varied significantly in both MG fruits of AV and Sel.14 (29 and 52 μg/100 g FW) as well as in RR fruits (25 and 38 μg/100 g FW folate) respectively (Figure 5.6 and 5.7). While other folate forms showed similar accumulation between AV and Sel.14; 5,10-Methenyl and 5-Formyl THF showed significant variations with 3 fold higher accumulation in AV leaf than in Sel.14. The distribution of folate derivatives was different in fruit tissue. Significantly higher fruit folate content of Sel.14 as compared to AV was mainly due to 5-methyl THF which was almost 1.8 and 2.5 times more than AV in MG and RR stages respectively (Figure 5.8A and 5.9A).

5.3.1 Level of pathway precursors- pterins and pABA in Sel.14 leaf and fruit:-

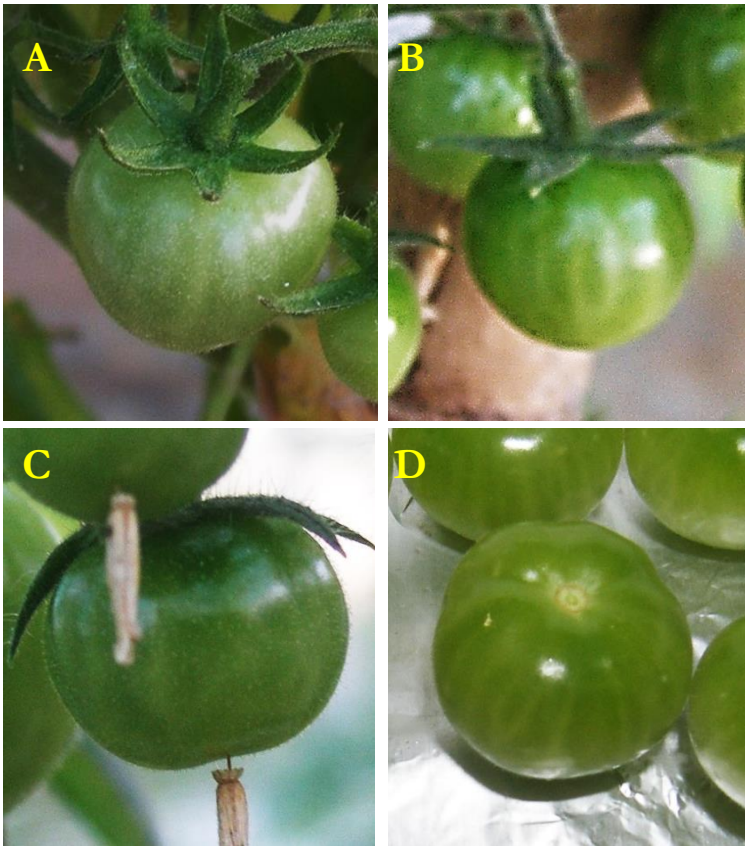


Figure 5.1. Segregating phenotypes observed for EC520046 mature green fruits. Plants were grown under green house. A= EC520046-1, light green color, B= EC520046-2, small cherry sized fruit, slightly darker, C= EC520046-3, prominent green shoulder, and D= EC520046-4, shiny fruit surface and slight notches.
Folate was extracted to establish the plant identity showing high folate character.

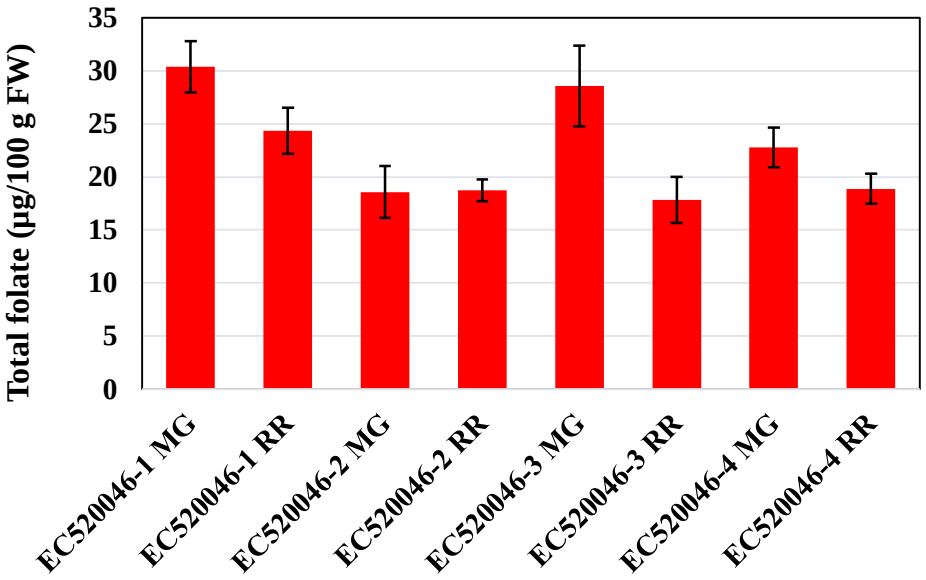


Figure 5.2. Folate level in fruits of four segregating phenotypes of EC520046

Table 5.1. Fruit diameter and weight data of Arka Vikas, Sel.14 and EC27910

Accession name	Horizontal diameter (mm)±S.E.	Vertical diameter (mm)±S.E.	Weight (g)±S.E.
Arka Vikas	31.55±1.98	40.66±1.76	31.20±4.53
Sel.14	33.63±0.97	30.88±1.55	22.94±2.47
EC27910	33.81±0.66	27.45±0.77	21.26±0.32

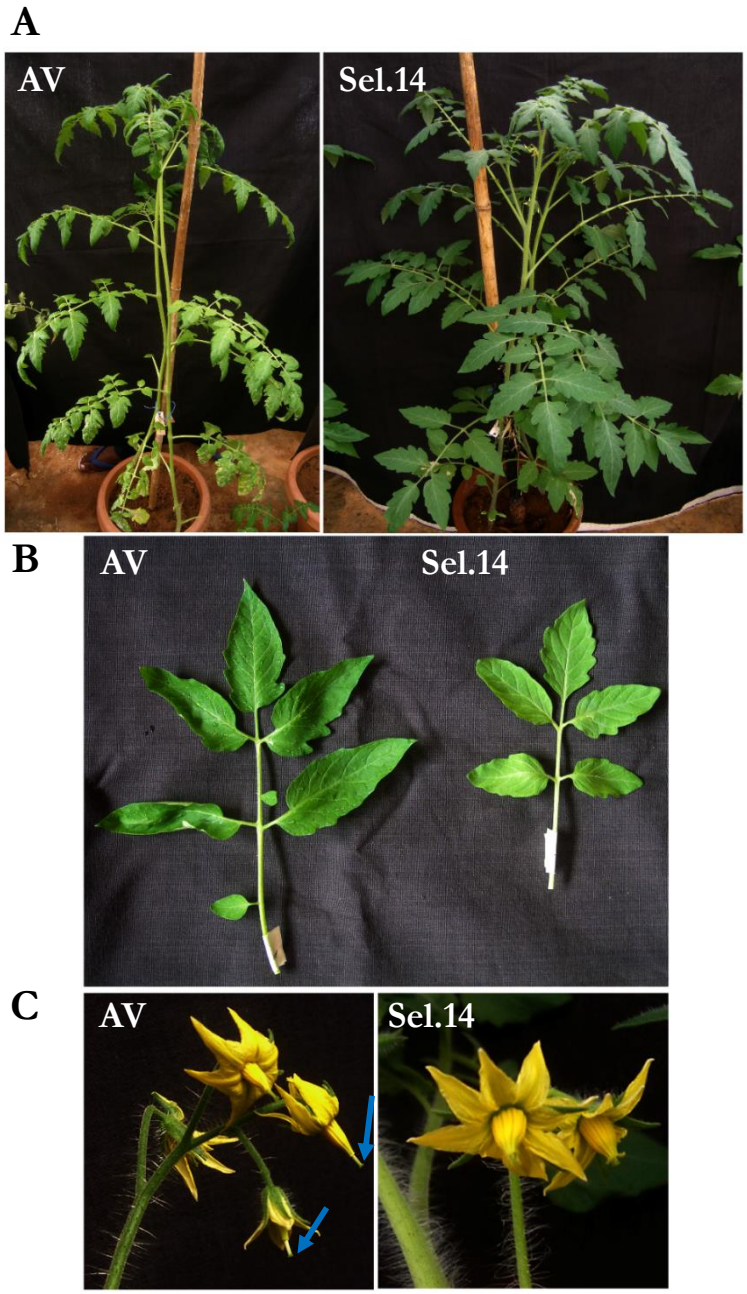


Figure 5.3. Phenotypic differences between WT and Sel.14 in (A) 2 months old plant architecture, (B) leaf of 4th node and (C) flower morphology. Plants were grown under green house. Leaves of Sel.14 were visually paler than the WT. Blue arrows indicate protruded stigma in semi-opened flower bud and fully opened flower of WT.

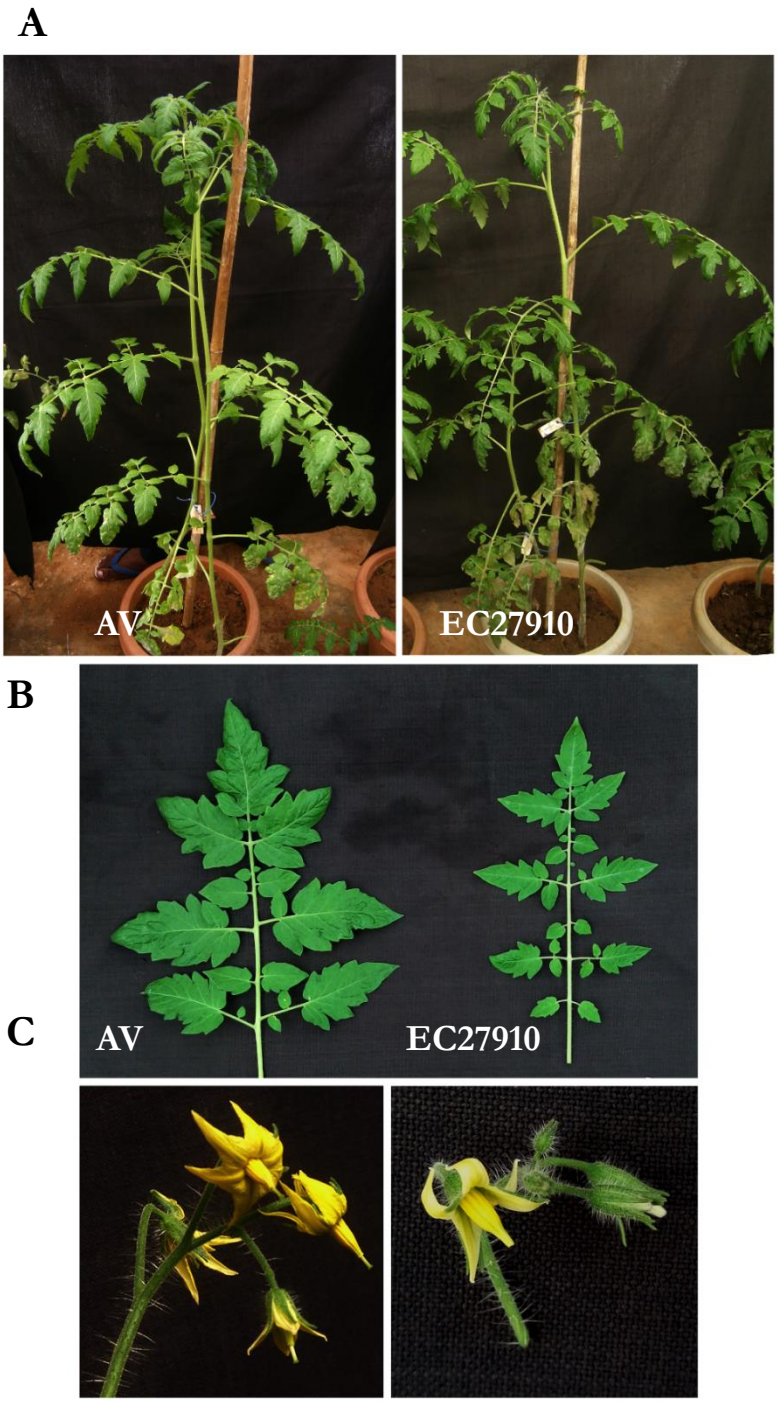


Figure 5.4. Phenotypic differences between WT and EC27910 in (A) 2 months old plant architecture, (B) leaf of 4th node and (C) flower morphology. Plants were grown in green house. Leaves of EC27910 showed smaller leaflets than the WT. Unopened flower buds in EC27910 were visually pale yellow as compared to WT.

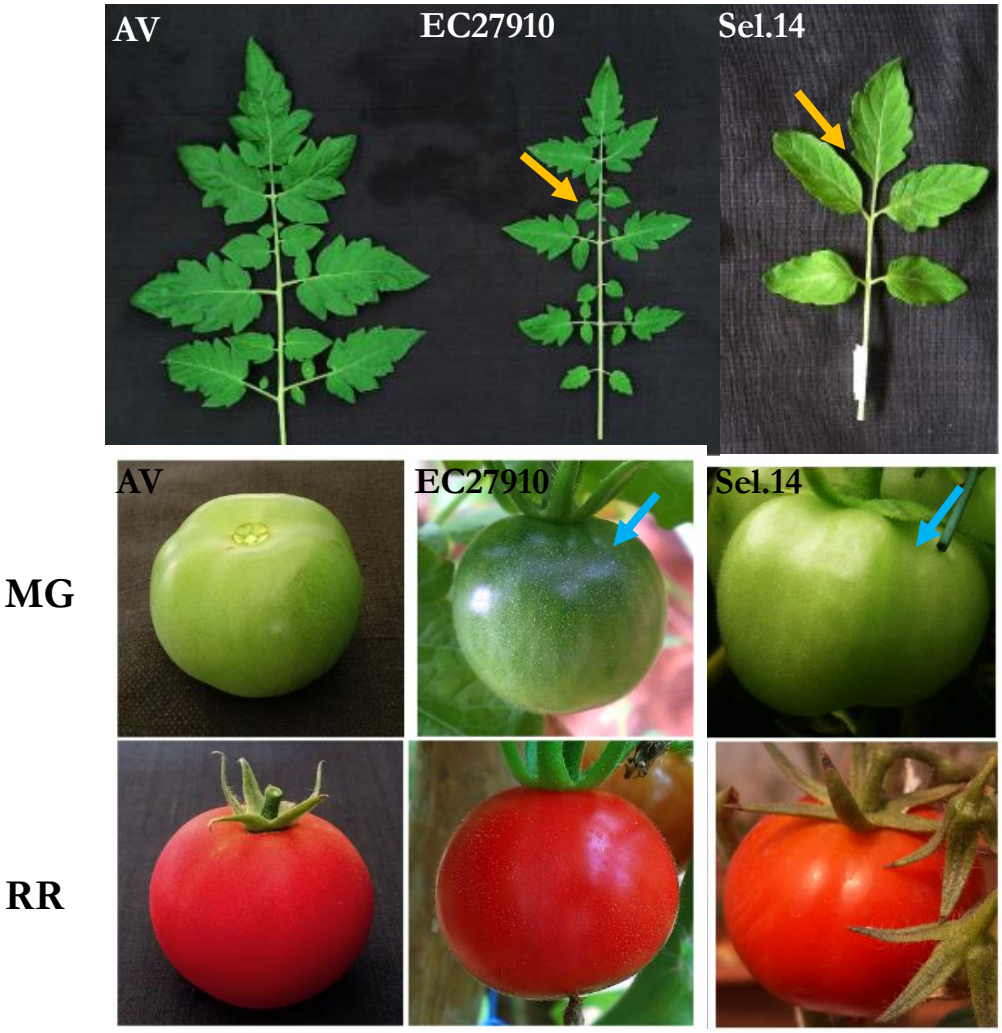


Figure 5.5. Phenotypic differences between leaf and fruit morphology in AV, EC27910 and Sel.14 in mature green (MG) and red ripe (RR) stage of fruit. **Yellow** arrows indicates smaller leaflet size in EC27910 and pale leaf color in Sel.14 compared to AV. **Blue** arrows indicates green shoulder in EC27910 and slight notch in the fruit of Sel.14 as compared to AV.

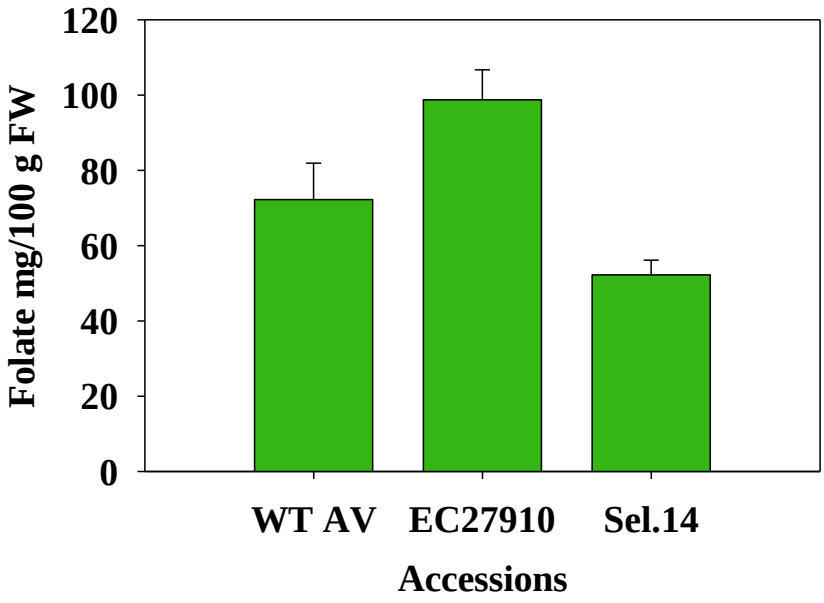


Figure 5.6. Folate content in leaf of Arka Vikas, EC27910 and Sel.14. Fully expanded leaves were harvested from 4th node of one and half month old plants. In case 4th leaf was lost to infection, 5th or 6th leaf was taken. Uniformity was maintained among all the replicates while harvesting of samples. Number of replicates n ≥ 3).

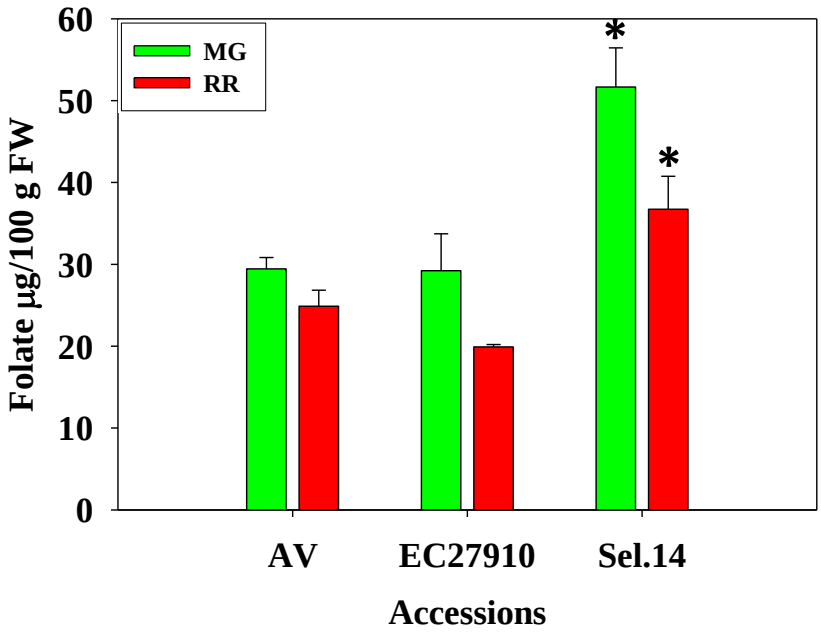


Figure 5.7. Folate content in MG and RR fruit of Arka Vikas, EC27910 and Sel.14. MG and RR fruits were harvested at 30-32±3 days after anthesis, and 8-10±3 days from MG respectively. Asterisks indicate statistically significant differences between WT and Sel.14 and WT and EC27910 of the same stage. (Student's t-test, ***** $p < 0.05$, number of replicates n ≥ 3).

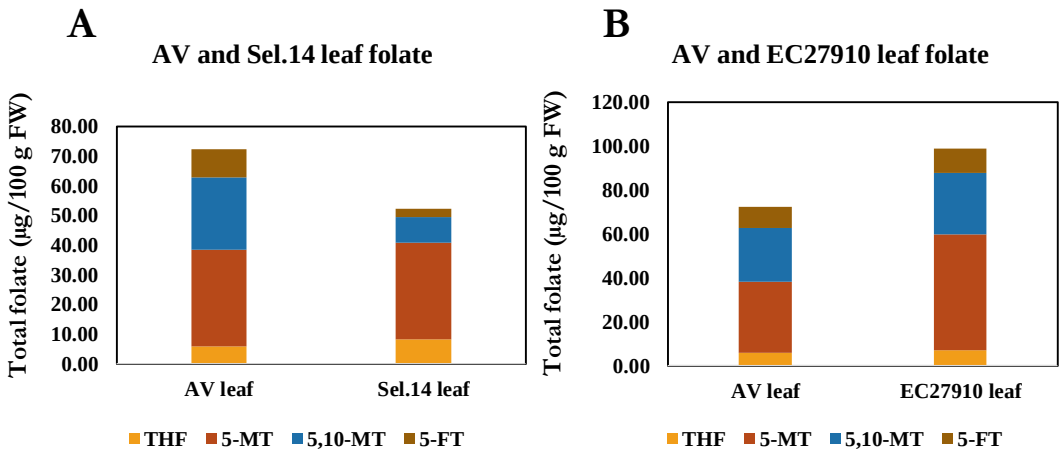


Figure 5.8. Distribution of various folate forms detected in leaf of (A) Arka Vikas and Sel.14, (B) Arka Vikas and EC27910. Fully expanded leaves were harvested from 4th node of one and half month old plants. In case 4th leaf was lost to infection, 5th or 6th leaf was taken. Uniformity was maintained among all the replicates while harvesting of samples. Number of replicates n ≥ 3).

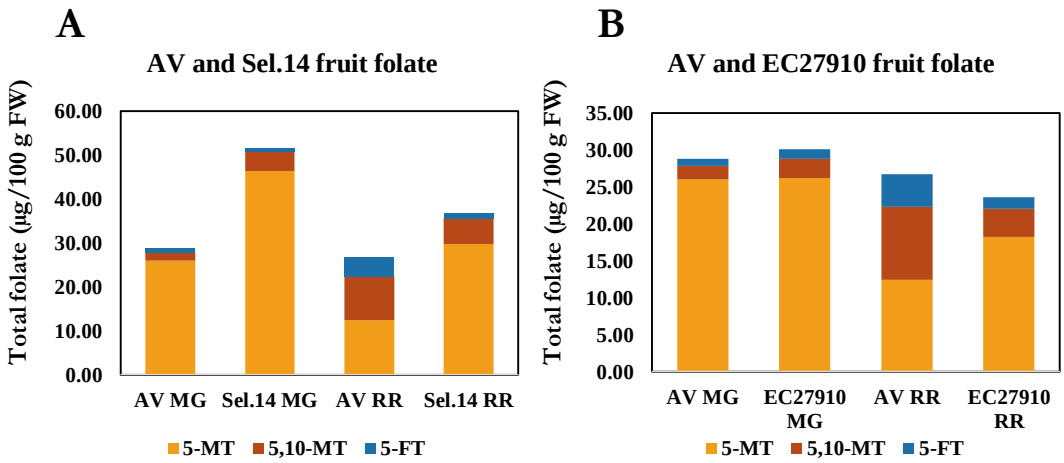


Figure 5.9. Distribution of various folate forms detected in fruit of (A) Arka Vikas and Sel.14, (B) Arka Vikas and EC27910. MG and RR fruits were harvested at 30-32±3 days after anthesis, and 8-10±3 days from MG respectively. Number of replicates n ≥ 3).

pABA levels in leaf tissue of Sel.14 was significantly less as compared to AV which correlates with the folate content in the leaves of these two accessions. Yet, both AV and Sel.14 showed similar levels of pABA in MG and RR fruit despite having disparity in total folate content. It is noteworthy that total folate content in Sel.14 was 1.8 and 1.5 times higher than AV in MG and RR stages, however the amount of precursor pABA did not directly correlate with it (Figure 5.10A, B).

Though total folate level in Sel.14 leaves was less than AV, yet level of pterin was not significantly different than AV. Similar pattern was also observed in MG fruits of Sel.14 where pterin level was not significantly different from AV. Notwithstanding these similarities the red ripe fruits of Sel.14 showed almost 13 fold higher pterin levels compared to AV (Figure 5.10A, B). It remains to be established whether a putative linkage is there between pterin and folate levels in RR fruits.

5.4 Low folate accession EC27910

EC27910 was procured from National Bureau of Plant Genetic Resources (NBPGR). Its passport data at the PGR portal database shows the variety name "Delicious" and it is a landrace from Japan. Plants were grown in green house and phenotypically characterized. The leaves and leaflets of EC27910 were smaller than that of AV. The stigma of EC27910 was not exerted as observed in AV flowers and the fruits were smaller than AV (Figure 5.4). Fruits of EC27910 showed prominent green shoulder in MG while the AV showed uniform light green color in the same stage (Figure 5.5). Table 5.1 shows a comparison of average fruit weight and diameter of AV, Sel.14 and EC27910.

5.4.1 Total folate content and various forms of folate in leaves and fruits of EC27910

Similar to the high folate accession Sel.14, the folate content in leaves and fruits of EC27910 was estimated and compared with AV. The total folate in leaves of Arka Vikas and EC27910 was 72 and 99 μg/100 g FW respectively (Figure 5.6). Folate content varied slightly in both MG of AV and EC27910 (29 and 30 μg/100 g FW) as well as in RR fruits (25, 24 μg/100 g FW folate) respectively (Figure 5.7). These results are in conformity with the folate levels in this line observed during large scale screening where the EC27910 and AV had nearly similar folate levels. In addition to difference in the total folate levels the distribution of different folate vitamers were also different in these tissues The variation between AV and EC27910 leaf folate content was mainly by 5-methyl THF which was 1.6 times more in EC27910 leaf than in AV while the levels of other folate vitamers was nearly

identical. The level of all three folate vitamers in the RR fruits of EC27910 were different than RR fruits Figure 5.8B and 5.9B).

5.4.2 Level of pterins and pABA in EC27910 leaves and fruit:- 'pathway precursors'

The level of pABA in AV and EC27910 was not different in leaf as well as MG fruit but showed significant reduction in EC27910 in RR fruit as compared to AV (Figure 5.10). Though MG fruits of EC27910 showed more pABA it was not statistically significant.

The leaves of EC27910 showed higher folate level than AV (not significant though), yet total pterin level in leaf was similar to AV. Pterin content in MG fruits was significantly higher (5 fold) in EC27910 as compared to AV but it was similar to it in RR fruit (Figure 5.10).

5.5 Comparative transcript profiling of folate biosynthesis pathway genes among AV, EC27910 and Sel.14 in leaf, MG and RR tissue

The leaf folate content in high and low folate accessions (with respect to fruit folate) showed exact reverse trend when compared with the AV leaf folate content with EC27910 showing highest (98.8±7.9 μg/100 g FW) and Sel.14 showing lowest (52.2±3.9 μg/100 g FW) leaf folate content with AV leaf content at 72.2±9.7 μg/100 g FW. Though, these differences were not statistically significant, yet we checked the folate pathway gene expression in leaf and MG and RR fruit in all the three accessions (Figure 5.11 and 5.12).

5.5.1 Overall expression pattern among accessions

The transcript abundance in leaf among AV, EC27910 and Sel.14 was significantly different only for *GGH2*, *ADCL1* and *ADCL2* genes. While *GGH2* expression was higher than AV in both EC27910 and Sel.14, *ADCL1* was higher only in EC279210. *ADCL2* expression was highest in Sel.14. *ADCL2*, which shows 20-100 fold less expression than *ADCL1* (Waller et al., 2010) followed similar trend in AV and EC27910 but its expression was substantial in Sel.14. Transcripts of all the other genes of pathway did not show statistically significant differences among the accessions (Figure 5.11).

In MG fruit, only *GCHI* expression was significantly higher in EC27910 compared to AV which correlated well with the pterin content in EC27910 MG fruit (Figure 5.12). In RR fruit, however, significantly lower expression was observed for *FPGSp* (plastidial

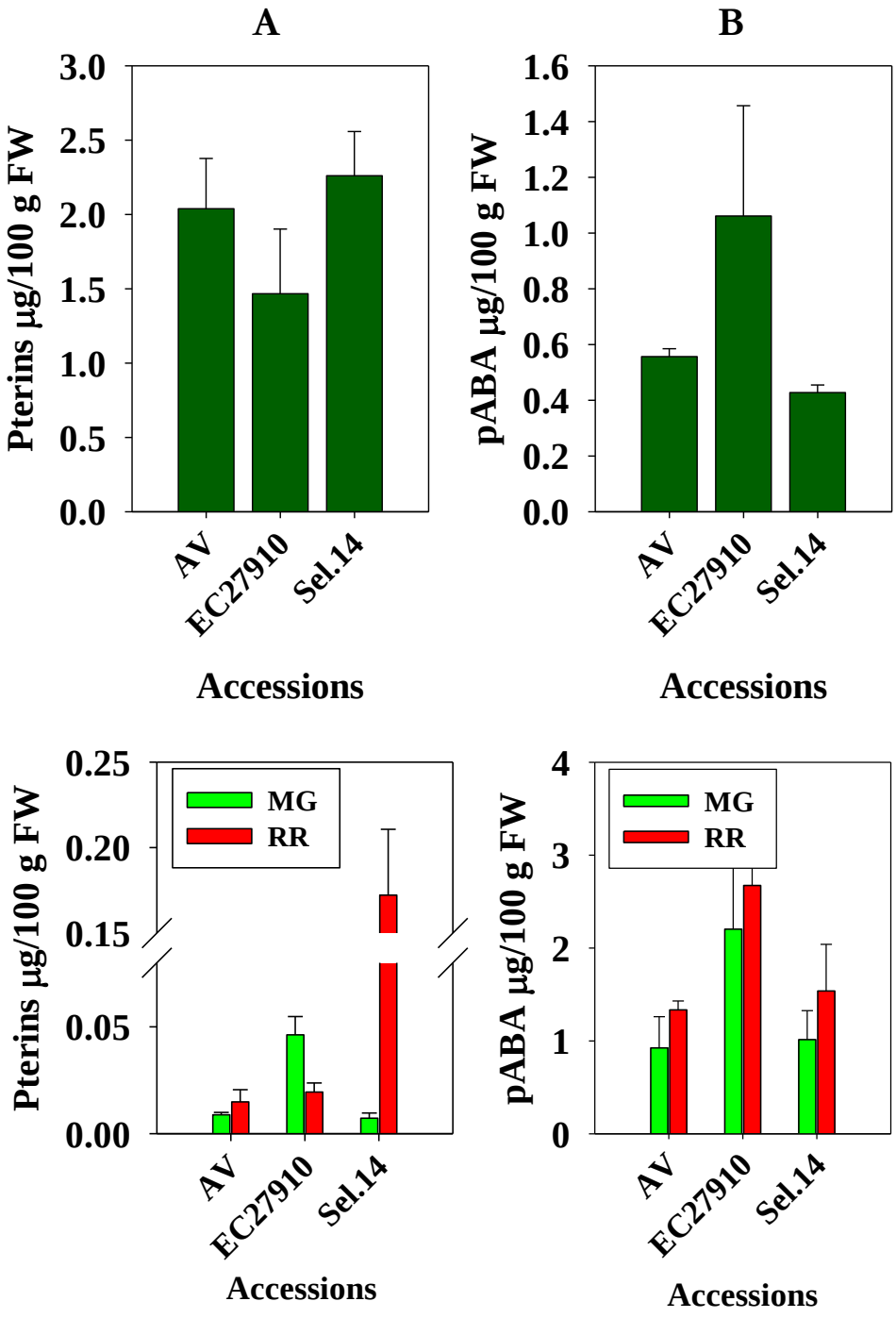


Figure 5.10. Estimation of folate intermediates, total pterin (A) and total pABA (B) in leaf and fruit (MG, RR) tissue of Arka Vikas, EC27910 and Sel.14 using LC-MS. Pterins are sum of 6-hydoxymethylpterin, carboxypterin, neopterin and monapterin. pABA values represent free pABA and pABA glucose ester. Asterisks indicate statistically significant differences between AV, EC27910, Sel.14 of the same stage. (Student's t-test, * $p < 0.05$, number of replicates $n \geq 3$).

form) in Sel.14 while the *ADCL2* expression was highest in the same accession (Figure 5.12).

5.5.2 Expression pattern between MG and RR stage

The gene expression did not vary significantly between MG and RR. Also, no sharp decline in gene expression was observed with the progression of ripening as reported previously by other groups (Basset et al., 2002; Basset et al., 2004; Waller et al., 2010). The significant decline was observed in *DHFS* expression from MG to RR progression in Sel.14 only. The expression of *DHFR* increased with ripening for both EC27910 and Sel.14.

Interestingly expression of *ADCL2* was consistently highest in leaf, MG and RR tissue of Sel.14, which is a high folate accession. *ADCL2* is not considered to be physiologically important due to its extremely low expression profiles (as compared to *ADCL1*) and absence of knowledge about its targeting locus (Waller et al., 2010). However, its expression was 2-5 fold lower than *ADCL1* in Sel.14 (Table 5.2).

5.6 DISCUSSION

The total folate content in leaf was lowest in Sel.14 while it was highest in EC27910 (not significant though). These levels were opposite to the pattern observed for the folate levels in the fruits of the respective varieties. Differences in folate content in leaves and fruits of these three may arise due to differential turnover of various oxidation states of THF like 10-formyl-THF (most oxidized), 5,10-methenyl, 5,10-methylene and 5-methyl-THF (most reduced) and their requirement by the cell. These folate forms are interconvertible and are generated during the serine, glycine, methionine and formate metabolism (Hanson and Roje, 2001). These interconversions can also arise because of intracellular activities as well as by modifications during the extraction and analysis of folates. However, considering that the extraction procedure remains the same for all the samples, it can be presumed that abundance of a particular THF form is indicative of a distinctive metabolic state of that particular sample. Likewise, this perhaps explains the differential accumulation of folate derivatives among the high, low and reference accessions; and in various tissues analyzed (leaf, MG and RR fruit).

In present study we only estimated the levels of oxidized pterins and total pABA, nonetheless, a full pathway intermediate profiling would be more appropriate. Our results show that levels of pterins and pABA were not directly proportional to the final folate content. This mismatch can be ascribed to the import of pathway precursors from their

respective site of synthesis (plastid and cytosol) to mitochondria and the activity of the mitochondrial enzyme (HPPK/DHPS) coupling the pterin and pABA to form dihydropteroate (DHP). Since the phosphorylation activity of HPPK/DHPS is sensitive to the oxidation state of pterins; for example hydroxymethylpterin, the oxidized form of hydromethyldihydropterin cannot be phosphorylated by HPPK (Díaz de la Garza et al., 2007).

Gene expression patterns in leaf of EC27910 which showed highest folate levels among the three accessions, corroborated well with its folate content. Except *dihydropteroate synthase (DHPS*), *aminodeoxychorismate lyase2* (*ACDL2*) and *γ-glutamyl hydrolase1 (GGH1*), expression of all other biosynthesis pathway genes peaked in the leaves of this accession in comparison to AV and Sel.14, though it was not statistically significant. Since *DHPS* can be inhibited by either DHP, DHF or THF concentration (Mouillon et al., 2002), higher expression of *DHPS* in low folate accession EC27910 is plausible, but same expression levels were not observed in AV where folate levels were almost similar to that of EC27910.

The expression patterns of folate biosynthesis and metabolism genes in fruits were distinctly variable (MG and RR) and did not show a fixed pattern. Contrary to the previous report (Waller et al., 2010) the expression of *aminodeoxychorismate synthase (ADCS*) did not decline with the ripening. Yet, similarity was observed in the absolute level of transcript abundance. Similar to Waller et al. (2010) in this study too *dihydrofolate reductase (DHFR*) was the maximally expressed gene. Another genes for which high transcript accumulation was observed was *FPGSp*. The pattern of *FPGSp* expression was similar to the earlier report of Waller et al. (2010). The probable reason for higher expression of *DHFR* was attributed to increased turnover of DHF caused by the high demand of thymidylate in a ripening fruit. Though *ACDL2* showed lowest absolute transcript abundance, its level was highest in the leaf, MG and RR in Sel.14, the high fruit folate accession. Not only was *ADCL2* expression was high in Sel.14, its absolute level was very near to that of *ADCL1* (Table 5.2). While the ratio of *ADCL1*/*ACDL2* expression in leaf and fruit of AV and EC27910 varied from 15 to 87, it only varied from 2 to 5 in Sel.14. Considering above accumulation of *ADCL2* in higher folate lines the presumption that *ACDL2* is physiologically insignificant may not be completely true and it might play some hitherto unreported role which needs further investigation. The expression of *FPGSp* in Sel.14 was in accordance with previous reports and its level showed negative relation with the high concentrations of folate substrates (Díaz de la Garza et al., 2007; Storozhenko et al., 2007; Hanson and Gregory III, 2011).

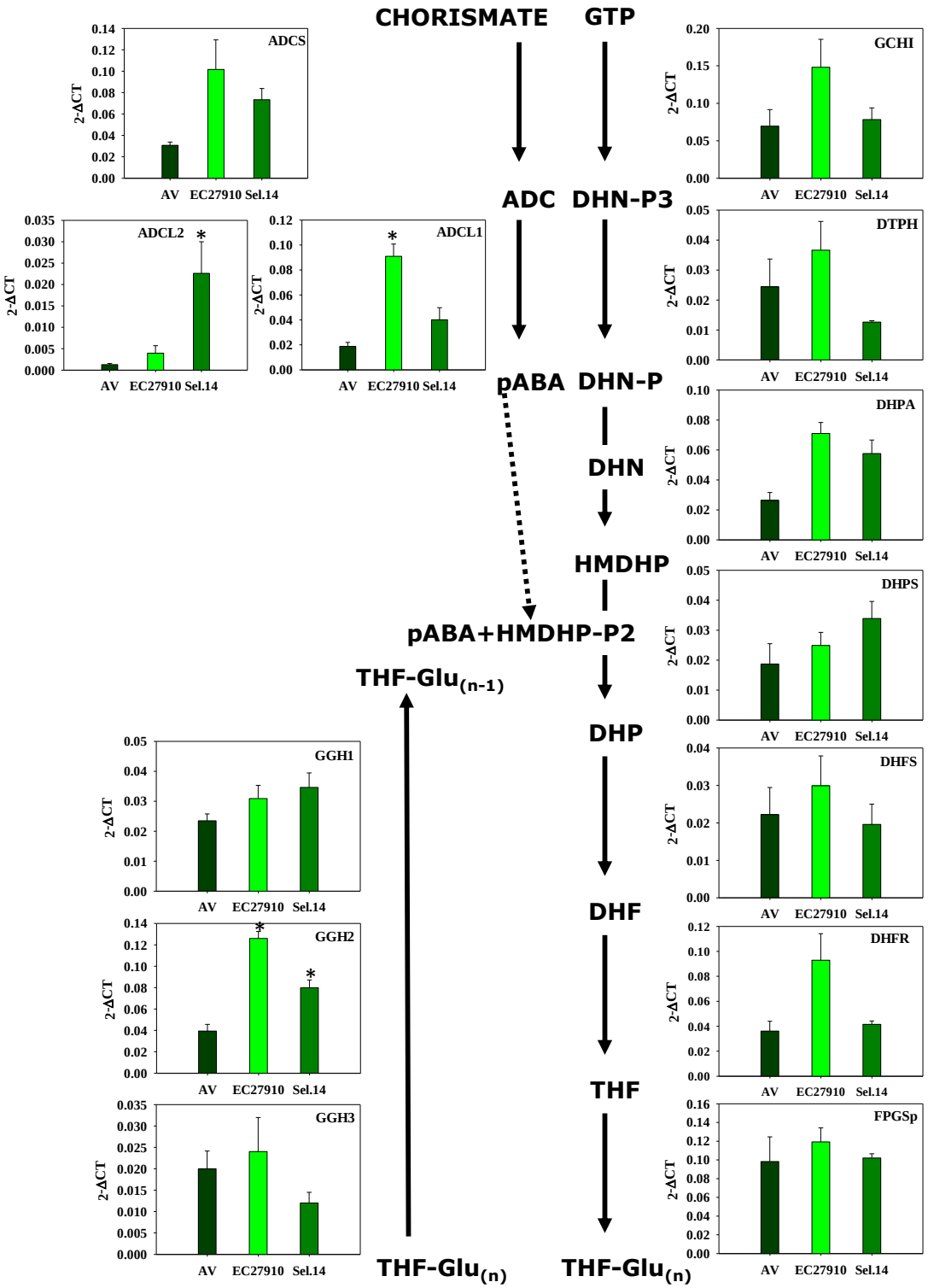

CHORISMATE
GTP
ADC
DHN-P3
pABA
DHN-P
DHN
HMDHP
pABA+HMDHP-P2
THF-Glu(n-1)
DHP
DHF
THF
THF-Glu(n)
THF-Glu(n)
ADCS
ADCL2
ADCL1
GCHI
DTPH
DHPA
DHPS
DHFS
DHFR
FPGSp
GGH1
GGH2
GGH3
2-ΔCT
AV
EC27910
Sel.14

Figure 5.11. Relative levels of transcripts involved in folate biosynthesis, polyglutamylation and deglutamylation in leaf of WT, EC27910 and Sel.14. The relative expression of transcript was plotted after normalization with internal controls (Actin and ubiquitin) using $2^{-\Delta cT}$ method. Asterisk represents significant difference from WT ($p < 0.05$). Number of replicates $n \geq 3$).
Abbreviations (enzymes):- ADCS= aminodeoxychorismate synthase, ADCL1, 2= aminodeoxychorismate lyases1 and 2, GCHI= GTP cyclohydrolase I, DTPH= dihydroneopterin (DHN) triphosphate diphosphatase, DHPA= dihydroneopterin aldolase, DHPS= dihydropteroate synthase, DHFS= dihydrofolate synthase, DHFR= dihydrofolate reductase, FPGSp= folylpolyglutamate synthase (plastidial isoform), GGH1, 2, 3= γ glutamyl hydrolase 1, 2 and 3.
Abbreviations (substrates/products):- ADC= aminodeoxychorismate, pABA= para aminobenzoic acid, DHN-P3= dihydroneopterin triphosphate, DHN-P= dihydroneopterin monophosphate, HMDHP= 6-hydroxymethyldihydropterin, DHP= dihydropteroate, DHF= dihydrofolate, THF= tetrahydrofolate, THF-$Glu_{(n)}$= tetrahydrofolate with *n* glutamate residues.

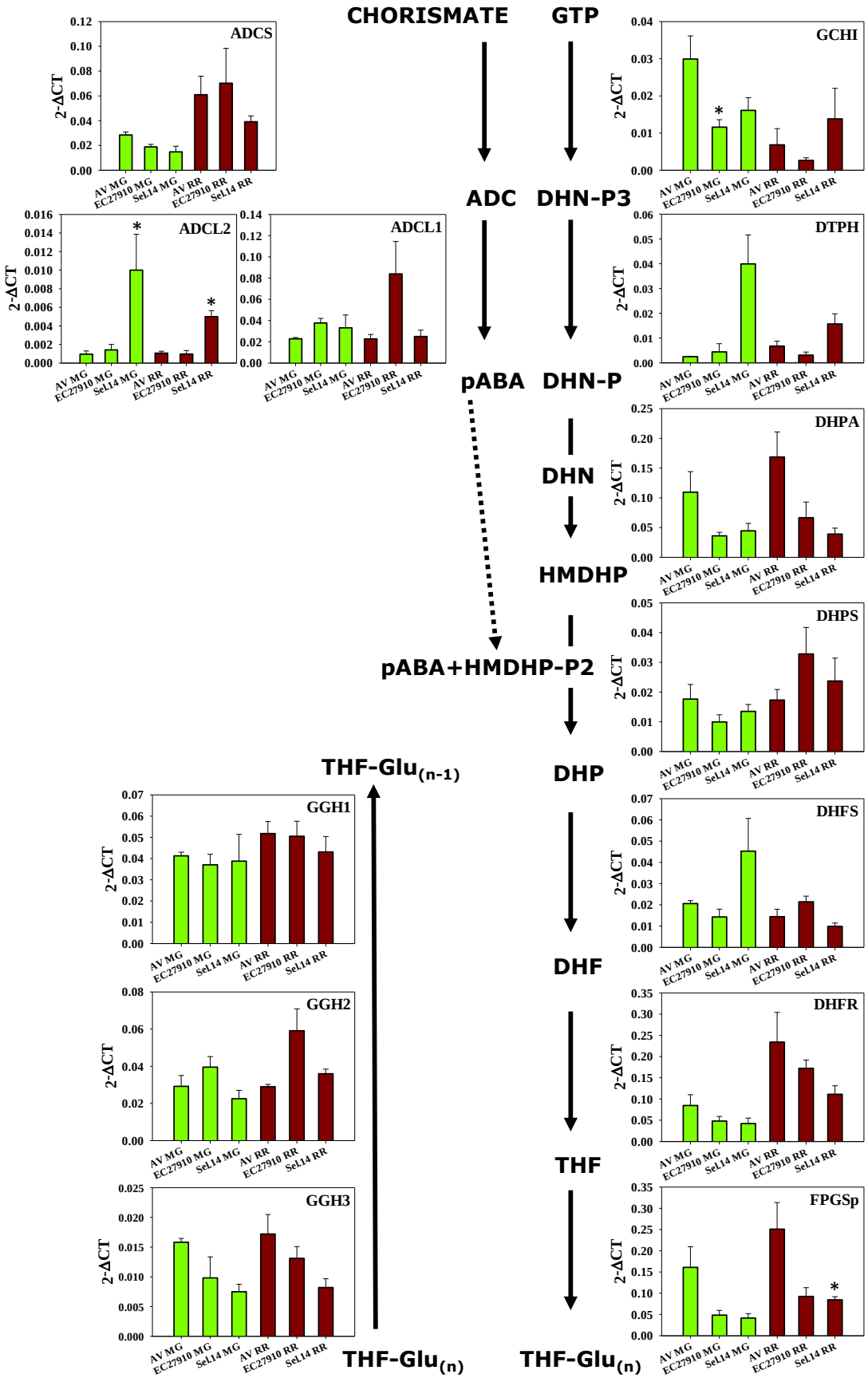

CHORISMATE
GTP
ADC
DHN-P3
pABA
DHN-P
DHN
HMDHP
pABA+HMDHP-P2
DHP
DHF
THF
THF-Glu(n-1)
THF-Glu(n)
ADCS
ADCL2
ADCL1
GCHI
DTPH
DHPA
DHPS
DHFS
DHFR
FPGSp
GGH1
GGH2
GGH3
2-ΔCT
AV MG
EC27910 MG
Sel14 MG
AV RR
EC27910 RR
Sel14 RR

Figure 5.12. Relative levels of transcripts involved in folate biosynthesis, polyglutamylation and deglutamylation in MG and RR fruit of WT, EC27910 and Sel.14. The relative expression of transcript was plotted after normalization with internal controls (Actin and ubiquitin) using $2^{-\Delta cT}$ method. Asterisk represents significant difference from WT ($p < 0.05$) of the same stage. Number of replicates $n \geq 3$).

Abbreviations (enzymes):- ADCS= aminodeoxychorismate synthase, ADCL1, 2= aminodeoxychorismate lyases1 and 2, GCHI= GTP cyclohydrolase I, DTPH= dihydroneopterin (DHN) triphosphate diphosphatase, DHPA= dihydroneopterin aldolase, DHPS= dihydropteroate synthase, DHFS= dihydrofolate synthase, DHFR= dihydrofolate reductase, FPGSp= folylpolyglutamate synthase (plastidial isoform), GGH1, 2, 3= γ glutamyl hydrolase 1, 2 and 3.

Abbreviations (substrates/products):- ADC= aminodeoxychorismate, pABA= para aminobenzoic acid, DHN-P3= dihydroneopterin triphosphate, DHN-P= dihydroneopterin monophosphate, HMDHP= 6-hydroxymethyldihydropterin, DHP= dihydropteroate, DHF= dihydrofolate, THF= tetrahydrofolate, THF-$Glu_{(n)}$= tetrahydrofolate with *n* glutamate residues.

Table 5.2. The differences in *ADCL2* expression among the accessions with respect to *ADCL*1 expression.

Tissue type	**Fold expression (*ADCL1/ADCL2*)**		
	Arka Vikas	EC27910	Sel.14
Leaf	15	23	2
Mature green fruit	24	27	3
Red ripe fruit	21	87	5

The absence of a synchronized control and direct relationship between folate level, pathway precursors and the expression of genes coding for them which has been reported by others as well (Waller et al., 2010) indicates towards more regulatory points other than the transcript or precursor accumulation (Blancquaert et al., 2013). This presumption was substantiated when the same two gene (*GCHI and ADCS*) overexpression strategy yielding massive folate enhancement in rice and tomato was not as successful in case of potato and Arabidopsis. The constraints regulating folate levels are yet to be discovered which cause or prevent folate accumulation. In particular the knowledge about the folate biosynthesis enzymes is extremely limited. It is well known that very often the transcript profiles do not reflect the protein levels and enzymatic reactions are governed by their own set of regulation such as feedback reactions, post-translational modification of proteins, transport of the metabolites etc.

Morandini (2013) emphasized on identifying the factors which could properly channelize the precursor metabolites thus creating a pull of demand for those metabolites instead of solely tinkering with rate limiting enzymes (overexpressing them or making free from natural feedback controls) which is very frequently not helpful and often results in dangerously higher levels of intermediate metabolites. This was observed in initial transgenic studies where *GCHI* expression lead to very high accumulation of pterin but only moderate increase in folate levels. Another aspect of flux which involves transporters is also not very well explored in case of folate metabolism. While the mammalian system is known to have 7 different types of transporters, only 3 have been characterized in plants so far. The genes identified based on homology searches could not be functionally validated as they did not transport folate or pterin molecules.

The coordinated regulation of folate biosynthesis genes is still not very well understood and it is believed that different mechanisms exist at both intra and inter species level. Considering above limitations and gaps in the existing studies it is evident that further characterization of these high and low folate lines involving quantification of pathway intermediates and total primary metabolites, flux monitoring and enzyme activity tests; individually as well in their F1 offspring may help in better understanding of the process of differential accumulation of folate in these lines.

Parallel to the characterization of high and low folate accessions, we also studied wild tomato introgression lines with respect to folate regulation. The usefulness of introgression lines in mapping of complex traits governed by QTLs rather than single gene

is well established for BRIX, vitamin C and vitamin E. Results of searching folate QTLs is described in the next chapter.

Chapter 6

Searching Folate regulatory bins using S. lycopersicum X S. pennellii introgression lines

6.1 INTRODUCTION

S. pennellii introgression lines (ILs) have been used for deciphering various complex traits in tomato which are governed by many factors instead of a single factor. In present study, large number of tomato natural accessions were screened for their folate content exhibiting a wide range of variation from 9-42 μg/100 g FW folate among 52 accessions. One high and one low folate accession selected among these, based on folate content were also preliminarily characterized (Sel.14 and EC27910, chapter 5). However, these studies though indicated that the regulation of folate level is complex in the nature but did not reveal the factors regulating differential folate accumulation in those accessions. The mismatch between folate levels and transcript profiles also indicated that this trait might be governed by more than one loci. For elucidation of complex traits, *S. pennellii* introgression lines (ILs) have been used in several studies as an exclusive resource.

The introgression lines harbor only a small portion of the donor parent *S. pennellii* genome and are otherwise isogenic to each other and M82 (the recurrent parent). Therefore, direct correlation of the trait variation observed in any IL compared to the M82 can be ascribed to the introgressed chromosomal segment of the donor parent. Parts of introgressed region may slightly overlap among the ILs covering a chromosome, yet there are unique non-overlapping regions which are not present in any other IL and are termed as bins. These bins are useful in further narrowing down the mapped region and ease the quantitative trait loci (QTL) identification. QTL mapping has helped in mapping of complex traits and in some cases the causal gene can also be identified.

At present there are no reports for mapping of QTLs governing folate levels. It was anticipated that screening of ILs for folate content might help in narrowing down the genomic regions regulating folate content in tomato fruits. Though we could not procure the *S. pennellii* fruits the results were compared with the recurrent parent M82. The folate data presented here are for 63 ILs and M82 with three or more replicates (Table 6.1). The position of various folate turnover genes in respective ILs has been shown in Figure 6.1.

6.2 RESULTS

6.2.2 Folate content in RR fruits of introgression lines

The folate content of 63 ILs varied from 8.2 to 25.5 μg/100 g FW folate with M82 showing 17.2 μg/100 g FW folate (Figure 6.2A). The mean and median folate content was 16.02 and 16.45 μg/100 g FW folate respectively. The predominant folate form in all the

ILs was 5-methylTHF followed by 5-formylTHF as the second most abundant form (Figure 6.2B). Folate content of 9 lines was significantly lower and 5 lines was significantly higher than M82 ($p<.05$) (Figure 6.4). The high folate trait was observed in IL 3-2, 9-1, 9-3, 10-1-1 and 12-2, while the low folate was observed in IL 4-2, 4-3-2, 7-2, 8-2, 8-3-1, 9-3-2, 10-2, 11-4-1 and 12-1.

6.2.3 Distribution of folate vitamers and high folate trait

Though, the predominant folate form among ILs was 5-methylTHF, it was not a fixed trend among the high folate ILs. The 5-methylTHF constituted 54% in the M82 and IL 12-2, it was ~60-80% in 3-2, 9-1 and 9-3. Another interesting observation was that IL 12-2 showed the same % profile of other vitamers as M82 despite being the highest folate IL. While IL 9-1 and 10-1-1 showed similar vitamer profiles, vitamer distribution of 3-2 and 9-3 was nearly same to each other (Figure 6.3).

6.2.4 Folate pathway gene expression among the selected ILs

The gene expression data of folate biosynthesis, polyglutamylation (Figure 6.5) and deglutamylation (Figure 6.6) for selected low and high ILs were retrieved from tomato functional genomics database (TFGD). The M82 RPKM values were subtracted from those of ILs and graphs were plotted. Majority of the genes showed downregulation of genes as compared to the M82. Also, many of the high and low folate ILs showed similar expression patterns.

6.3 DISCUSSION

6.3.1 Initial screening of tomato introgression lines to identify folate QTL

The folate content variations among ILs in red ripe fruits was not conclusive for mapping of a folate reducing QTL, since many ILs showed significantly lower folate than M82 representing the introgression in different chromosomes. However, careful observation of folate content of ILs of chromosome 12 (12-1, 12-1-1, 12-2 and 12-3, 12-3-1, 12-4 and 12-4-1) revealed that the increase and decrease in folate content was associated with IL12-2 and IL12-1 respectively. Though, folate content of other ILs of chromosome 12 was not significantly different from M82, yet, except 12-2, all showed a similar trend of lower folate content, which showed higher folate.

Therefore we mapped the location of folate related genes on the IL map. We found that among the 20 different genes related to folate biosynthesis and transformations, only two of the many serine hydroxymethyl transferases (SHMT); SolyC12g095930.1 and

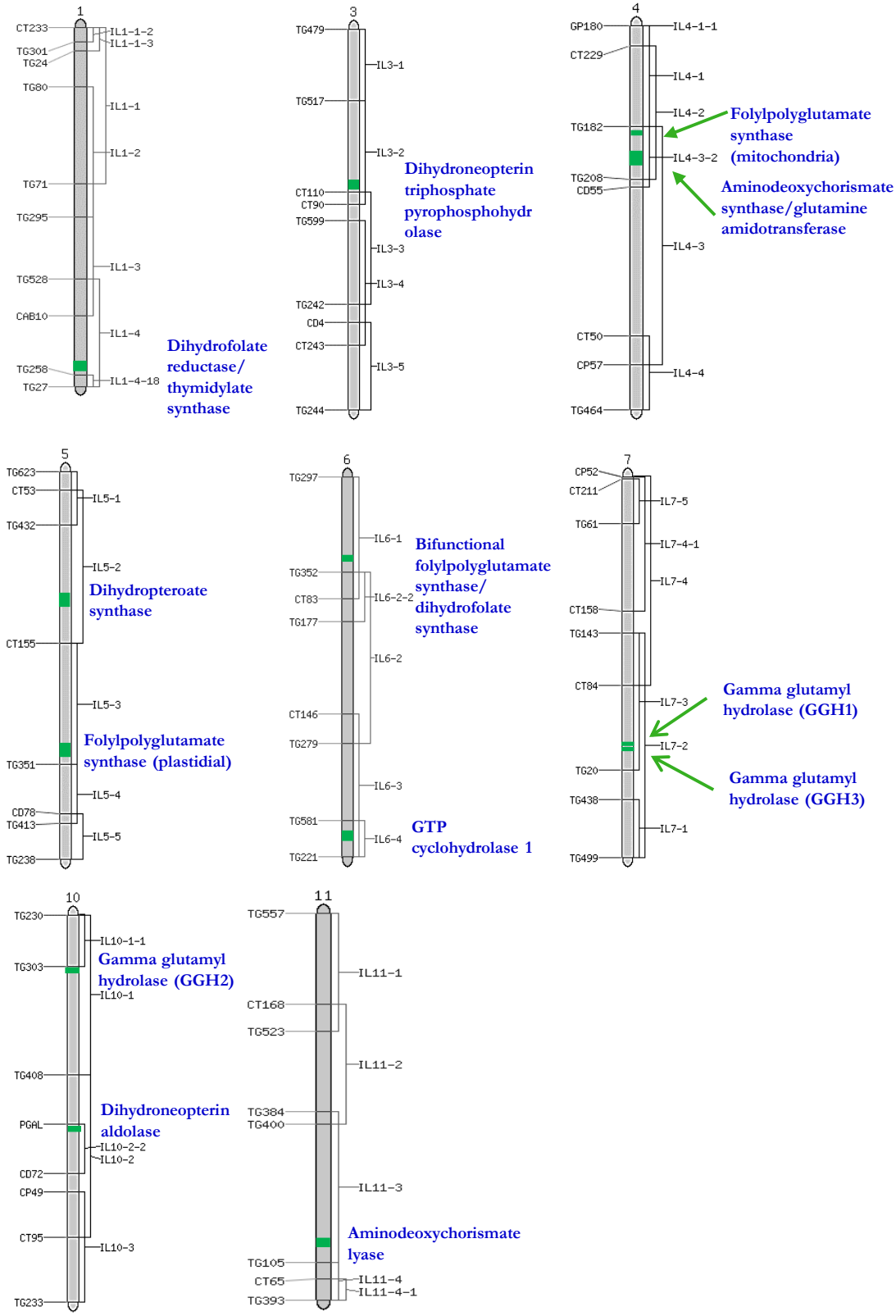


Figure 6.1. Position of folate biosynthesis genes shown on respective ILs in proximity to the nearest marker.

Table 6.1. The number of ILs used in the study with their respective replicates. The data is consolidated from 2012-13 and 2015-16 sessions of growing ILs in the green house.

Number of ILs	**Representation in fruit folate data**
63	3 or more replicates
Donor and recurrent parents	
M82	3 or more replicates
S. pennellii	No fruit

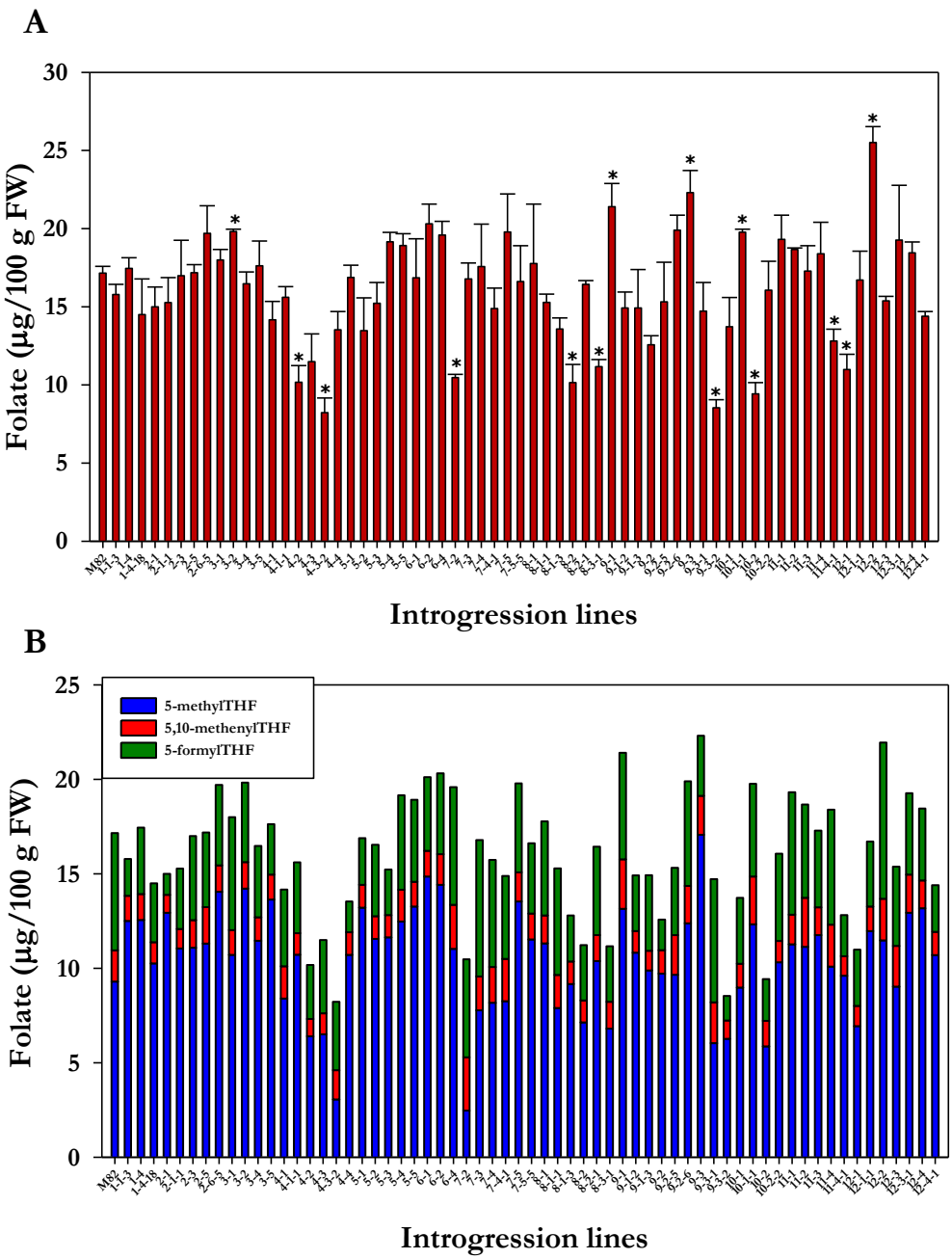


Figure 6.2. Folate content in red ripe (RR) fruits of 63 ILs as compared to M82. ILs were grown in green house. A= total folate content across the ILs and in M82. B= stack graph showing distribution of folate vitamers across ILs. Asterisk represents significantly different value as compared to M82 ($p<.05$).

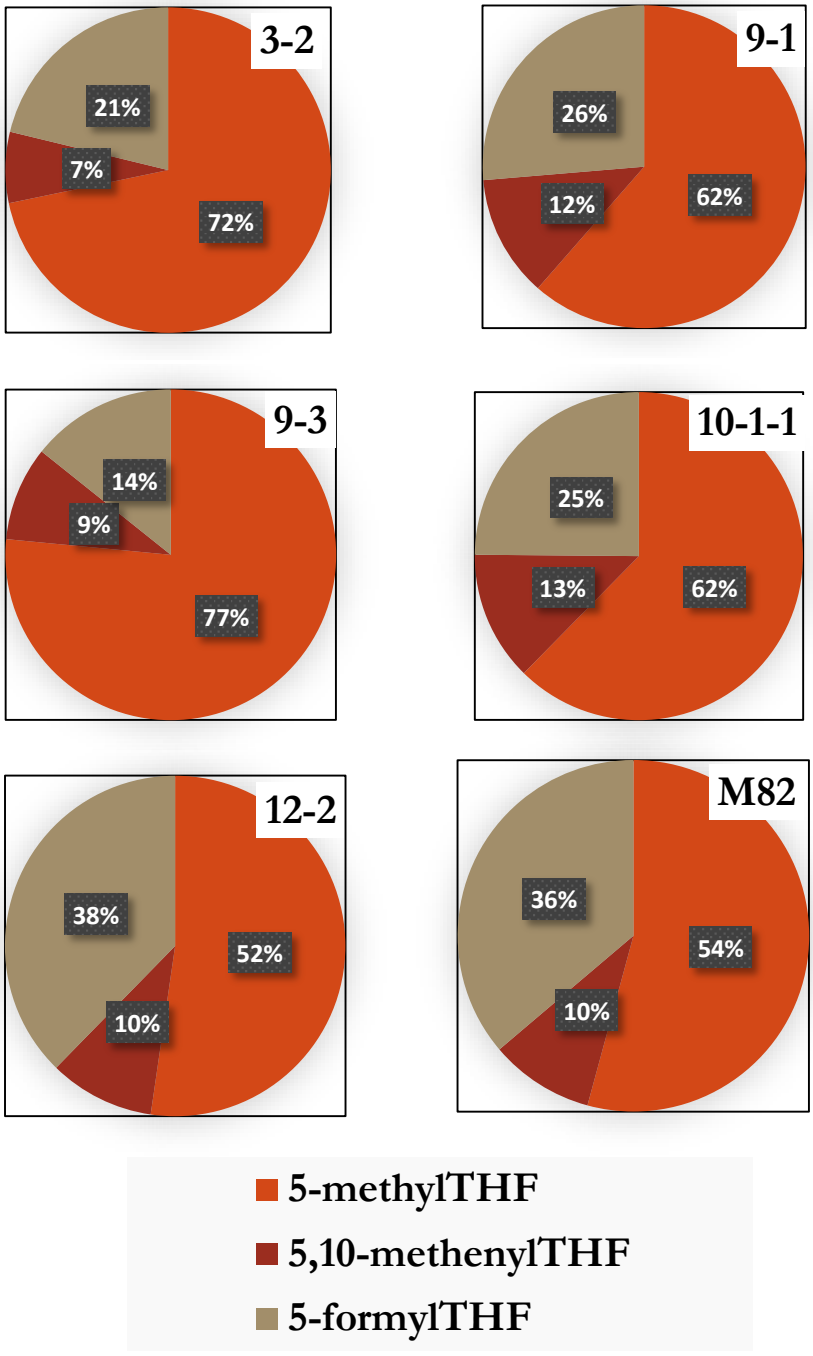


Figure 6.3. Percentage contribution of folate forms among the high folate ILs and parent accession.

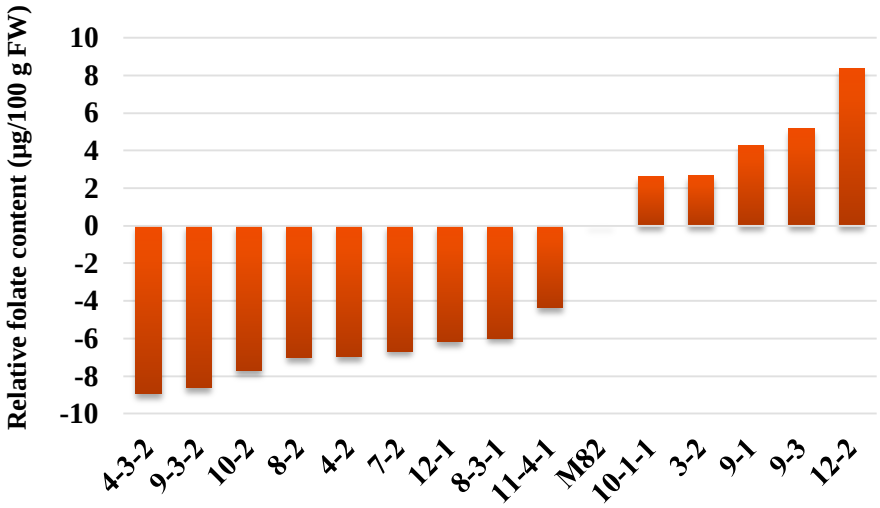


Figure 6.4. Relative folate content in red ripe fruits represented after subtracting the mean folate value of M82. The value of M82 was considered zero and lines with significantly higher and lower ($p<.05$) folate as compared to M82 were shown on + and – sides of y axis respectively. ILs were grown in green house.

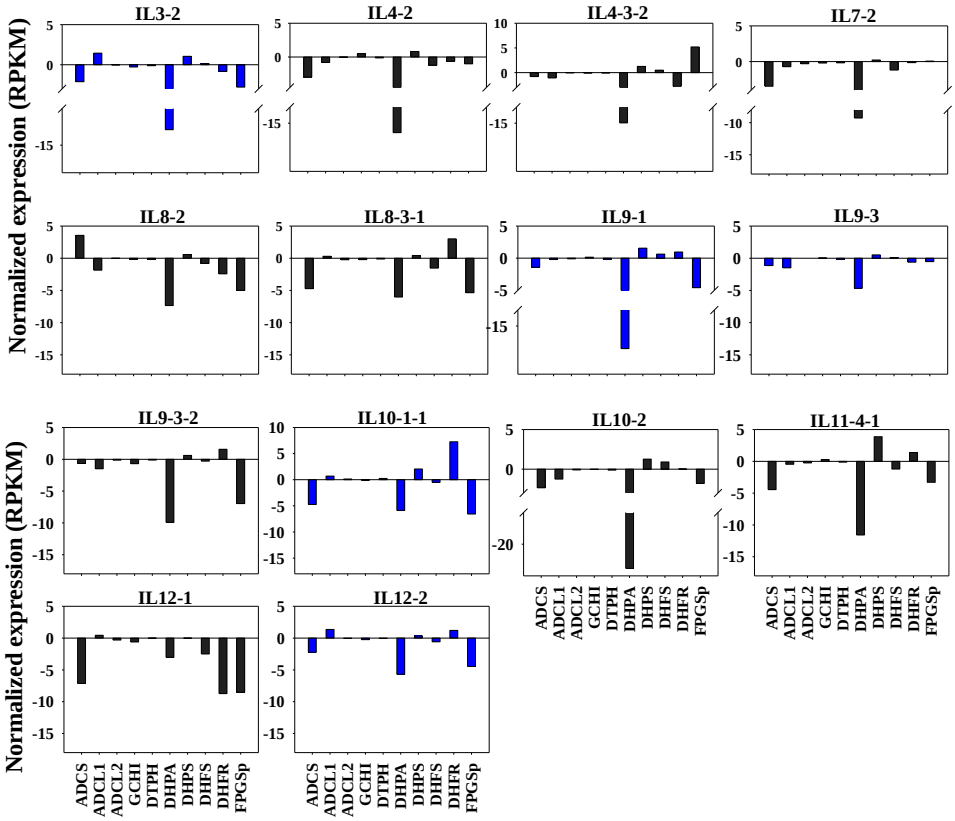


Figure 6.5. Gene expression data of folate biosynthesis and polyglutamylation genes after normalization with M82 in significantly low and high folate ILs compared to M82. The data were retrieved from tomato functional genomics database (http://ted.bti.cornell.edu/). **Blue bars** indicate expression in higher folate ILs, **black bars** indicate lower folate ILs compared to M82

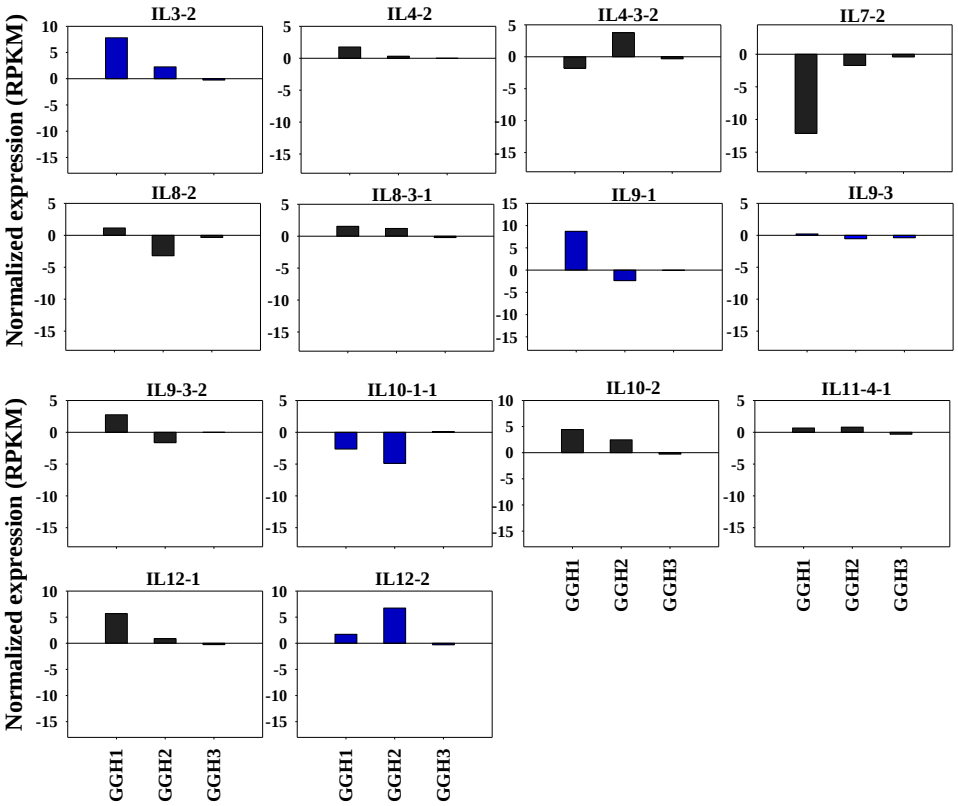


Figure 6.6. Gene expression data of folate deglutamylation genes after normalization with M82 in significantly low and high folate ILs compared to M82. The data were retrieved from tomato functional genomics database (**http://ted.bti.cornell.edu/**).
Blue bars indicate expression in higher folate ILs, **black bars** indicate lower folate ILs compared to M82

SolyC12g098490.1 were mapped to IL12-4 (the low folate IL). SolyC12g095930.1 was mapped to the overlapping bin of 12-3 and 12-4 while SolyC12g098490.1 is present in 12-4 region.

Following reactions are catalyzed by SolyC12g095930.1-

1. 5,10-methenyltetrahydrofolate + H2O → 5-formyl-tetrahydrofolate + 3 H^+
2. D-alanine + pyridoxal 5'-phosphate = pyruvate + pyridoxamine 5'-phosphate
3. DL-allothreonine = acetaldehyde + glycine
4. L-serine + tetrahydrofolate ↔ glycine + 5,10-methylene-THF + H_2O;

While, SolyC12g098490.1 catalyzed only one reaction-

1. L-serine + tetrahydrofolate ↔ glycine + 5,10-methylene-THF + H_2O.

Apart from versatility of SolyC12g095930.1 catalyzed reaction another interesting observation was dissimilarities in the expression profile of both the genes. The expression data was obtained from tomato functional genomics database (TFGD, http://ted.bti.cornell.edu/cgi-bin/TFGD/digital/experiment.cgi?ID=D003). The normalized expression (RPKM) value for SolyC12g095930.1 was found to be highest and second highest in 12-3 and 12-4 respectively. On the other hand, RPKM value of SolyC12g098490.1 in these two ILs was one of the lowest. These genes will be further explored for their role in folate metabolism.

The putative high folate QTL 12-2 could not be further narrowed down since it encompasses an extended chromosomal length, delimited by 12-1 and 12-3-1 on either ends. It is a long stretch of 25 centimorgan (cM) from 13 to 38 cM and therefore finer resolution will help in identifying the underlying genes conferring the high folate phenotype to IL12-2.

6.3.2 Distribution of folate vitamers among high folate ILs

The folate vitamer distribution was studied among the significantly high folate ILs with the aim to identify QTLs for these vitamers. It was observed that the percentage contribution of each folate forms among these ILs varied from each other. The IL12-2 showed the relative proportion of vitamers similar to M82 but the absolute values of each vitamer caused enhanced total folate in 12-2 as compared to M82. IL9-1 and 10-1-1

showed similar total folate content and vitamer profile while IL 3-2 and 9-3 showed similar percentage of only 5-methylTHF form. These results indicate that the introgressed portions are not solely responsible for modulating the folate levels and epistatic or pleiotropic interaction might be involved (Quadrana et al., 2014; Chapman et al., 2012).

6.3.3 Folate turnover gene expression pattern

The M82 normalized expression data did not establish any clear pattern between the folate content and transcript abundance. In general, dihydroneopterin aldolase (*DHPA*) expression was downregulated in both high and low folate ILs, however, maximum decline was observed in IL 10-2 where *DHPA* is mapped. Also, *FPGSp* (plastidial folylpolyglutamate synthase) was downregulated in all except IL 4-3-2. Apart from these genes, other genes showed similar up/downregulation among ILs irrespective of their folate content. The *GGH2* downregulation in 10-1-1 correlated with the high folate content of this line (Orsomando et al., 2005). In contrast, upregulation of *GGH2* was observed in IL 12-2 (high folate). Folate pathway transcripts did not show correlation with the folate content which was also observed in case of high and low folate accessions of tomato (chapter 5) characterized in present study. The data indicates that more than the directly related genes of folate biosynthesis and turnover, regulatory genes such as transcription factors are participating in it. Transcription factors, in contrast to the structural genes control multiple steps of metabolic pathway (Broun, 2004; Sagor et al., 2015) and are able to modulate metabolite levels through a synchronized control of various factors.

The above results indicate that the underlying genes in the identified folate QTLs need further investigation. Moreover, the mapping resolution has considerably increased after the creation of smaller sub-lines (Alseekh et al., 2013) and recently by the backcrossed inbred lines (BILs) (Ofner et al., 2016). With the clues obtained from our results, these resources can be used for targeting the factors modulating folate levels in tomato.

The approaches described till now involved phenotype (folate content) based screening and search for regulatory factors. In parallel, we also checked genic diversity of folate turnover genes using EcoTILLING and TILLING. The results are elaborated in the next chapter.

Chapter 7

Assessment of genic polymorphism in folate turnover genes using EcoTILLING and TILLING

7.0 INTRODUCTION

In order to feed burgeoning world population, it has been estimated that the cereal production needs to increase by 30% till 2050 (Allen et al., 2011). Along with increasing food production to meet the above demand, it is equally important to provide the population micronutrients enriched affordable foods. However, most cultivated crop varieties have lost their genetic variability due to domestication. Therefore it is imperative to create variability by mutagenesis for generating as well as studying the impact of new alleles on desired traits. Among the vegetable plants, tomato is considered a moderate source of folate. However, so far no concerted efforts have been made to use mutagenesis to improve the folate levels in any of the crop plants. Unlike yield or pigment related traits which are phenotypically scorable, forward screening for improved folate content in a large mutagenized population is not a practical option since extraction and estimation of folate is both, burdensome and expensive. In view of this constraint, in present study, an EMS mutagenized tomato cv. Arka Vikas population was used for finding out mutations in genes regulating folate metabolism by TILLING (Targeting Induced Local Lesions In Genomes).

TILLING is a reverse genetics approach that has been used for several crops including tomato to find alleles of interest in a given gene and its variation. EcoTILLING also identifies genic polymorphism in the natural germplasms. Usage of TILLING in tomato have identified several novel alleles regulating traits such as yield, shelf life, stress resistance and accumulation of beneficial nutrients such as carotenoids and sugars (Minoia et al., 2010; Menda et al., 2004; Rigola et al., 2009; Piron et al., 2010; Okabe et al., 2011). In present study, using TILLING mutations in folate biosynthesis, degradation or salvage related genes were screened in an EMS-mutagenized population of tomato.

TILLING technique was later adapted to discover polymorphisms in natural populations and termed as EcoTILLING (Comai et al., 2004). This technology is applicable to any organism even if they are heterozygous or polyploid. EcoTILLING employs the comparison of homologous loci, detects gene variants and helps in better functional annotation of a gene. Since sequencing entire population's genome for natural accessions of any single crop is expensive and not always feasible, EcoTILLING provides a great tool for deciphering the possible outcomes of gene variants and how they evolved in response to the different climatic conditions. This approach can be used effectively for detecting single nucleotide polymorphisms (SNPs), small insertions and deletions, and variations in microsatellite repeat number. TILLING and EcoTILLING though similar in

technique, address different aspects. While EcoTILLING is for finding natural SNPs, TILLING is for novel mutations that may not exist in the population.

The *FPGS* and *GGH* genes selected in this study polyglutamylate and deglutamylate folate respectively and thus affect folate storage and homeostasis. The influence of FPGS and GGH activity on total folate level has been studied by means of overexpression, T-DNA insertion or RNAi lines only (Akhtar et al., 2008; Akhtar et al., 2010; Waller et al., 2010). At present no study is available that examined the effect of mutation or polymorphism in *FPGS* and *GGH* genes on folate levels.

FPGS is involved in the addition of glutamate residue to THF and its derivative molecules. Since polyglutamates are found in various cellular compartments, it is expected that FPGS are present at many locations in the cell. In Arabidopsis, 3 isoforms of *FPGS* are known; plastidial, mitochondrial, and cytosolic while two or more are mostly present in other plants (Mehrshahi et al., 2010). In tomato, only mitochondrial (*FPGSm*) and plastidial (*FPGSp*) isoforms are reported (Waller et al., 2010). *FPGSm* (SOL ID Solyc04g016550) and *FPGSp* (SOL ID Solyc05g052920) are present on 4th and 5th chromosome of tomato genome and are 9842 and 9702 bp long. Both genes are characterized by the long intronic stretches and very short exons. The cDNA lengths for these genes are only 1369 and 1981 bp which is only 14 and 20% of the total gene length respectively.

Similarly, *GGH* gene which has three isoforms reported in tomato viz. *GGH1* (SOL ID Solyc07g062270), *GGH2* (SOL ID SolyC10g007410), and *GGH3* (SOL ID Solyc07g062280) and are presumably vacuolar (as no target peptide could be detected) (Akhtar et al., 2008). GGH are involved in the removal of glutamate residue from THF and its derivative molecules and also from pABA-Glu which results from the natural breakdown of folates after the pterin molecule is cleaved away. This is an important reaction as movement of folates usually takes place as monoglutamates and also, during the salvaging of various moieties which can then be re-used in folate biosynthesis. *GGH1*, *GGH2* and *GGH3* are present on 7th, 10th, and 7th chromosome and are 4222, 6777 and 3386 bp long respectively. Their corresponding exon coverage is 31, 22 and 40%.

In the following sections the results of screening polymorphism and mutations in *FPGS* and *GGH* using EcoTILLING and TILLING are presented.

7.1 RESULTS

7.1.1 EcoTILLING Population size and scoring of SNPs

A population of 391 tomato accessions was screened for *FPGS* and *GGH* gene and their isoforms using EcoTILLING in this study.

7.1.2 Grouping of accessions in haplotypes and confirmation of polymorphisms

In EcoTILLING, when different lanes in a LI-COR gel showed exactly the same band size/number of bands at the same position, that indicated similar position of SNPs. All the accessions with similar SNP(s) were categorized as one haplotype (Figure 7.1.1). Haplotypes (HT) were given number from 1 to as many as was the case for a particular gene. HT one always comprised of reference cultivar and all those accessions which did not have any SNP. Likewise, SNPs for *FPGSm*, *FPGSp*, *GGH1*, *GGH2* and *GGH3* were scored and categorized into haplotypes (Table 7.1.1).

Since LI-COR gels have very high resolution and even a 10 bp difference in the sizes of the fragments can be detected and therefore probable positions of the SNPs (exonic/intronic) were assessed visually. This helped in selecting for those accessions where cuts were similar and these were grouped in a single haplotype. The SNPs were reconfirmed using genomic DNA from corresponding accessions for each haplotype using EcoTILLING protocol as described in chapter 3. After successful confirmation, the same DNA was sent for sequencing to ascertain the location and nature of SNP. The sequencing results were analyzed with the help of Chromas LITE version 2.1.1. The chromatogram was examined both for homozygous and also for the heterozygous changes, the later needed more careful examination as these can be missed during base calling. The sequencing reads from Chromas LITE were then aligned with that of wild type sequence or reference sequence using MultAlin (http://multalin.toulouse.inra.fr/multalin/multalin.html) and exact location and type of SNP was noted (Figure 7.1.2).

7.1.3 Distribution and nature of SNPs among screened genes

Polymorphisms in *FPGSm* gene were grouped into 6 HTs but members of only HT3, HT5, and HT6 could be successfully sequenced. No SNP was found in EC14078 of HT6, HT3 and HT5 showed a common A2486G substitution. Maximum number of accessions (28) were present in HT3 followed by 24 accessions in HT5. All the SNPs were present in introns.

SNPs in *FPGSp* constituted 4 HTs, however only EC27910 of HT2 could be sequenced and had an intronic SNP C3513G. This SNP was shared by 14 accessions in HT2. EC27910 also showed low folate levels which were almost similar to our reference

cultivar Arka Vikas. Though, this SNP was intronic, this accession was taken for further characterization as it showed low folate levels in fruits

GGH1 and *GGH2* SNPs constituted 6 and 3 HTs respectively. Both the genes showed relatively less number of SNPs as compared to the *FPGS*. Sequencing was not successful for many representatives of *GGH1* but based on the position of bands on the gel, all the SNPs were intronic. In *GGH2*, EC362949 of HT2 showed two substitutions at G4922A and G5326A out of which G4922A was in exon 6. Since folate levels of this line were not very different from the reference cultivar and the estimations did not involve 3 or more replicates of RR fruits (due to losses to abiotic and biotic factors), this line was not advanced any further.

SNPs in *GGH3* gene were grouped into 4 HTs and the members of all HTs were sequenced successfully. HT 2, 3 and 4 had 28, 12 and 9 accessions respectively. Strikingly, EC520046 and EC1087 of HT2 and HT4 showed exactly the same substitutions. The similarity may have arisen from inaccurate analysis of LI-COR gel image where some of the bands in EC520046 were difficult to distinguish (due to hazy gel profile) and therefore these accessions were placed in different HTs. Maximum number of SNPs were present in LA4104 of HT3 where high sucrose accumulation trait has been introgressed from *S. chmielewskii*. Since *S. chmielewskii* is a wild species of tomato, a high number of polymorphism in this accession could be a result of this introgression. Members of each haplotype were sequenced and the results are presented in table 7.1.2.

7.1.4 Exonic SNPs and their possible effects on gene function

Most of the accessions showed multiple SNPs, and only few of them were in exons. Both non-synonymous (causing change in amino acid) and synonymous (not causing change in amino acid) SNPs were analyzed. Non-synonymous SNPs were analyzed with SIFT (*s*orting *i*ntolerant *f*rom *t*olerant, http://sift.bii.a-star.edu.sg/) and categorized as tolerated or deleterious based on the SIFT score cut-off. A SIFT score <.05 signifies potential disruption in protein function due to the amino acid change while changes with SIFT score >.05 are considered tolerated (Figure 7.1.3, 7.1.4).

Synonymous SNPs were also analyzed using the codon usage table (http://faostat3.fao.org/home/E) and frequencies of wild type codon and altered codon were compared (Figure 7.1.4). The changes in codon usage are discussed in following section.

Folate was estimated in RR fruits of accessions bearing SNPs in *FPGS* and *GGH* isoforms (Figure 7.1.5) but the experimental results were not conclusive due to lack of

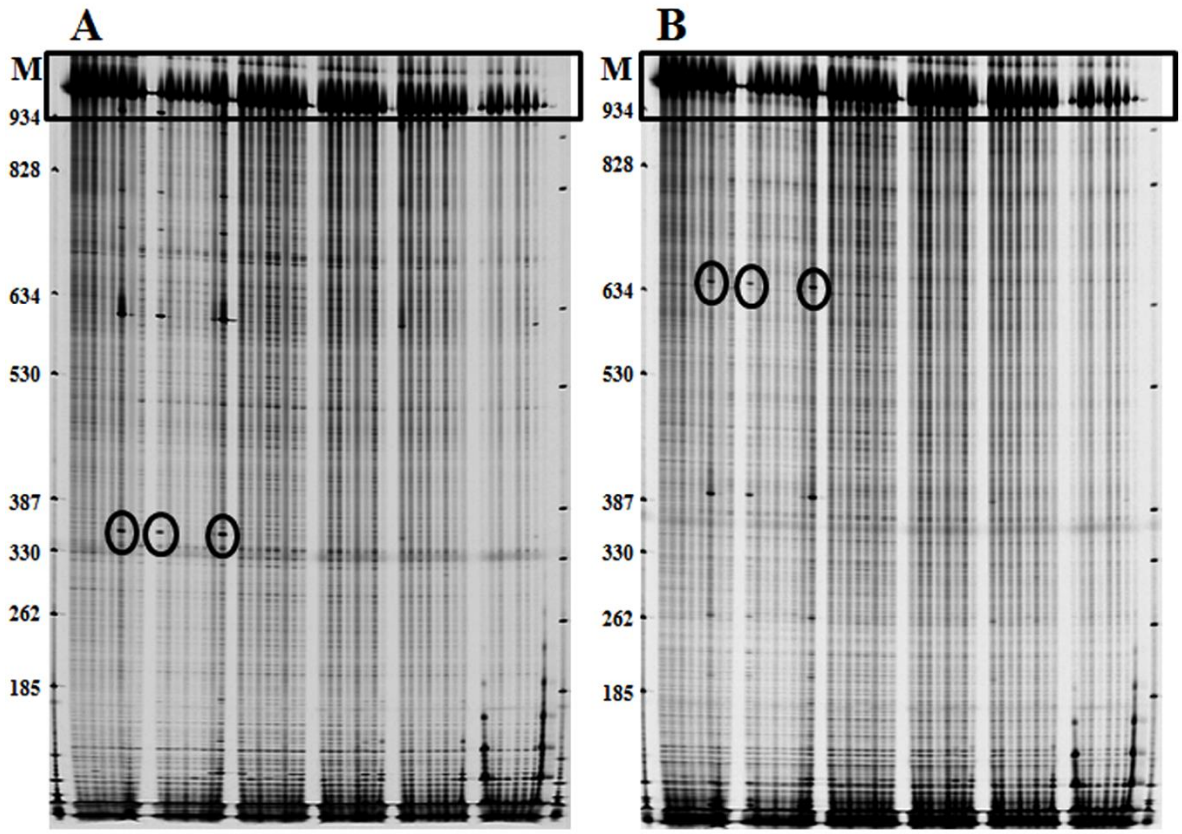


Figure 7.1.1. Detection of polymorphisms in *GGH3* gene in natural accessions. Gel images from the IRD700 (A) and IRD800 (B) channels of LI-COR analyzer. Each lane displays the 970 bp amplified region of *GGH3* gene digested with CEL I (box). Cleaved products, indicated by circles in both the channels of LI-COR. Molecular weight marker in bp indicated on right side of each panel.

Table 7.1.1. Summary of haplotypes scored on screening of genes involved in folate biosynthesis pathway. 391 tomato accessions were screened for SNPs.

Gene	Haplotype, 1 (no SNPs)	Haplotype 2	Haplotype 3	Haplotype 4	Haplotype 5	Haplotype 6
FPGSm	Arka Vikas and all other, accessions	EC252, EC487, EC1087	EC279088, EC398711, EC433607, EC3668, EC6845, EC7912, EC15127, EC369020, EC398684, EC241446, EC529083, EC443369, EC398699, EC398710, EC398716, EC398717, EC490130, EC520059, EC32557, EC33878, EC520046, EC520075, EC398688, EC398697, EC538139, EC531805, EC529081, EC168283	WIR3768	EC7317, EC7785, EC20639, EC27885, EC26150, EC27995, EC14073, EC27910, EC8591, WIR3957, EC520076, WIR13717, WIR3928, Type-1, TLBR-2, Castle-, Rock, H-88-87, EC520077, EC326139, EC370867, EC1914, 2-141, LA0517	EC14078, A.Arka
FPGSp	Arka Vikas and all other, accessions	EC3216, EC12689, EC27910, EC529083, EC6053-1, EC13904, EC30303, EC1914, H-24, H-88-78-5, M-88-78-3, P.Pink, IIHR2201, Sankranti	CLN2998	WIR3957, N2501	-	-
GGH1	Arka Vikas and all other, accessions	EC4506-2, EC6845 EC12689, EC13736 EC15127, EC398600, EC398711, EC398715, EC 338717, EC 362941, EC-27960, VLT-34, Nandi, TLBR-4, P.Pink	EC520079	EC487, EC8936	EC520076, WIR-3928, EC520077, BL-1208, IIHR2201, EC326146, EC531805	EC1087, EC14073
GGH2	Arka Vikas and all other, accessions	EC362949, EC529083, EC538141, EC521083, EC398695, EC26684	EC8936, WIR3768, EC2798	-	-	-
GGH3	Arka Vikas and all other, accessions	EC2790, EC3216, EC6192, EC12689, EC362949, EC320583, EC320571, EC546727, EC177297, EC528388, EC538146, EC521083, EC490128, EC520075, EC398699, EC398704, EC13904, EC30303, EC32557, EC520046, EC501577, EC368883, EC529085, EC526146, EC531805, EC529081, EC2798, EC129604	EC8936, WIR3768, EC20636, EC20639, EC26150, EC34477, EC163598, EC8936, EC34480, EC2977, EC520079, LA4104	EC520076, WIR3928, Type-1, EC528367, EC398405, BL1208, IIHR2201, EC321425, EC1087	-	-

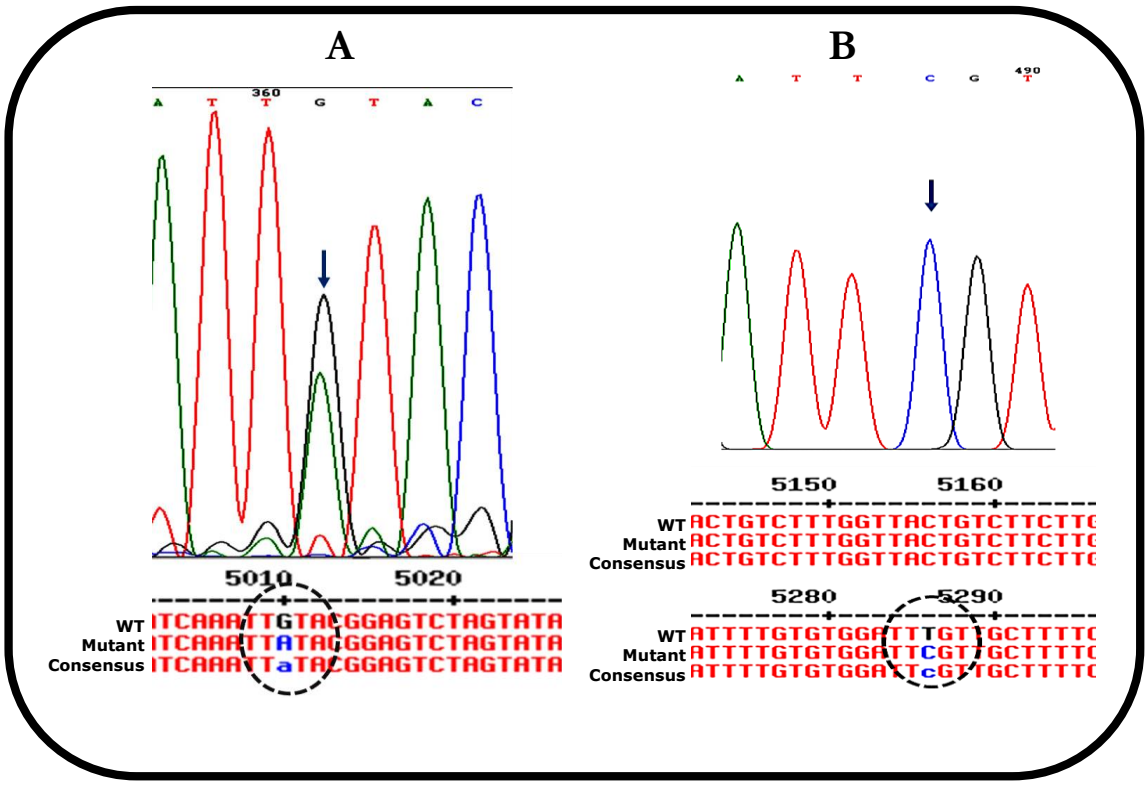


Figure 7.1.2. Confirmation of SNPs through sequencing. A= a heterozygous SNP (G to A) is visible after careful observation of chromatogram which, masked by the wild type peak, was not picked up in base calling. B= a homozygous SNP (T to C) is clear, unmasked and correctly base called (dark blue arrows).

Table 7.1.2. SNP confirmation of selected haplotype accessions through sequencing

Gene	Haplotype, 1 (no SNPs)	Haplotype 2 Sequenced accession and base change	Haplotype 3 Sequenced accession and base change	Haplotype 4 Sequenced accession and base change	Haplotype 5 Sequenced accession and base change	Haplotype 6 Sequenced accession and base change
FPGSm	Arka Vikas and all other, accessions	NA	EC398717, A2486G	NA	EC28356, A2486G, C2724T	EC14078, no SNPs found in sequencing
FPGSp	Arka Vikas and all other, accessions	EC27910, C3513G	NA	NA	-	-
GGH1	Arka Vikas and all other, accessions	VLT-34, Nandi no SNPs found in sequencing	NA	NA	WIR-3928 T2888C* G2984A C3139T	NA
GGH2	Arka Vikas and all other, accessions	EC362949, G4835A*, G5239A	NA	-	-	-
GGH3	Arka Vikas and all other, accessions	EC520046, T2166C, A2425:, T2554C, A2617T*, T2740C	LA4104, C1965T, C1976T, T2020C, G2169A, C2170T, C2184T, A2234T, A2290G, T2309C, G2383A*, C2499T	EC1087 T2166C, A2425:, T2554C, A2617T*, T2740C	-	-

NA= Sequencing was not successful after many attempts, -=no members in that haplotype, *=exonic SNP. **Blue fonts** show the accessions which was used for sequencing from a particular haplotype.

Solanum lycopersicum **[gbpln]: 1452 CDS's (634390 codons)**

fields: [triplet] [amino acid] [fraction] [frequency: **per thousand**] ([number])

```
UUU F 0.60 26.0 ( 16504)  UCU S 0.26 21.2 ( 13464)  UAU Y 0.60 18.6 ( 11799)  UGU C 0.62 10.8 (  6883)
UUC F 0.40 17.5 ( 11123)  UCC S 0.12  9.9 (  6268)  UAC Y 0.40 12.4 (  7855)  UGC C 0.38  6.7 (  4223)
UUA L 0.15 14.4 (  9111)  UCA S 0.25 20.7 ( 13102)  UAA * 0.40  0.9 (   598)  UGA * 0.37  0.9 (   553)
UUG L 0.25 24.2 ( 15336)  UCG S 0.07  5.6 (  3521)  UAG * 0.23  0.5 (   340)  UGG W 1.00 13.5 (  8563)

CUU L 0.26 24.9 ( 15814)  CCU P 0.39 19.2 ( 12192)  CAU H 0.67 15.5 (  9827)  CGU R 0.15  6.9 (  4404)
CUC L 0.12 11.2 (  7127)  CCC P 0.12  5.7 (  3645)  CAC H 0.33  7.8 (  4926)  CGC R 0.07  3.1 (  1946)
CUA L 0.11 10.0 (  6348)  CCA P 0.39 19.2 ( 12205)  CAA Q 0.60 21.0 ( 13315)  CGA R 0.12  5.4 (  3414)
CUG L 0.11 10.5 (  6665)  CCG P 0.09  4.6 (  2920)  CAG Q 0.40 14.0 (  8861)  CGG R 0.07  3.1 (  1981)

AUU I 0.50 28.2 ( 17914)  ACU T 0.39 19.9 ( 12612)  AAU N 0.64 30.5 ( 19339)  AGU S 0.19 15.2 (  9669)
AUC I 0.25 14.0 (  8906)  ACC T 0.17  8.6 (  5464)  AAC N 0.36 17.3 ( 11002)  AGC S 0.11  9.3 (  5897)
AUA I 0.25 14.0 (  8893)  ACA T 0.35 17.9 ( 11333)  AAA K 0.50 31.1 ( 19724)  AGA R 0.35 16.4 ( 10398)
AUG M 1.00 24.7 ( 15651)  ACG T 0.09  4.6 (  2939)  AAG K 0.50 31.0 ( 19651)  AGG R 0.25 11.9 (  7541)

GUU V 0.43 28.0 ( 17762)  GCU A 0.45 30.7 ( 19474)  GAU D 0.72 39.3 ( 24917)  GGU G 0.34 23.9 ( 15192)
GUC V 0.15 10.1 (  6382)  GCC A 0.15 10.1 (  6426)  GAC D 0.28 15.0 (  9509)  GGC G 0.14  9.7 (  6165)
GUA V 0.17 11.2 (  7126)  GCA A 0.33 22.2 ( 14107)  GAA E 0.57 34.8 ( 22053)  GGA G 0.37 25.6 ( 16258)
GUG V 0.25 16.0 ( 10181)  GCG A 0.08  5.2 (  3327)  GAG E 0.43 26.6 ( 16906)  GGG G 0.15 10.8 (  6839)
```

Coding GC 42.52% 1st letter GC 50.16% 2nd letter GC 39.87% 3rd letter GC 37.53%
Genetic code 1: Standard

Figure 7.1.3. Codon usage table to analyze the effect of synonymous changes based on the frequency of original codon and the codon resulting as a consequence of polymorphism.

Gene name	Synonymous change	Non synonymous change	Effect
GGH1	**-**	**V220A**	**Tolerated**
GGH2	**-**	**S237N**	**Deleterious**
GGH3	**A276=**	**-**	**Codon frequency 22.2→30.7**
GGH3	**Q255=**	**-**	**Codon frequency 14→21**

Figure 7.1.4.. Codon usage and SIFT analysis results of exonic SNPs identified in natural accessions. The SIFT cut-off score below 0.05 indicates deleteriousness of the SNP.

sufficient numbers of biological replicates (lost due to biotic and abiotic factors). However, EC27910 which showed an intronic SNP in *FPGSp* gene was carried forward based on the results obtained in LC-MS based folate estimation (chapter 4). This accession showed low folate levels therefore was further characterized to understand the mechanism of folate regulation (chapter 5).

7.1.4.1 DISCUSSION

Though, EcoTILLING is a variant of TILLING it allows detection of natural polymorphism in genes in large population. The species where chemical mutagenesis is not successful, EcoTILLING can be used as an alternate method understanding the gene function (Barkley and Wang, 2008). The natural variations in the gene identified by EcoTILLING can be correlated with the phenotype for deducing the gene function. The examination of natural polymorphism in the germplasm of crop species revealed that most of the natural polymorphism is present in intronic regions. Considering that any variation in the exons if not useful to organism is eliminated from the population, it is expected that variations in the introns would be higher than the exons. In addition most of the eukaryotic DNA is non-coding including introns which further explains the higher amount of polymorphism in introns. In 268 *Beta vulgaris* accessions, EcoTILLING of FL1a, FL1b, FT1a, FT1b and BTC1 genes revealed the presence of 21 SNPs. Interestingly, 18 SNPs were present in introns (Frerichmann et al., 2013). Similarly EcoTILLING in poplar revealed that rate of polymorphism in introns was 3.5 times higher as compared to exons (Gilchirst et al., 2006). A possible explanation for relatively few SNPs in the exon could be the extreme selection pressure which restricts the genetic diversity. Given that folates regulate very crucial metabolic reactions, it is not surprising that very less variability was present in terms of both number of SNPs as well as exonic SNPs. Another reason could be that only a small portion of the gene(s) was screened (approx. 800-1300 bp region) for all the genes with the intervening introns being much longer than the exons. There are no reports in the plant system regarding the SNP frequency for *FPGS* and *GGH* genes which makes it very difficult for us to assess the exact recalcitrance of these genes towards change.

The exonic SNPs were classified as synonymous (not causing amino acid change) or non-synonymous (causing amino acid change). These were analyzed based on their codon usage frequency and SIFT score respectively. Synonymous SNPs can also be effective sometimes in altering the translation efficiency through codon bias. An interesting observation was that in case of all the synonymous changes, the altered codon frequency

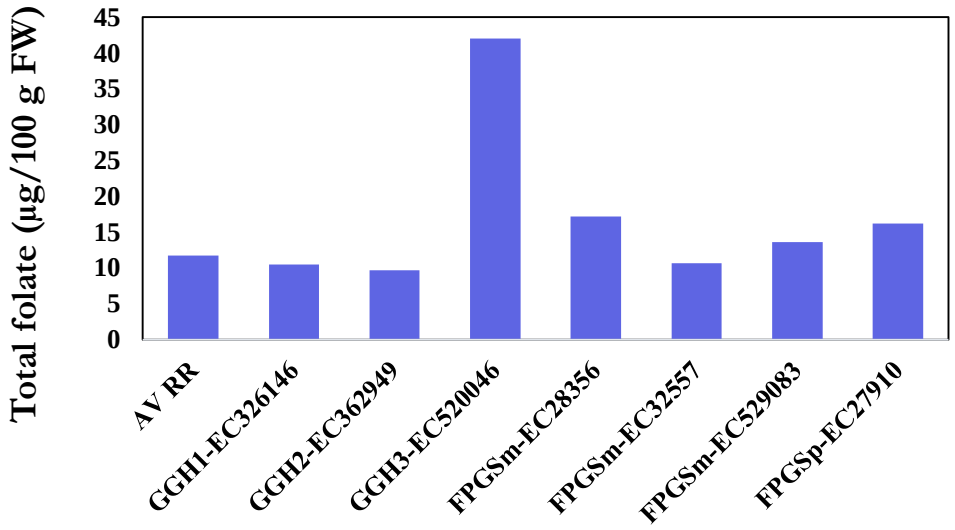


Figure 7.1.5. Folate content in red ripe (RR) fruits of EcoTILLING SNP harboring accessions. Respective gene names have been shown before the accession code. The data are for less than 3 replicates for each accession.

was increased from that of original codon which means that the reference genome (Heinz 1706) or Arka Vikas which are domesticated cultivars, have lost the better suited codons during the process of domestication, a loss of function caused by the biased domestication. It has been shown that human activities greatly affect and usually decrease the genetic diversity or cause huge changes in specific allele frequencies (Brown, 1992). This is quite possible as yield, shelf life, disease resistance, color, taste and aroma were the sole criteria during the course of tomato domestication. Even nowadays, tomatoes are not selected for their folate content due to the lack of any visual or molecular marker associated with this trait. These results provide interesting insights as the accessions showing favored codons could be thoroughly characterized and amount of proteins of biosynthesis genes could be compared with the wild type. In case of non-synonymous SNPs, accession EC362949 which showed many SNPs in *GGH2* gene, had a deleterious SIFT score (0.02) prediction for S237N change. No effect on total folate content was observed (Figure 7.1.5) in EC362949 as compared to AV, which may be due to functional redundancy in both, *FPGS* and *GGH* isoforms (Orsomando et al., 2005; Akhtar et al., 2008; Akhtar et al., 2010; Mehrshashi et al., 2010). Why few genes have large intron and short exons is not known. It is possible that the unusually large introns might serve some special purpose in the evolution of these genes which needs to be examined in details. It is also desirable to screen the accessions with intronic polymorphism for the presence of transcript variants as the intronic SNP in yellow flesh mutant of tomato encoded a defective phytoene synthase gene (Yuan et al., 2008). Our study also indicates that a better approach towards finding exonic polymorphism would be to screen coding DNA that can allow easy detection of polymorphism in such highly fragmented genes. Considering that the exonic polymorphism in these genes was low the tomato genome sequences available in public domain were used to find out the extent of polymorphism in these genes (section 7.3).

7.2 TILLING

A population of 2304 tomato mutagenized lines (120 mM EMS) was screened for mutations in *FPGSm*, *GGH1* and *GGH2* genes using LI-COR based TILLING. LI-COR based detection depends on efficient analysis of the gel picture using appropriate software and scoring of bands (putative mutations/ SNP/s). These bands represent the size of cut fragment as a result of CEL1 mismatch assay. This size of band in IR700 channel image of the LI-COR gel indicates the approximate distance of the SNP/mismatch from the forward primer binding site. In the same lane in IR800 channel image indicates the approximate distance of the SNP/mutation from the reverse primer binding site. The total sizes of both the bands for a true mutation/SNP should add up to the size of the main product, otherwise it could be an artifact. All such bands were scored for each lane (representative of individual mutant line).

7.2.1.1 Mutants identified in mitochondrial *FPGS* (*FPGSm*) gene

The screening of 3 row plates containing 8 fold pooled DNA of 2304 mutagenized tomato lines, seven lanes exhibited multiple cuts with complementary pairs on the gels indicating presence of the mutations. The veracity of these cuts being genuine mutations was confirmed by analyzing mutations in the respective column plates. In case the complementary mutations could not be confirmed by analyzing corresponding column plate, individual DNA of all 8 plants constituting the given well in the particular row plate well were screened and the individual mutant line was identified. Thereafter, a minimum of 5 individual M_3 plants were grown from seed packs corresponding to the M_2 leaf material of the respective line for identifying the mutant line. The screening of *FPGSm* gene showing mutations in tomato plant numbers were AV-M_2M_3-45B, AV-M_2M_3-384B and AV-M_2M_3-689A but, when DNA was isolated from corresponding M_3 seedling and presence of mutation was rechecked, mutation was confirmed only in lines 384B and 689A indicating that the mutation in AV-M_2M_3-45B is a false positive. The sequencing of *FPGSm* gene PCR product from the genomic DNA of lines AV-M_2M_3-384B and AV-M_2M_3-689A, presence of 5 mutations was confirmed in these two lines. The Sanger sequencing detected all the mutations that were present in the amplified region of the *FPGSm* gene notwithstanding the fact that some time the mutation may not be detected on LI-COR gel due to less intensity of cleaved fragment in the gel. The identified mutations were A2486G, A2496G, C2724T, T2899C, and C3015T (Table 7.2.1a). All five mutations were present in intron and represented transition, whereas no transversions were observed. It was surprising that two distinct lines showed same set of mutations.

7.2.1.2 Mutants identified in *GGH1* gene

The screening of *GGH1* gene resulted in identification of 16 lanes showing similar type of complementary cut fragment on the LI-COR gel. However, after confirmation in the column plate and using the individual DNA in some cases, only 8 mutations were confirmed. When M_3 seedling DNA from some of these lines was sequenced, all identified mutant lines displayed an identical change at A2739G. This mutation too was present in the intron.

7.2.1.3 Mutants identified in *GGH2* gene

The screening of *GGH2* gene resulted in identification of 3 lanes showing complementary cut fragments on the LI-COR gel and two of them were identical. After confirmation of plants in the column plate, individual M_3 plants were grown and the mutation was reconfirmed by Sanger sequencing. Mutant line number AV-M_2M_3-1017A and AV-M_2M_3-363B had similar mutation T5286C in *GGH2* gene. Line number AV-M_2M_3-273C showed a mutation G5004A in *GGH2* gene. All three mutations were present in the intron.

All the mutant lines when identified, represented M_3M_2 DNA as shown in figure 3.2, chapter 3. All subsequent characterization started from M_4M_3 seeds and resulting generation. For simplicity, only the mutant line numbers will be used hence forward.

7.2.2 Characterization and advancing of mutant lines to next generation

After identifying the M_3 plants bearing a given mutation these lines were examined whether any of these mutations affected the folate levels in fruits. The estimation of total folate content in the red ripe fruits of all these mutant lines did not reveal any significant difference from the level of folate present in red ripe fruits of parental line Arka Vikas (Figure 7.2.2). Normally intronic mutations do not affect the gene function directly, but there are reports where intronic mutation may affect function by determining gene splicing (Irimia et al., 2014). Therefore, to evaluate whether any of these mutations can affect the gene function, these mutations were analyzed with the help of NetGene2 Server (http://www.cbs.dtu.dk/services/NetGene2/). The NetGene2 server predicts how an intronic mutation can affect the splicing process. The mutations detected in *FPGSm*, *GGH1*, and *GGH2* gene were compared with the wild type sequence of the respective genes. NetGene2 predicted that there was a probability of additional acceptor splice site in line number 273C at 4983 position in *GGH2* gene because of the G5004A change. Barring this, other mutations showed no effect on gene function. Considering above possibility,

Table 7.2.1a. The mutants in *GGH1*, *GGH2* and *FPGSm* genes isolated using TILLING. The mutant line numbers with site of mutation shown in same color possessed identical mutation.

Gene name	Mutations detected	Mutations confirmed at plant level	Mutant line number (AV-M_2M_3)	Site of mutation	Type of mutation
GGH1	16	8	565C, 679B, 1234B, 127A, 157A, 208A, 953B, 1584C	T2739G	Intronic
GGH2	3	3	1017A, 363B	T5286C	Intronic
			273C	G5004A	Intronic
FPGSm	7	2	384B, 689A	A2486G A2496G C2724T T2899C C3015T	Intronic

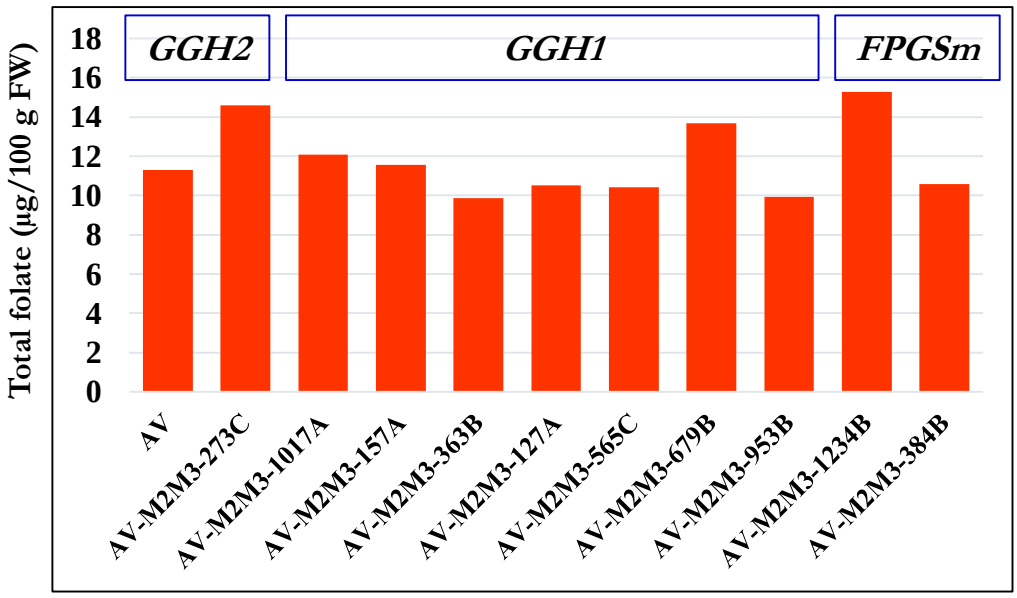


Figure 7.2.2. The level of folate n red ripe (RR) fruits of different mutant lines for *GGH1, GGH2* and *FPGSm* genes.
Respective gene names have been shown above the bar of each line. Data represent less than 3 replicates for each mutant line only means are presented and SE values are not shown.

the 273C line number was further characterized. The remaining mutant lines were not taken forward for further analysis.

7.2.2.1 Characterization of *GGH2* mutant line 273C

The identification of plants bearing *GGH2* mutation revealed that the mutation was in heterozygous state. Since recessive mutation shows phenotype only in the homozygous state, the plants bearing homozygous mutation were identified by advancing them to next generation. Out of ten seeds of 273C M_3 line grown for mutation detection in *GGH2*, 3 seeds failed to germinate, and out of germinated 7 seeds only one seedling contained the mutation (G5004A). This seedling was grown to obtain homozygous mutant plants in M_4 generation. However even in M_4 generation none of the plants bore homozygous mutation and plants were advanced to M_5 generation.

Both in M_3 and M_4 generation the heterozygous plants yielded very few seeds and seed germination was also poor. Out of 38 M5 seeds (due to paucity of seeds in line 273C) 6 seeds did not germinate. Out of 32 seedlings, 15 were heterozygous for mutation and only one seedling was homozygous for mutation. The DNA was also extracted from six ungerminated seeds, two were heterozygote and one was homozygote. The parent being heterozygous for mutation the expected ratio is 1:2:1 for WT:heterozygote:homozygote. However, out of 38 plants only two were homozygote, among which one DNA represented that of un-germinated seed. This indicated highly skewed ratio of genetic segregation (19:17:2) and that the homozygote mutations may be largely lethal in the nature. (Table 7.2.1b). The single homozygous mutant individual isolated in M_5 generation was characterized. This line was crossed with AV for eliminating background mutations and F1 seeds were obtained.

7.2.2.2 Phenotypic characterization of 273C

The EMS-mutagenesis randomly targets genomic DNA for mutations. Based on frequency of mutation we have calculated that on average a single EMS-mutagenized tomato plant likely contains about 2000 background mutations. This necessitates elimination of background mutations at the same time retaining the desired mutation by repeated backcrossing. Nonetheless, a preliminary characterization of homozygous plant helps to determine if the mutation and phenotype has any linkage. If multiple alleles in same gene have same phenotype, this can establish genotype phenotype linkage. Since so far no *GGH* mutant has been reported there is no *a priori* information about expected phenotypic changes.

Overall growth and morphological characters of line 273C did not significantly differ from AV. The leaves of the 273C showed a slight downward curling and the leaflet size was smaller than the wild type (Figure 7.2.3). Few differences were observed in flower and fruit morphology. The flowers of AV characteristically show exerted style and stigma from the anther cone. Interestingly, the exerted stigma in 273C was much longer and was at an angle which pierced the anther cone at the point of exit (Figure 7.2.4B). The damage to the anther cone resulted in complete failure of pollination. The manual pollination was required for fruit set and obtaining the seeds. The successful manual pollination also indicated that the pollens were fertile. The fruit phenotype was also different in the surface morphology from wild type. The fruits of 273C had a distinctive very glossy surface as compared to the wild type fruit at both MG and RR stages of development (Figure 7.2.4 D and F). Another difference was in the fruit firmness (data not available due to technical problem). Fruits of 273C were extremely soft right after the turning stage which continued till RR stage.

7.2.2.3 Full gene (*GGH2*) screening of mutant line 273C

Based on CODDLE prediction for most mutagenic region only 836 bp of *GGH2* gene was screened by TILLING. To find out the presence of mutations in the remaining portion of *GGH2* gene 6 sets of primers were designed covering the entire *GGH2* gene and both the UTRs (8370 bp in total) (Table 7.2.2, Figure 7.2.5) and PCR products were sequenced. It was also considered that 273C line phenotypic traits might be because of some other mutation in *GGH2* gene. In parallel, same primer sets were used to amplify the genomic DNA of AV and the amplicons were sequenced. The sequencing reads were aligned using MultAlin. The alignment showed that no other mutation was present in the line number 273C.

7.2.2.4 Folate content of mutant line 273C in M_5 generation

Total folate content was estimated in the 4th leaf, MG and RR fruits of 273C WT-homo, heterozygous and homozygous mutant plants and compared with that of WT parent (AV). No significant difference was found in the folate levels in leaves of 273C compared to wild type. Only in one plant that is heterozygous 273C RR folate levels was different from wild type parent. The others lines did not show any significant difference in the folate level (Figure 7.2.6 A, B). The relative proportions of folate derivatives was also similar to the wild type with a considerable increase in 5,10-Methenyl THF in the RR fruit than in MG fruit. This rise was observed in 273C as well (Figure 7.2.7A, B). Though, mutant plants did not show any change in folate levels, the homozygous mutant plant was

Table 7.2.1b. The **s**eed germination pattern of line 273C with mutation in *GGH2* gene from M_3 to M_5 generations.
Note:- * first leaf of seedlings from germinated seeds. ** the embryo was excised removing the seed coat and used for DNA isolation

Generation	**Germination status**	**Mutant:WT**
M3	7/10	1 (het):6
M4	8/10	5 (het):3
M5 38/50 yielded DNA, others lost to fungal infections	32 germinated seedlings DNA*	16 (15 het+1 homo):16
	6 un-germinated embryo DNA**	3 (2 het+ 1 homo):3
Overall mutation ratio in M5 generation **(hetero:homo: WT)=17:2:19**		

Figure 7.2.3. Phenotype of AV and 273C homozygous mutant individual in M5 generation. A, C and E show whole plant and canopy structure and leaf phenotype of Arka Vikas. B, D and F whole plant and canopy structure and leaf phenotype for 273C. Blue arrows indicate downward curling of mutant plant leaves. Both wild type and mutant leaf pictures are from 7th node.

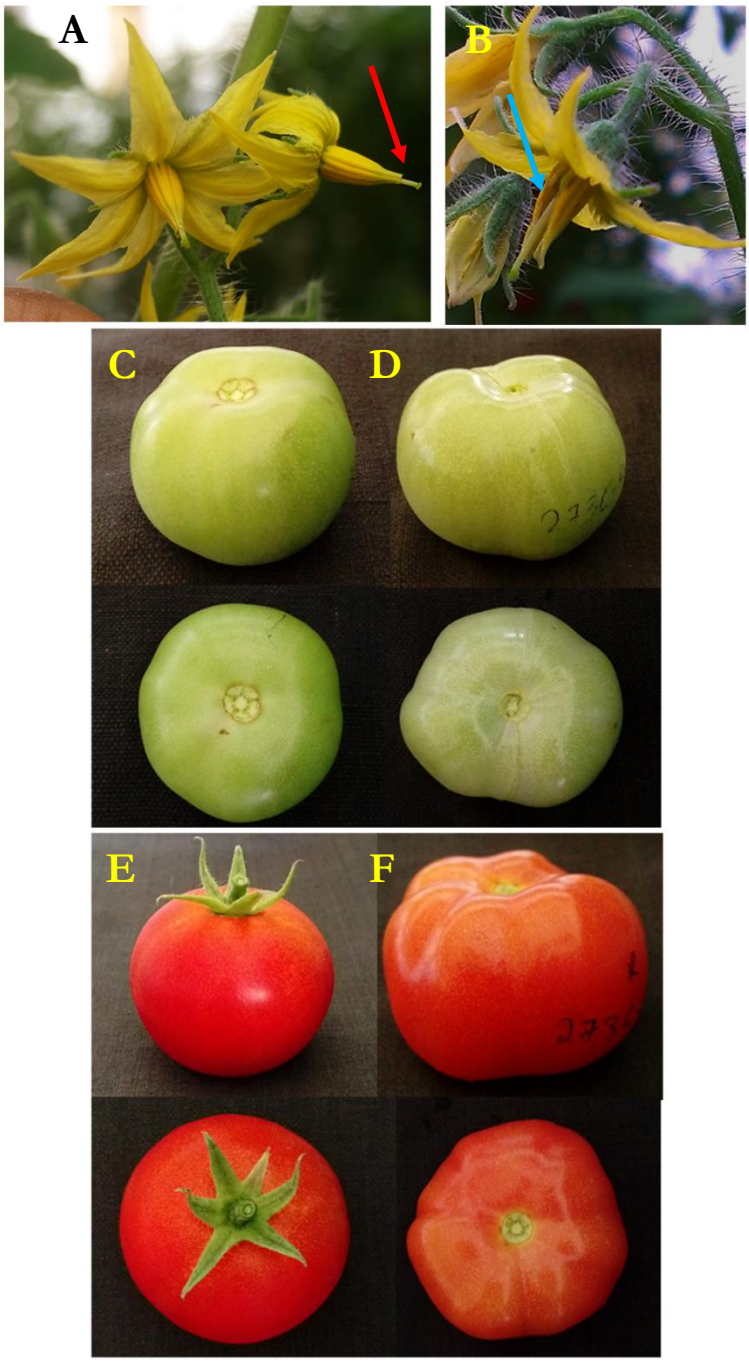


Figure 7.2.4. Phenotype of AV and 273C homozygous mutant individual in M5 generation. A, C and E show flower, MG and RR fruits of Arka Vikas. B, D and F show the same for 273C. The stigma in 273C exerted (blue arrow) out of the anther cone by a higher extent than AV (red arrow). Anther cone in 273C was broken/disrupted at an angle in most of the flowers so that manual pollination was required for fruit set. Fruits of this mutant line showed a very glossy surface as compared to the AV right from the mature green stage till red ripe stage.

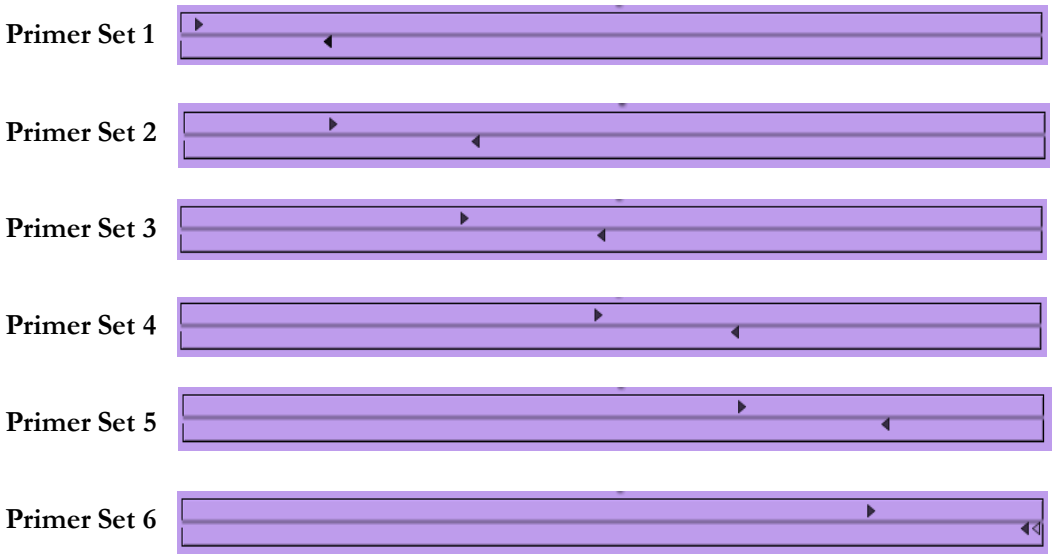


Figure 7.2.5. Full length *GGH2* gene was amplified using 6 set of overlapping primers that also included 5′ UTR and 3′ UTR. A total of 8370 bp region of *GGH2* gene were screened.

Table 7.2.2. The Span of primer sets used for screening of full *GGH2* gene in mutant line 273C, including both 5′ and 3′ untranslated regions (UTRs)

Primer set	GGH2 gene region covered
1	-1184 to +301 bp
2	+222 to +1670 bp
3	+1501 to +2933 bp
4	+2717 to +4269 bp
5	+4188 to +5644 bp
6	+5441 to +7030 bp (stop codon at +6319 bp)

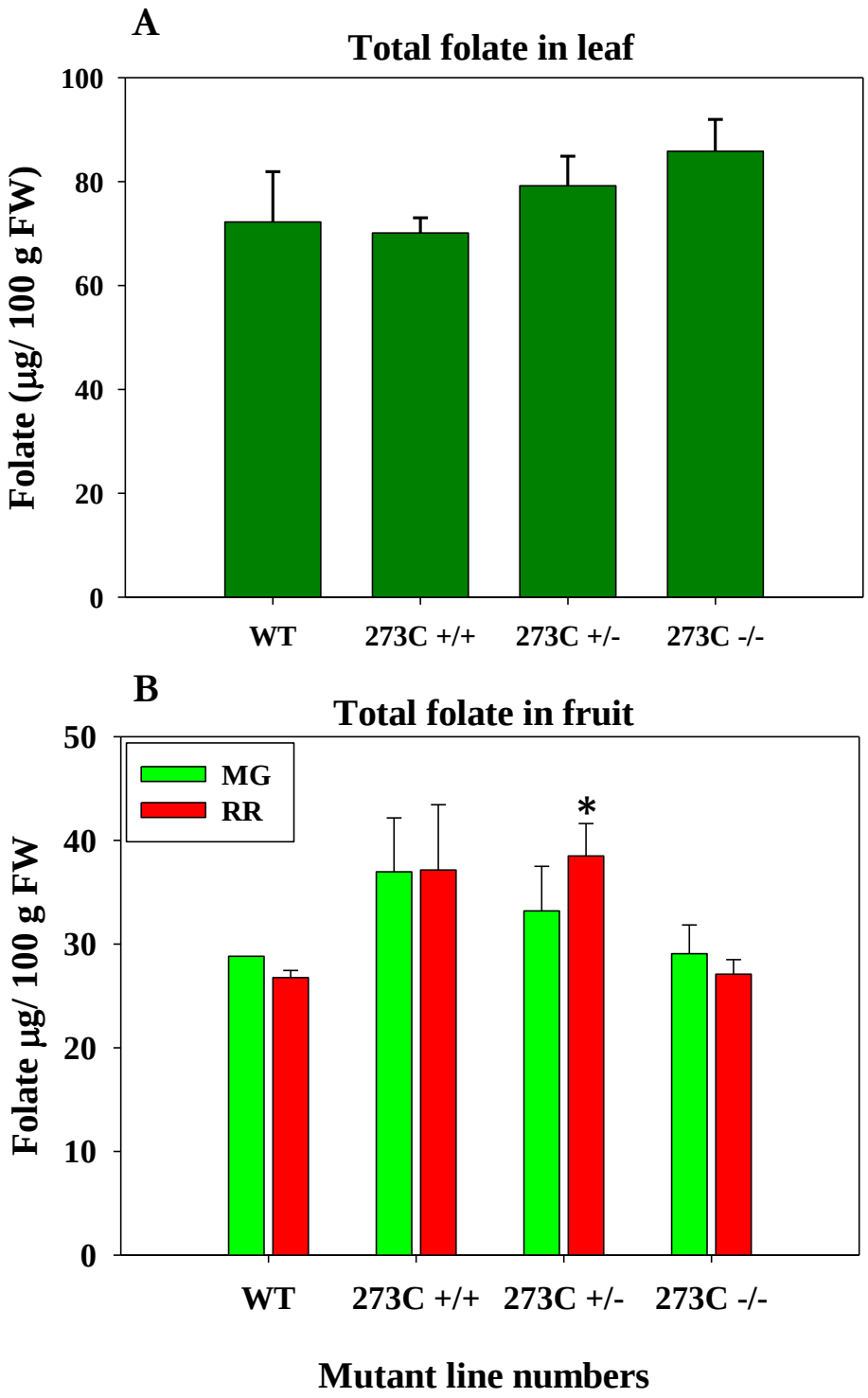


Figure 7.2.6. Folate content in *GGH2* mutant line 273C in M5 generation. A=total folate in leaf, B= A=total folate in fruit. Non mutant sibling (+/+) of the same line was also studied to ascertain the effect of background mutations.

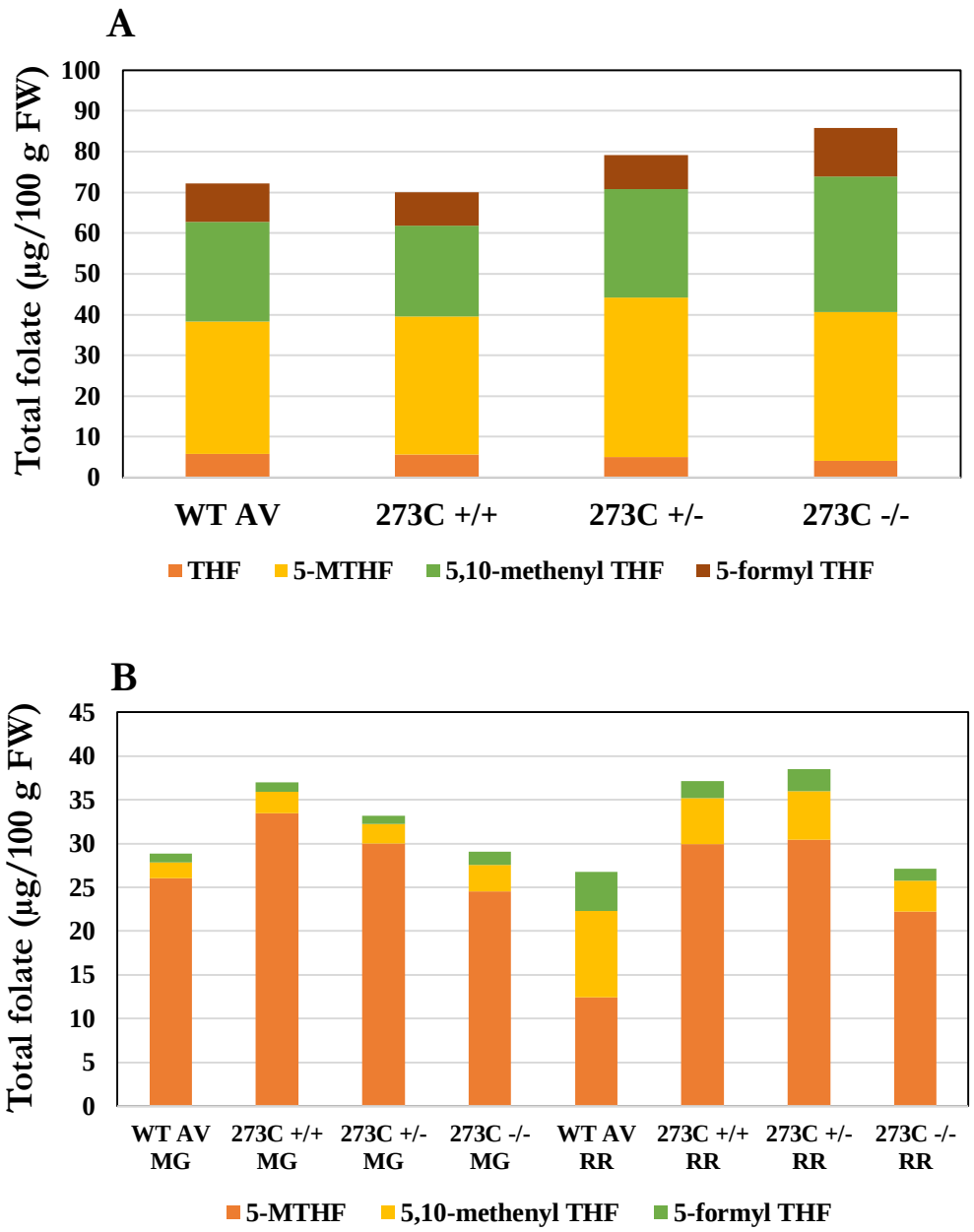


Figure 7.2.7. Relative proportion of different folate forms in *GGH2* mutant line 273C in M5 generation. A=folate forms in leaf, B=folate forms in fruit.
WT (+/+) sibling (which did not contain this mutation) of the same line was also studied to ascertain the effect of background mutations.

crossed with the wild type to eliminate background mutations to examine the possibility if the background mutations have adversely affected the folate levels.

7.2.2.5 Amplification of 273C cDNA using *GGH2* specific primers

Since this mutation was intronic and there was a prediction of extra acceptor splice site, we amplified the *GGH2* cDNA in 273C leaf, MG and RR fruit along with wild type using two primers sets covering entire cDNA. However, this gene did not show the presence of any splicing variant (Figure 7.2.8).

7.2.3 TILLING by sequencing

A population of 768 mutagenized (120 mM EMS) lines of tomato was screened for presence of mutation using next generation sequencing (NGS) of the PCR products amplified for select genes. The approach involved pooling of 768 mutagenized plant DNA into 3 different types of pools with 64, 48, and 48 fold pooling depth namely D, C (column) and R (row) pools (Tsai et al., 2011). After sequencing, the mutant allele's presence in all the three pools confirmed the trueness of mutation. One advantage of NGS is it identifies the mutation thus obviates the need of mutation validation by Sanger sequencing. *GGH1* and *FPGSp* genes were examined for mutation using NGS approach which yielded 4 mutations present in 3 plants (Table 7.2.3). Two mutations were found in the same plant in *GGH1* gene (G2984A and C3139T) while two additional mutations were found in *FPGSp* gene (C3513G and G3643A) in two different plants. Most importantly G3643A in the *FPGSp* gene was the only exonic mutation identified among all the isoforms of *FPGS* and *GGH* genes. However, the G3643A mutation in the *FPGSp* gene was synonymous (L164=) therefore did not result in any amino acid change. Nonetheless, examination of codon usage frequency in tomato revealed a drastic difference between the original and mutated codon frequencies (UUG=24.2 to UUA=14.4) which may affect its translation efficiency (Table 7.2.3). This mutant was identified very recently and therefore its characterization is underway.

7.2.3.1 Mutation frequency observed in the genes

LI-COR and NGS based TILLING approaches, though restricted to a smaller region of the gene (predicted by CODDLE) comprising 800-1300 bp, showed nearly the same mutation frequency (Table 7.2.4). Mutation frequency of one mutation per 224 to 3048 kb was observed.

7.1.3.1 DISCUSSION

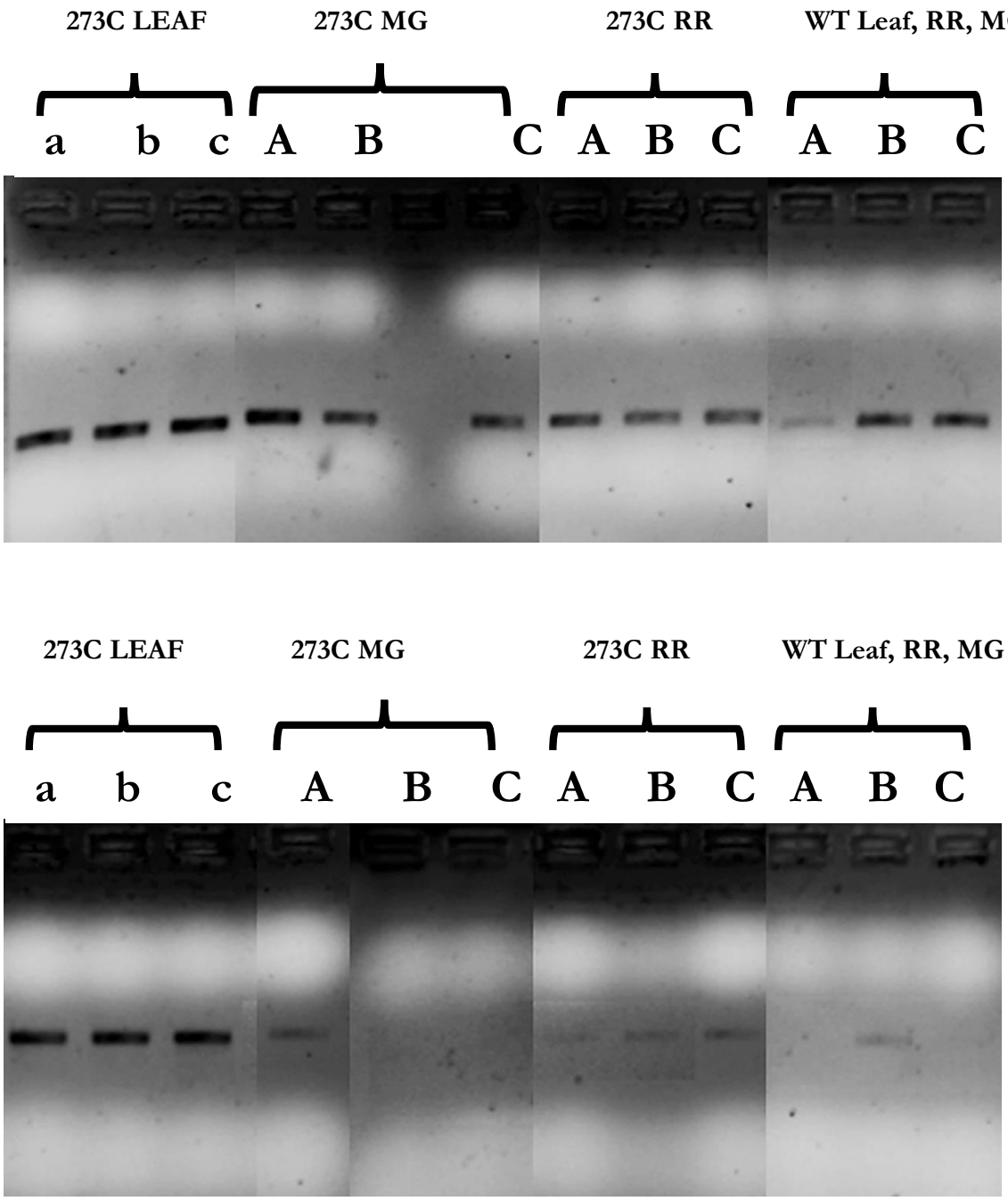


Figure 7.2.8. Amplification of *GGH2* cDNA in mutant line 273C in M5 generation.
273C LEAF a, b and c= leaf cDNA of -/-, +/+ and +/- 273C respectively.
A, B and C= fruit cDNA of -/-, +/+ and +/- 273C respectively in mature green.
A, B and C= fruit cDNA of -/-, +/+ and +/- 273C respectively in red ripe.
WT Leaf, RR, MG A, Band C= cDNA of leaf, RR and MG of Arka Vikas.
Upper panel=Amplification with primer set 1, Lower panel=Amplification with primer set 2. Primer sequences are-

SET 1

Forward 5′TGTTCAATACTCAAAGACCACACA3′, Reverse 5′GTTAAAAGTTCAAAACCCAAGCAG3′

SET 2

Forward 5′GCAAGAAGGGTCTCTACTTTGAA3′, Reverse 5′CAAACACTTGGAAAGCATAAAGG3′

Table 7.2.3. The mutants in *GGH1* and *FPGSp* isolated using NGS based TILLING. Both mutations of *GGH1* were in the same mutant line AV-M_2M_3 413A whereas both the mutations of *FPGSp* were present in two separate mutant individuals.

Gene name	Confirmation	Mutation site	Mutant line number (120 mM EMS AV-M_2M_3)	Type of mutation
GGH1	Confirmed	G2984A	413A	Intronic
	Confirmed	C3139T	413A	Intronic
FPGSp	Confirmed	C3513G	849A	Intronic
	Confirmed	G3643A (L164=)	**1302D**	**Exonic**

Table 7.2.4. Overall mutation frequency observed in each gene screened through TILLING using either LI-COR or NGS approach.

Gene	Population size	Screening method	Mutation frequency
GGH1	2304	LI-COR/Cel1	1/3048 kb
GGH2	2304	LI-COR/Cel1	1/963 kb
FPGSm	2304	LI-COR/Cel1	1/461 kb
GGH1	768	NGS	1/234 kb
FPGSp	768	NGS	1/224 kb

$$\text{Mutation frequency} = \frac{\text{Total number of mutations identified}}{(\text{Number of individuals screened x Number of base pairs screened})}$$

The screening of the FPGS and GGH genes revealed the presence of only intronic mutations. Among these only one line 273C was predicted to form a splice site variant and mutation in the identified M_3 line was in heterozygous state. To judge whether this putative mutation may have phenotype this line was characterized in detail. It is reported that the GGH activity is related to cellular folate homeostasis (Orsomando et al., 2005; Akhtar et al., 2010). The examination of fruit folate levels in M_3 generation of 273C line did not show any difference from wild type. However, its fruits were relatively small and contained very few seeds, a character observed in tomato transgenic lines overexpressing *GGH2* under the influence of E8 promoter (Akhtar et al., 2010). The mutant seeds also show high percent of mortality during germination in M_3 and M_4 generation.

However, to ascribe a phenotype to a given gene it is essential to demonstrate that its genetic segregation is tightly linked with the phenotype. The crossing with parental line also eliminates the undesired background mutations which are spread all over the genome in case of an EMS-mutagenized line. The advancing of the originally isolated heterozygous mutant M_3 line till M_5 generation allowed identification of one homozygous mutant. Apart from leaf and fruit morphology, no differences were perceptible between the WT hetero, and homozygous mutant individuals as compared to wild type parent with respect to total folate content or fruit weight. Normally intronic mutations do not affect the phenotypes, however in several instances it is reported that intronic mutations cause phenotypic changes. For example in yellow flesh tomato mutant an intronic SNP disrupts the biosynthesis of carotenoids (Yuan et al., 2008). Even synonymous mutations can affect gene functions by regulating mRNA splicing, stability and translation regulation where a preferred synonymous codon is more efficiently translated (Shabalina et al., 2013).

Considering the NetGene2 prediction that this mutation was intronic and there was possibility of extra acceptor splice site, the *GGH2* cDNA in 273C leaf, MG and RR fruit along with wild type was examined. However it did not show the presence of any splicing variant (Figure 7.2.8). Though this largely eliminated the possibility of this mutation affecting the folate level, yet the remote chances of mutation affecting phenotype cannot be completely ruled out. It is reported that the alternate splicing event requires certain internal cues and it might not be a regular occurrence in the cell (Staiger and Brown, 2013) despite the presence of potential site for that. Unlike studies with RNAi where, downregulation of *GGH* caused significant increase in folate level in tomato (Akhtar et al., 2010) in the line 273C, folate levels were similar to the wild type. Considering all these facts together, the results indicate that the *GGH2* intronic mutation is not directly involved in

regulation of total folate content. The observed leaf, flower and fruit variations appear to be due to other mutations.

Compared to 2300 individuals examined for mutation using LI-COR based TILLING, a relatively small population of 768 EMS mutagenized tomato lines was examined for mutations using the next generation sequencing. The main variations from LI-COR based TILLING were- (1) in the size of amplified fragment (600 bp instead of 1500 bp in LI-COR), (2) pooling depth (max. 64 fold instead of 8 fold in LI-COR) and (3) identification of exact location of mutation while in LI-COR based TILLING, exact location is confirmed by Sanger sequencing of the individual putative mutant DNA. *GGH1* and *FPGSp* genes were selected for NGS TILLING. Two mutations were identified in *GGH1* gene which were located in intron. In *FPGSp* too, 2 mutations were identified which were confirmed through CEL1 mismatch assay. One of the mutations was located in exon but it was a synonymous change (L164=). Despite being a synonymous change, this mutation might affect the translation efficiency due to much lower frequency of the changed codon as compared to the original codon and is being characterized.

The examination of frequency of mutations in folate polyglutamylation and deglutamylation genes *FPGS* and *GGH* indicated that these genes were recalcitrant to mutagenesis. Out of 12 different single base changes identified among four genes/isoforms, only one mutation was located in exon. The other intronic mutations barring G5004A in *GGH2* gene were also not predicted to cause any splicing aberrations through the NetGene2 server prediction tool. All the *FPGS* and *GGH* isoforms are characterized by very short exons and fairly longer introns. Although, GC:AT ratio was similar in both introns as well as exons which are the typical changes expected in an EMS population, still most of the mutations were present in introns only. There could be some significance of longer introns other than causing the splicing variants for the rarity of exonic mutations in these genes and requires novel experimental designs for further exploration.

The absence of mutation in folate metabolism genes highlights the importance of these genes to maintain the cellular homeostasis in the cell. Since folate is a critical molecule needed for several metabolic reactions in the cell, its deficiency will be lethal to the organism. Even over-production of the folate can also affect the folate metabolism. Therefore, the living organisms maintain folate levels within optimal range and any deviation beyond it leads to lethality. The absence of any reported mutants in folate biosynthesis and metabolism barring one in *FPGS* in Arabidopsis and maize strongly

points the importance of folate for the survival of plants. This is further exemplified with the observation that only exonic mutation identified in this study in *FPGSp* gene was synonymous in nature.

The range of mutation frequency was observed in the *GGH1*, *GGH2*, *FPGSm* and *FPGSp* genes was analyzed with reference to genome size. The observed range was one mutation per 224 to 3048 kb in these genes. While it was 1 mutation per 765 kb-3427 kb in a 0.5% (w/v) EMS mutagenized Micro-Tom population (Okabe et al., 2011) for SlETR1 and SlSSADH genes, the frequency increased after doubling the EMS dose to 1% (w/v) to 1 mutation per 612 kb-685 kb for the same genes respectively. However, in case of SlPL gene frequency decreased from 1 mutation per 865 kb-1175 kb. For the same mutagen concentration 1% (w/v), the frequency varied among three cultivars from 1 mutation per 322 kb, 734 and 734 kb for Red setter, Micro-Tom and TPPADASU varieties respectively (Okabe et al., 2013).

Mutation frequency is subject to various factors which can affect the TILLING output such as population size, gene coverage, cultivar, EMS concentration and screening method. Though, this study used 120 mM EMS (equivalent to 1.49% [w/v]) for mutagenesis, the observed mutation frequencies for polyglutamylation and deglutamylation genes were generally low. This could be because of two main reasons, the overall recalcitrance of cultivar Arka Vikas towards mutagenesis and the critical roles of genes in the folate metabolism selected in the study. Since, large variations were scored even in *GGH1* mutation frequency alone between LI-COR and NGS screening methods, this could be explained by the differences in population size and amplicon size in both the methods.

The rarity of tightly linked exonic polymorphism or mutations in folate turnover genes could be due to the vital functions of these genes. There are reports that the genes involved in tomato fruit development, ripening, light perception and signaling exhibit low variability (Mohan et al., 2016) and no variability was scored in *PSY1* and *RIN* genes. However, screening of smaller genic fragment could also be a probable cause for such results.

Since we observed a scarcity of variations in our study involving EcoTILLING and TILLING, we checked the overall variations occurring in the folate turnover genes in recently sequenced genomes of tomato through *in silico* approach. The data is publicly available, including both, cultivated varieties and wild relatives. Only exonic, non-synonymous SNPs were examined which are discussed in the following section.

7.3 *In silico* survey of 444 accessions of tomato for polymorphism in polyglutamylation (*DHFS*, *FPGS*) and de-glutamylation (*GGH*) genes

7.3.1 Diversity of non-synonymous SNPs detected in all the accessions

The completion of tomato genome sequence in the year 2012 (Tomato Genome Consortium, 2012) followed by resequencing of large number of tomato cultivars and natural accessions to find out novel alleles and also to closely decipher the phylogenetic lineage of tomato and its close relatives (Aflitos et al., 2014; Lin et al., 2014). The genome sequences of different cultivars that are currently available in public domain were examined for polymorphism in *FPGS* and *GGH* gene isoforms. In conjunction with *FPGS* and *GGH* gene(s) an additional gene involved in the folate metabolism-dihydrofolate synthase (*DHFS*) was also examined. The first glutamate residue addition by DHFS to dihydropteroate (DHP) leads to formation of dihydrofolate (DHF) and is an important step of folate biosynthesis. Since the genome sequencings were carried out in two independent studies the reported gene polymorphisms are separately summarized for each study as there were differences in the sequencing depths and selection basis of sequenced accessions. The protein sequence variations resulting due to the non-synonymous SNPs were aligned which constituted the protein groups for each gene. For visualization purpose, only those amino acid residues where changes existed in at least one of the groups were shown along with the similarity tree.

Selection of material in the 84 accessions consisted of first set of 54 accessions of tomato landraces and heirloom cultivars of *S. lycopersicum* and S. *lycopersicum* var. cerasiforme, which were taken from EU–SOL tomato core collection (https://www.eu-sol.wur.nl). The second set of 30 accessions (4 *S. pimpinellifolium,* 2 *S. peruvianum,* 2 *S. chiemliewskii,* 3 *S. cheesmaniae,* 2 *S. neorickii,* 2 *S. arcanum,* 3 *S. huaylasense,* 2 *S. chilense,* 7 *S. habrochaites,* 2 *S. pennellii,* and 1 *S. galapagense*) included wild relatives of tomato. Overall, these accessions represented the full range of expected genetic variation within *S. lycopersicum* (Aflitos et al., 2014). On the other hand, 360 tomato accessions were collected from TGRC (Tomato Genetics Resource Center), USDA (US Department of Agriculture), EU-SOL (European Union Solanaceae project), INRA (National Institute for Agricultural Research) and IVF-CAAS (Institute of Vegetables and Flowers, Chinese Academy of Agricultural Science) and included 10 wild tomato accessions (1 *S. habrochaites*, 3 *S. cheesmaniae*, 1 *S. galapagense*, 3 *S. peruvianum*, 1 *Solanum neorickii* and 1 *S. chilense*), 53 PIM (S. *pimpinellifolium*) accessions, 112 CER (*S. lycopersicum* var. cerasiforme) accessions, 166 BIG

(big-fruited *S. lycopersicum*) accessions (2 accessions were excluded for extreme phenotype segregation) and 17 modern commercial hybrids (F1) (Lin et al., 2014).

In the 84 tomato genomes project, an extensive genotyping of 7000 EU-SOL accessions was performed by using 384 SNP markers to finally select the 84 accessions which embodied full genetic diversity of the crop. This selection strategy allowed the enrichment of genetic diversity which is evident from the results (Table 7.3.1). Though the number of accessions in the 84 genome sequencing project was nearly 4 times less than the 360 genome project, yet, greater diversity was observed in 84 genome project in the analyzed genes. The exception being the *FPGSm* and *GGH1* which consistently showed huge number of SNPS and resulting protein changes in both the projects. Though the number of protein groups formed for DHFS, FPGSm, FPGSp, GGH1, GGH2 and GGH3 were 21, 30, 19, 29, 20, 25 and 5, 41, 6, 35, 11 and 19 in 84 and 360 accessions (Figures 7.3.1 to 7.3.6), these classifications are not comparable. For example 360 accession study revealed only three changes in amino acids for DHFS compared to 47 observed in 84 accession study. Though 360-accession showed only three changes two changes at position 9, 33 and 417 did not overlap with 84-accesssion study. Similarly a higher number reported for position 222 to 397 in the C-terminal of the protein. It is well established observation that in most proteins N-terminal is more conserved and has catalytic activity whereas C-terminal is more variable and subject to post-translational modifications (Wall et al., 2016).

The less number of classes for FPGSp is very similar to that described for DHFS above with only six changes in amino acids in 360-accessions compared to 38 amino acid changes in 84-accessions. The higher number of classes in the GGH1 is apparently due to presence of extra SNPs detected leading to changes in amino acids at position 127 and 272 in 360-accessions. Similarly in GGH2 three extra amino acids at position 226, 237 and 313 in 360-accession increased the number of protein classes to 11, though only seven amino acid changes were found. The detection of 19 classes for GGH3 in 360-accession likewise may be related to presence of extra amino acid at position 313.

The reason for more amino acid changes detected in 84-accession compared to 360-accessions though not completely known but can be ascribed to design of the respective study. The 84-accession study started with a much wider base of germplasm using SNPs to select lines with maximum diversity. Secondly 84-accession study included 30 wild relatives compared to 10 used in 360-accessions leading to greater degree of genic

Table 7.3.1. Overall summary of phylogenetic groups for each gene and number of amino acid changes per gene

84 Accessions		
Gene name/protein length	**Number of groups**	**Total number of residues with change**
***DHFS*/554**	21	48
***FPGSm*/401**	30	100
***FPGSp*/543**	19	37
***GGH1*/340**	29	59
***GGH2*/344**	20	30
***GGH3*/337**	25	46
360 Accessions		
***DHFS*/554**	5	4
***FPGSm*/401**	41	19
***FPGSp*/543**	6	6
***GGH1*/340**	35	14
***GGH2*/344**	11	7
***GGH3*/337**	19	7

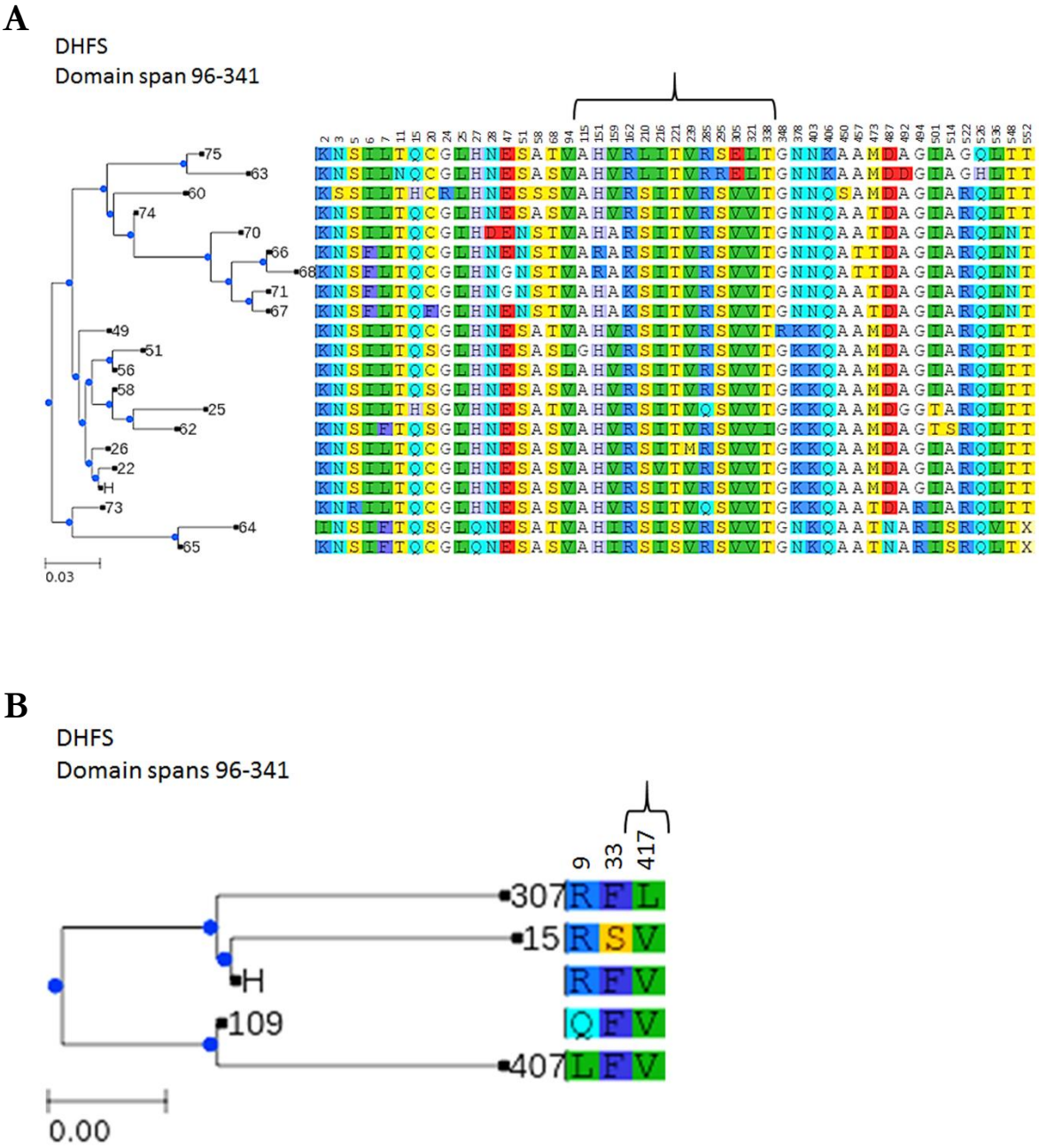


Figure 7.3.1. Protein similarity tree co-aligned with the amino acid positions in DHFS protein which showed change in minimum one of the groups of tree. 'H' represents the Heinz 1706 (reference genome) group while numeral values represent the number of any of the accessions of that particular group. The accessions names corresponding to each digit can be found in Annexure 2, 2.1, 3 and 3.1 for 84 and 360 accession respectively.
A=Changes observed in 84 accessions, B=Changes observed in 360 accessions.
Braces show the changed residue position which falls in the catalytic domain.

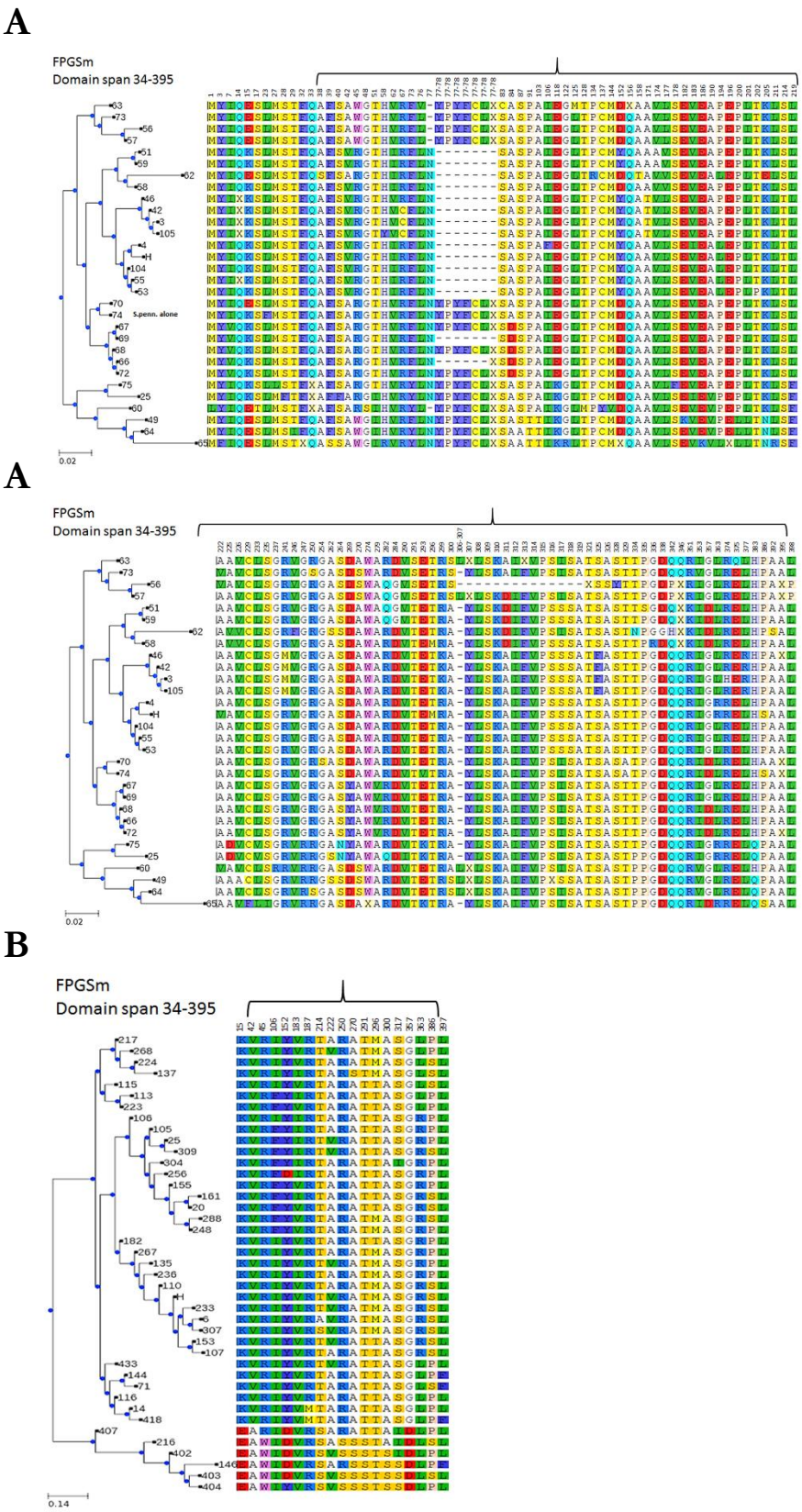
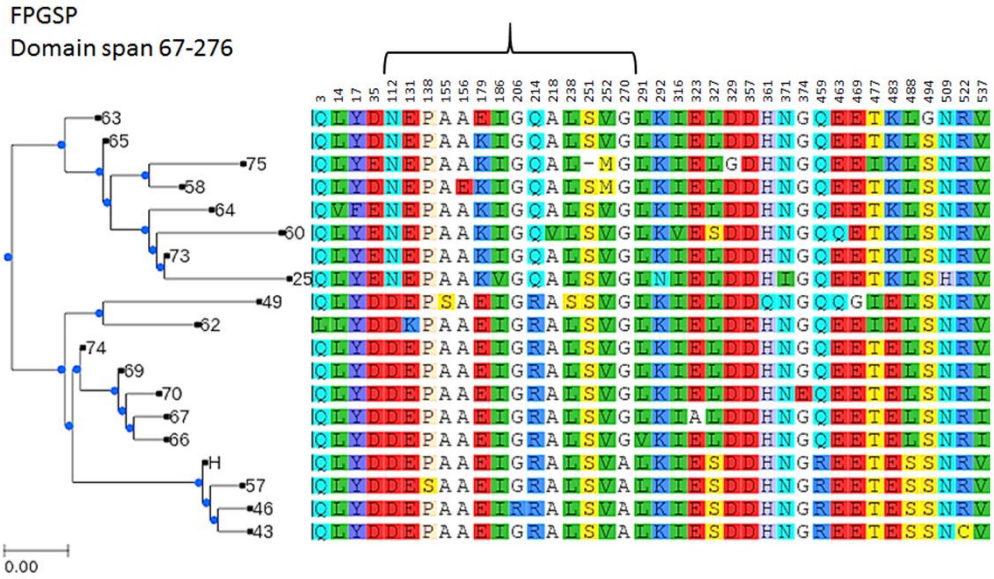


Figure 7.3.2. Protein similarity tree co-aligned with the amino acid positions in FPGSm protein which showed change in minimum one of the groups of tree. 'H' represents the Heinz 1706 (reference genome) group while numeral values represent the number of any of the accessions of that particular group. The accessions names corresponding to each digit can be found in Annexure 2, 2.1, 3 and 3.1 for 84 and 360 accession respectively. A=Changes observed in 84 accessions, B=Changes observed in 360 accessions.
Braces show the changed residue position which falls in the catalytic domain.

A



B

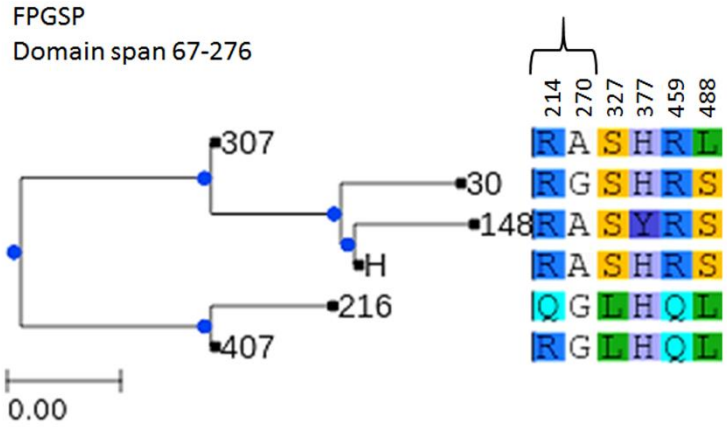


Figure 7.3.3. Protein similarity tree co-aligned with the amino acid positions in FPGSp protein which showed change in minimum one of the groups of tree. 'H' represents the Heinz 1706 (reference genome) group while numeral values represent the number of any of the accessions of that particular group. The accessions names corresponding to each digit can be found in Annexure 2, 2.1, 3 and 3.1 for 84 and 360 accession respectively. A=Changes observed in 84 accessions, B=Changes observed in 360 accessions.
Braces show the changed residue position which falls in the catalytic domain.

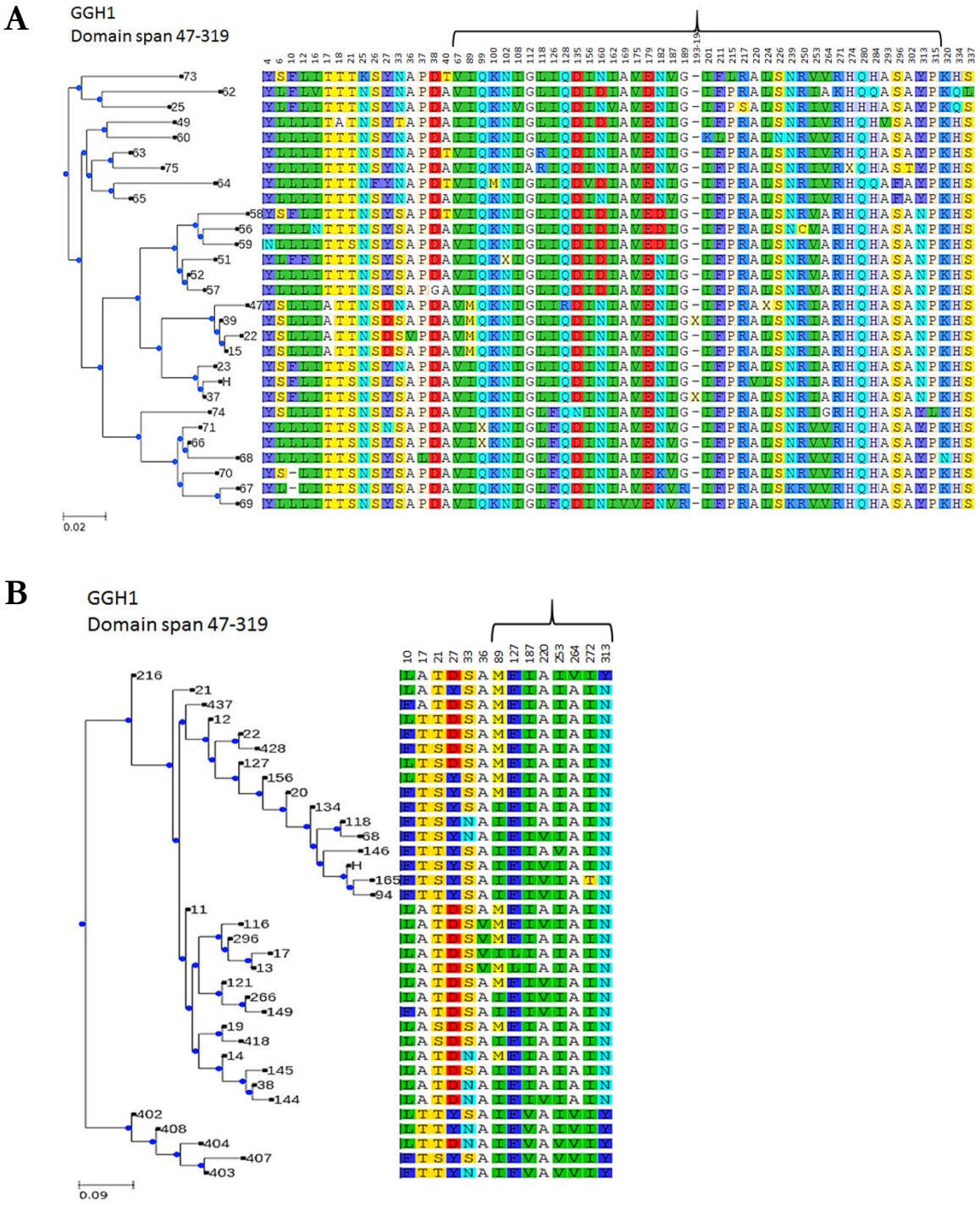


Figure 7.3.4. Protein similarity tree co-aligned with the amino acid positions in GGH1 protein which showed change in minimum one of the groups of tree. 'H' represents the Heinz 1706 (reference genome) group while numeral values represent the number of any of the accessions of that particular group. The accessions names corresponding to each digit can be found in Annexure 2, 2.1, 3 and 3.1 for 84 and 360 accession respectively. A=Changes observed in 84 accessions, B=Changes observed in 360 accessions.
Braces show the changed residue position which falls in the catalytic domain.

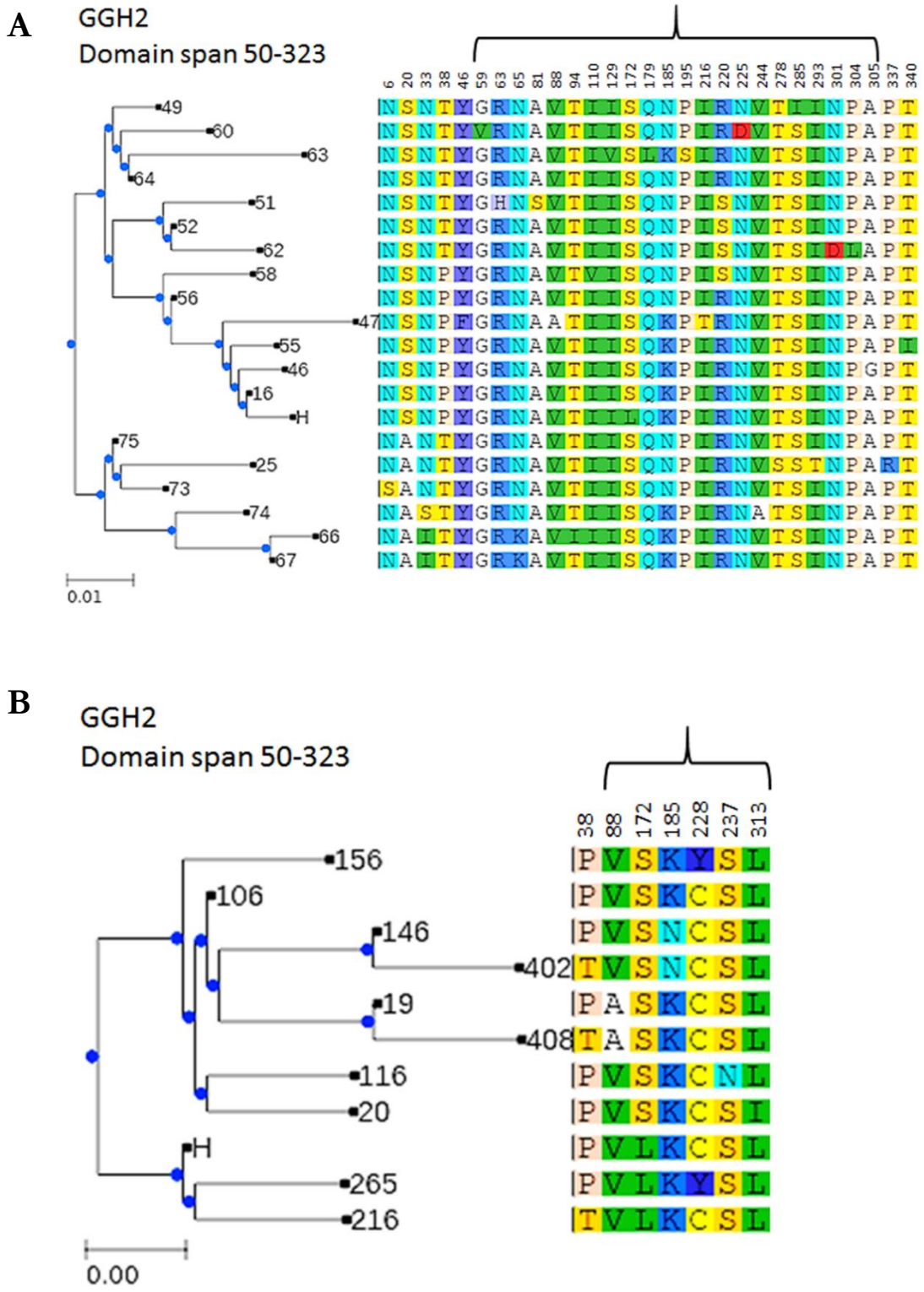


Figure 7.3.5. Protein similarity tree co-aligned with the amino acid positions in GGH2 protein which showed change in minimum one of the groups of tree. 'H' represents the Heinz 1706 (reference genome) group while numeral values represent the number of any of the accessions of that particular group. The accessions names corresponding to each digit can be found in Annexure 2, 2.1, 3 and 3.1 for 84 and 360 accession respectively.
A=Changes observed in 84 accessions, B=Changes observed in 360 accessions.
Braces show the changed residue position which falls in the catalytic domain.

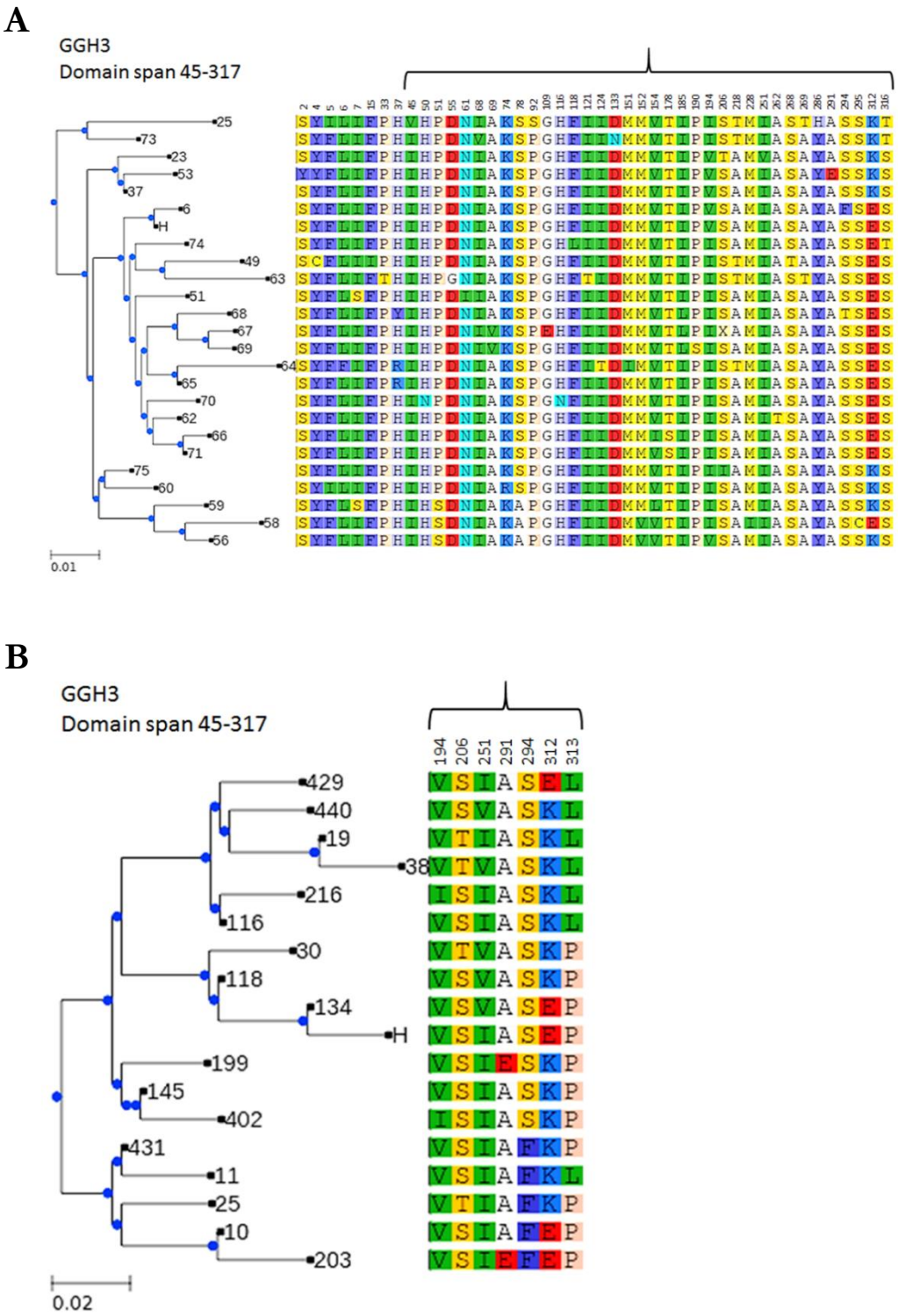


Figure 7.3.6. Protein similarity tree co-aligned with the amino acid positions in GGH3 protein which showed change in minimum one of the groups of tree. 'H' represents the Heinz 1706 (reference genome) group while numeral values represent the number of any of the accessions of that particular group. The accessions names corresponding to each digit can be found in Annexure 2, 2.1, 3 and 3.1 for 84 and 360 accession respectively. A=Changes observed in 84 accessions, B=Changes observed in 360 accessions.
Braces show the changed residue position which falls in the catalytic domain.

Table 7.3.2. SNP output of wild relatives in 84 tomato accessions

Gene	*S. pimpinellifolium*	*S. cheesmaniae*	*S. pennellii*	*S. habrochaites*
DHFS	1 group, no stop codon	1 group, no stop codon	2 groups, no stop codon	5 groups, no stop codon
FPGSm	3 groups, stop codon at 14th amino acid in 2 groups	3 groups, stop codon at 14th amino acid in 2 groups	2 groups, no stop codon	6 groups, no stop codon
FPGSp	3 groups, no stop codon	1 group, no stop codon	2 groups, no stop codon	4 groups, no stop codon
GGH1	4 groups, no stop codon	3 groups, no stop codon	2 groups, no stop codon	6 groups, no stop codon
GGH2	4 groups, no stop codon	3 groups, no stop codon	2 groups, no stop codon	2 groups, no stop codon
GGH3	1 group, no stop codon	3 groups, no stop codon	2 groups, no stop codon	6 groups, no stop codon

Table 7.3.3. SNP output of wild relatives in 360 tomato accessions

Gene	*S. pimpinellifolium*	*S. cheesmaniae*	*S. habrochaites*
DHFS	3 groups, no stop codon	1 group, no stop codon	1 group, no stop codon
FPGSm	13 groups, no stop codon	2 groups, no stop codon	1 group, no stop codon
FPGSp	2 groups, no stop codon	1 group, no stop codon	1 group, no stop codon
GGH1	16 groups, no stop codon	1 group, no stop codon	1 group, no stop codon
GGH2	7 groups, no stop codon	2 groups, no stop codon	1 group, no stop codon
GGH3	7 groups, no stop codon	1 group, no stop codon	1 group, no stop codon

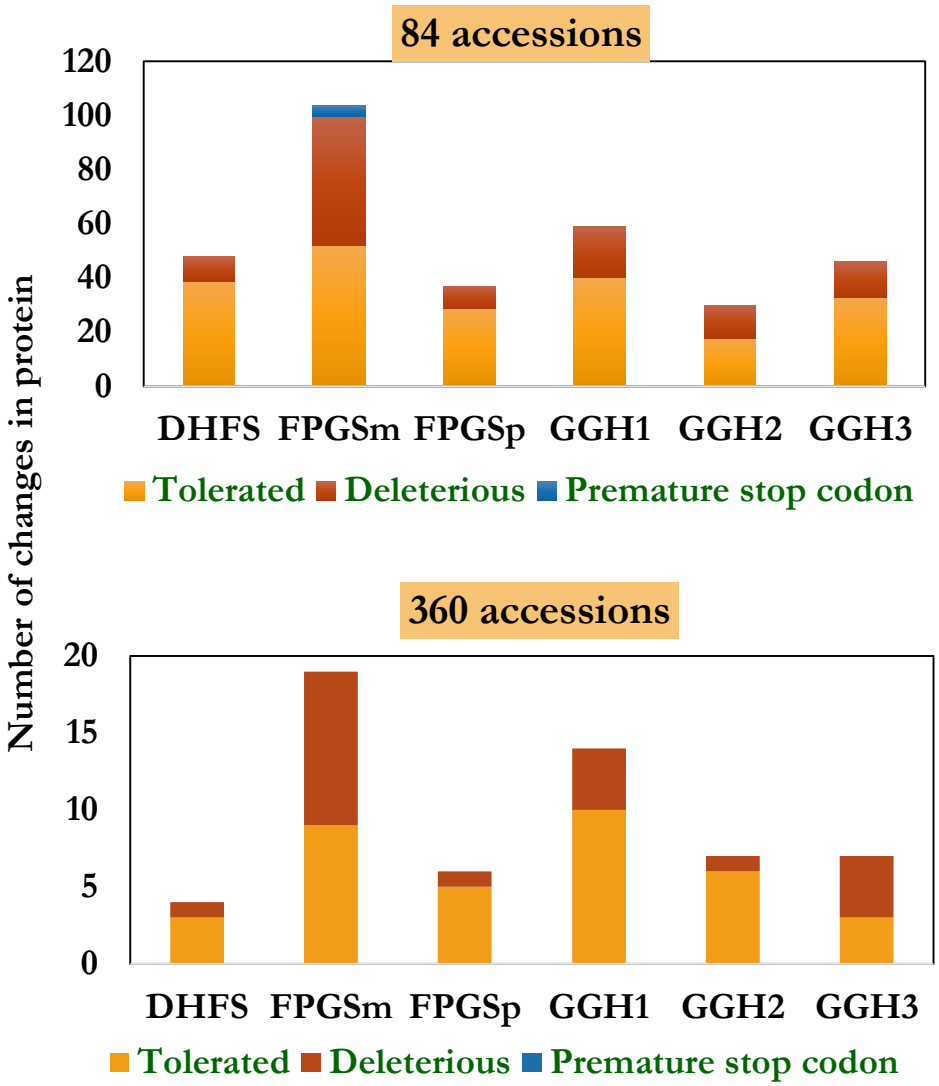


Figure 7.3.7. Effect of comparison of tolerated, deleterious and truncation causing non-synonymous SNPs in 444 tomato accessions (*S. lycopersicum* and wild species) in *DHFS*, *FPGSm*, *FPGSp*, *GGH1*, *GGH2* and *GGH3* genes.

polymorphism in the population. In addition the 360-accession study included 166 accessions with big fruits which further narrowed the genetic diversity.

The above analysis revealed that FPGSm followed by GGH1 showed maximum number of changes in amino acid residues. Though not comparable it is interesting to note that these results are broadly similar to our results on EcoTILLING of these two genes where too maximum number of haplotypes were recorded in *FPGSm* and *GGH1* genes (Table 7.1.1). The occurrence of this diversity probably stems from presence of the multiple copy of these genes where any loss of function in these genes can be compensated by *FPGSp*, *GGH2* and *GGH3* genes.

This is an interesting observation since *FPGSp* and *GGH2* respectively are able to take over the functions of above two genes thus maintaining the critical balance of folate mediated reactions in the cell. However, the genetic elasticity exhibited by *FPGSm* and *GGH1* in terms of polymorphism could not just be their dispensable status due to redundancy of their isoforms and might be of some significance that needs more detailed examinations.

7.3.2 Diversity of non-synonymous SNPs detected in four wild relatives of tomato

After examining the number of non-synonymous polymorphisms in all 444 accessions, the polymorphisms of 4 wild relatives of tomato namely *S. pimpinellifolium*, *S. cheesmaniae*, *S. pennellii* and *S. habrochaites* as representatives of red, orange, green and green fruited wild varieties of tomato was analyzed. The *S. pennellii* was selected because we earlier used *S. pennellii* introgression lines in M82 cultivar for folate profiling in this study. As the selection of accessions indicates; 84 genomes had more accessions of *S. cheesmaniae*, *S. pennellii* and *S. habrochaites*, except *S. pimpinellifolium* which was represented by 53 accessions in the 360 genomes as opposed to 4 accessions in 84 genomes. This is reflected in the protein similarity trees also where the number of groups corresponded well with the number of accessions sequenced for each wild relative. Surprisingly, only two accessions of *S. pimpinellifolium* and *S. cheesmaniae* each showed the gain of stop codon in FPGSm which again showed maximum range of protein dissimilarity groups resulting due to the non-synonymous SNPs (Table 7.3.2, 7.3.2). Figure 7.3.7 shows the ratio of deleterious to tolerated (D:T) SNPs which again establishes the highly elastic nature of *FPGSm* and *GGH1* where D:T was 1:1 and 1:2 respectively.

Taken together, these analyses can be used for better understanding of genetic regulation of folate metabolism by characterization of accessions showing stop codons or

amino acid changes that can change the function of proteins. These results provide promising information for assessment of protein function and related folate turnover.

However the filtering of synonymous SNPs and also restricting to those SNPs that are present only in coding sequences underrepresent the overall diversity of genic polymorphism. At the same time application of above filters allowed examination of only those changes that may have affected the protein function. Nevertheless, it should be noted that even synonymous changes as well as intronic SNPs can affect the protein function, however the probability of these effects is much lower than the nonsynonymous SNPs.

Chapter 8

Summary

Summary

Vitamins have a unique and indispensable role in metabolism acting as cofactors in biochemical reactions along with the enzymes carrying out these reactions. Folates (B_9) are important vitamin of B complex group. Major role of folates appears in metabolic reactions requiring single carbon (C1) chemical groups like methyl, methylene, methenyl, and formyl etc. Folates exist in their reduced tetrahydrofolate (THF) derivative forms and donate or receive these groups to the corresponding enzymes (Cossins, 2000). The reactions requiring folates are primary and critical for any cell, as they encompass biosynthesis of purine, methionine, serine and glycine (Blancquaert et al., 2010); the processes associated with the cell division. The folate synthesis is restricted to bacteria and plants and other organisms obtain folate through dietary sources. Folate deficiency, a major worldwide problem has profound health effects on the deficient individual and on the developing fetus in case of expecting mothers. Folate deficiency can occur because of different eating habits (diets mainly consisting of starchy and one type of food), socio-economic status (inability to purchase fruits, vegetables and health supplements) and ongoing health condition of the individuals (alcohol consumption and liver diseases).

The recommended dietary allowance (RDA) for folate is 400 and 600 µg per day for normal adults and pregnant women (https://ods.od.nih.gov/factsheets/Folate-HealthProfessional/#h10). Folates are enriched in spinach, green leafy vegetables, pulses, beans, lentils, broccoli, asparagus, citrus fruits, liver based delicacies and fish meat (Kader et al., 2004; Delchier, et al, 2016). However, the bulk of population obtains folate from corn, wheat, rice and cassava based foods which contain very low amount of folate per serving. Many people do not regularly consume green vegetables due to their unavailability throughout the year and folate rich food items are more expensive. This results in chronic deficiency condition that depletes the liver folate store (which can fulfill the folate requirement of entire body for 4-5 months through enterohepatic circulation) (Hercberg and Galan, 1992). Common folate deficiency symptoms in humans and animals are neural tube defects (anencephaly, *Spina bifida*), cardiovascular diseases, cognitive dysfunction and increased risk of colorectal cancer.

To tackle folate deficiency, USA and Canada have mandated food fortification with folic acid (synthetic forms of folate) since 1980s. The alternate approaches to ensure proper folate intake is by oral supplementation of folic acid pills and through usage of folate rich dietary sources. These two approaches have their own limitations and side effects as food

fortification programs and folic acid supplementation can result in excess intake of folate. Therefore, folate intake using folate-enriched food sources is the harmless and suitable option.

A number of studies measuring the folate levels in the items constituting regular diet of different regions have cataloged the folate rich and folate deficient items (Pfeiffer et al., 1997; Ruggeri et al., 1999; Ndaw et al., 2001; Ringling and Rychlik, 2013; Iwatani et al., 2003). These studies are important because a high folate accession of common crops can be directly released into the market and can be further exploited for breeding purposes to make them suitable for certain geographical environments. At the same time it can provide adequate amount of folate in regular daily servings.

Though, there are several naturally folate rich plant and animal food sources, however, there is a need for improving folate levels in the fruits and vegetables which are cheaper, grown universally and accessible to every strata of society. Tomato, in India is the second largest crop both production and consumption wise and therefore an apt target for fortification efforts. To achieve folate enhancement, either genetic engineering approaches can be used or a large number of plant products can be screened for their folate content. The former approach which mostly utilized the overexpression of folate pathway initial genes (*GCHI=GTP cyclohydrolaseI* and *ADCS=aminodeoxychorismate synthase*) (Hossain et al., 2004; Diaz de la Garza et al., 2004; Diaz de la Garza et al., 2007; Storozhenko et al., 2007; Nunes et al., 2009) has substantially improved the folate levels. However, genetic engineering approaches have two major constraints- (1) it results in genetically modified (GM) plants which are still not accepted by the society and (2) the similar overexpression approaches were not successful for other crops except rice, tomato, Mexican common bean and to some extent in lettuce.

Therefore, in this study we attempted to search for folate enriched tomato accessions through folate screening. Two methods of folate screening were used, namely microbiological assay (MA) and LC-MS. MA estimation of total folate content among 113 tomato accessions ranged from 5-60 μg/100 g FW with reference cultivar Arka Vikas folate content being 40 μg/100 g FW. Most accessions showed folate content ranging from 20-40 μg/100 g FW which constituted 51% of all the accessions screened while 36% accessions showed total folate ranging from 5-20 μg/100 g FW. Only 13% accessions exhibited more than 40 μg/100 g FW of total folate. Similar variations in folate levels were reported by Akilanathan et al. (2010) where they found 5 fold differences between different

cultivars of tomato, grape, guava and banana. These variations in folate level could be either cultivar based or because of human errors during the sample harvesting in picking up the uniform stage of fruit or during the assay itself. Our analysis showed that though, MA is a cheap and high throughput method, it is prone to inaccurate estimation. It also gives higher values than the chromatography based methods (Phillips et al., 2010). However, it gives a preliminary idea about the folate content in a large population which can be validated by other methods of estimations.

In subsequent experiments folate was estimated using LC-MS in tomato samples in next season. Total folate including tetrahydrofolate, 5-methyl tetrahydrofolate, 5-formyl tetrahydrofolate and 5,10-methenyl tetrahydrofolate was estimated in RR and mature green (MG) fruits samples from 52 and 27 corresponding accessions respectively. The folate content ranged from 9-42 μg/100 g FW and 18-58 μg/100 g FW in RR and MG stage respectively. The most abundant folate derivative was 5-methyl tetrahydrofolate which constituted 74-94% and 59-88% of total folate in MG and RR tomato fruits which is in congruence with the previously established distribution of folate forms (Rébeillé et al., 2006). Second most abundant folate detected in both MG and RR tomato fruits was 5-Formyl tetrahydrofolate. In literature, tomato folate content is reported to be 9-29 μg/100 g FW and 4.1-35.3 μg/100 g FW (Bekaert et al., 2007; Iniesta et al., 2009). Our results showed a broader spectrum of folate content in tomato which might be because of more number of accessions used in this study. We also observed that folate content in tomato either remained unaltered or decreased with the ripening process. Similar patterns have been found by many other groups and there is no tangible pattern established so far between folate content and ripening (Iniesta et al., 2009; Strålsjö et al., 2003). Sel.14 showed 2.4 fold more folate than AV and is a promising line.

The underlying reasons causing differential accumulation of folate were investigated in one high (Sel.14) and one low (EC27910) folate accessions along with AV. The characterization studies included estimation of folate forms, total pABA and pterin (main precursors), and gene expression of folate biosynthesis and turnover genes. However, no clear pattern was observed between the precursor or transcript abundance and folate content and analyses such as enzyme assays and metabolite profiling might be helpful in understanding the reasons of high or low folate accumulation.

Parallel to this, tomato wild relative *S. pennellii* introgression lines were studied to identify the folate QTLs and regulatory bins. Several high and low folate QTLs were

identified. IL 12-2 showed high folate content while other 12th chromosome ILs showed low folate character which significantly delimited the size of the folate QTL and is a promising result which will be pursued further. The low folate character was observed in many ILs of different chromosomes and therefore regulatory bins for low folate trait appeared to be distributed among many chromosomes. It is likely that the regulatory bins are present on many chromosomes as folate plays critical role in the metabolism. The folate biosynthesis gene expression data did not correlate with the folate content of ILs showing significantly higher or lower folate than the parent cultivar M82. These results corroborated with the results obtained in expression profiling of high and low folate natural accessions of tomato (Sel.14 and EC27910), indicating that the transcript abundance was not directly linked with the folate accumulation and there could be post-transcriptional changes regulating the folate content in fruits.

The genic variations of folate turnover genes (causing Polyglutamylation and deglutamylation of folate derivatives), *folylpolyglutamate synthases (FPGS)* and *γ-glutamyl hydrolases (GGH)* were studied. The naturally occurring genetic polymorphism was scored among 391 accessions using EcoTILLING. Simultaneously, an ethyl methane sulphonate (EMS) mutagenized tomato cv. Arka Vikas population was also screened for mutations using TILLING. The genes selected for both the approaches were *folylpolyglutamate synthase* (*FPGS*) and *γ-glutamyl hydrolase* (*GGH*) isoforms. *FPGS* has two isoforms in tomato, plastidial (*FPGSp*) and mitochondrial (*FPGSm*) located on 5th and 4th chromosomes respectively. Similarly, *GGH* isoforms *GGH1*, *GGH2* and *GGH3* are present on 7th, 10th and 7th chromosomes respectively. Since the basic methodology for both EcoTILLING and TILLING is same except for population and DNA pooling, common primer sets were designed for all the genes using PRIMER3 software for the CODDLE (Codons Optimized to Discover Deleterious Lesions) predicted gene region of approximately 1000 bp length.

EcoTILLING in *FPGSm* and *FPGSp* genes showed multiple SNPs in many accessions. Accessions showing similar SNPs were classified into one haplotype (HT) and thus *FPGSm* and *FPGSp* constituted six and four haplotypes respectively where HT 1 represented reference cultivar and all those accessions which did not have any SNP in the screened region for every gene. None of the sequenced representatives of these haplotypes showed exonic SNPs. EC27910 which showed an intronic SNP in *FPGSp* also showed low folate and was further pursued. Since fruits of these accessions were not available in three or more replicates, therefore no correlation between SNPs and folate levels could be inferred. Similarly, *GGH1, GGH2* and *GGH3* SNPs were grouped in 6, 3 and 4 haplotypes

including the HT 1. All the haplotypes showed multiple SNPs out of which, few were present in exon (WIR-3928 in *GGH1*; EC362949 in *GGH2* and EC520046 and EC1087 in *GGH3*). However, these plants were not examined in following generations due to uncertainty about correlation between folate content and the scored SNP. The prevalence of intronic SNPs indicated that deleterious SNPs are removed from the population as a result of purifying selection therefore no SNPs were detected that led to changes in the protein coding region.

LI-COR and NGS based TILLING was performed on 120 mM EMS re-mutagenized population which largely yielded intronic mutations in *GGH1, GGH2, FPGSm* and *FPGSp*. Only one mutation in *FPGSp* was detected in exon and was synonymous mutation. The absence of mutations in exons reflects that severe mutations in these genes might be lethal for the plant and therefore were eliminated from the population. In conclusion, the folate pathway genes were not amenable to the mutagenesis possibly due to their vital role in metabolism causing such rarity of exonic mutations.

It was speculated that the overall lack of exonic polymorphism and mutations in *FPGS* and *GGH* genes detected through EcoTILLING and TILLING approaches was due to the critical nature of these genes. However, this could be also due to the screening of only a small region of gene. Therefore, we also looked at the SNP data of resequencing of 84 and 360 tomato genomes. Only non-synonymous SNPs were considered and the resulting protein sequences were grouped and aligned on the basis of similarity. We also analyzed the deleterious or tolerated nature of these SNPs. Interestingly, *FPGSm* and *GGH1* isoforms showed more variations. Most of the variations were centered among the wild relatives which constituted a major percentage of 84 genome data whereas, less variability was obtained among the 360 genome data due to less representation of wild species in those accessions. Also, there were many unique, non-overlapping changes between the protein similarity tress of 84 and 360 accessions, indicating towards the tremendous variability existing among the cultivated varieties as well. Thus results of *in silico* analysis provided insights for selecting the accessions with drastic changes and study of its effect on total folate turnover.

In conclusion, we identified a folate enriched accession which can be used as a breeding parent as well as thorough characterization of this accession can unravel the factors behind its high folate trait. Further analyses like enzyme activity and metabolite profiling can be used for better understanding of the mechanisms. The regulatory bins

identified are valuable leads towards mapping of high folate trait to a smaller region and can help in revealing the governing factors. The results of *in silico* analysis can be used for selecting the accessions showing compromised gene functions of the turnover genes and studying its effects. Together, these results indicate that despite functional redundancy among the *FPGS* and *GGH* isoforms, each isoform is important and therefore cannot accumulate many deleterious changes. The accessions showing deleterious SNPs in the catalytic domain of the protein can be studied in detail that how these SNPs are affecting the enzyme activity and folate levels. The analysis of results obtained in this study along with the folate QTLs can reveal more about regulation of folate turnover.

References

Aflitos S, Schijlen E, Jong H, Ridder D, Smit S, Finkers R, Wang J, Zhang G, Li N, Mao L (2014) Exploring genetic variation in the tomato (Solanum section Lycopersicon) clade by whole-genome sequencing. The Plant Journal **80:** 136-148

Akhtar TA, McQuinn RP, Naponelli V, Gregory JF, Giovannoni JJ, Hanson AD (2008) Tomato γ-glutamylhydrolases: expression, characterization, and evidence for heterodimer formation. Plant physiology **148:** 775-785

Akhtar TA, Orsomando G, Mehrshahi P, Lara-Núñez A, Bennett MJ, Gregory III JF, Hanson AD (2010) A central role for gamma-glutamyl hydrolases in plant folate homeostasis. The Plant Journal **64:** 256-266

Akilanathan L, Vishnumohan S, Arcot J, Uthira L, Ramachandran S (2010) Total folate: diversity within fruit varieties commonly consumed in India. International journal of food sciences and nutrition **61:** 463-472

Alban C, Job D, Douce R (2000) Biotin metabolism in plants. Annual review of plant biology **51:** 17-47

Allen AM, Barker GL, Berry ST, Coghill JA, Gwilliam R, Kirby S, Robinson P, Brenchley RC, D'Amore R, McKenzie N (2011) Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (Triticum aestivum L.). Plant Biotechnology Journal **9:** 1086-1099

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

Ansari R, Mahta A, Mallack E, Luo JJ (2014) Hyperhomocysteinemia and neurologic disorders: a review. Journal of Clinical Neurology **10:** 281-288

Appling DR (1991) Compartmentation of folate-mediated one-carbon metabolism in eukaryotes. The FASEB journal **5:** 2645-2651

Arcot J, Shrestha A (2005) Folate: methods of analysis. Trends in Food Science & Technology **16:** 253-266

Arms EM, Bloom AJ, Clair DAS (2015) High-resolution mapping of a major effect QTL from wild tomato Solanum habrochaites that influences water relations under root chilling. Theoretical and Applied Genetics **128:** 1713-1724

Barkley N, Wang M (2008) Application of TILLING and EcoTILLING as reverse genetic approaches to elucidate the function of genes in plants and animals. Current genomics **9:** 212-226

Bartley GE, Scolnik PA (1995) Plant carotenoids: pigments for photoprotection, visual attraction, and human health. The Plant Cell **7:** 1027

Basset G, Quinlivan EP, Ziemak MJ, de la Garza RD, Fischer M, Schiffmann S, Bacher A, Gregory JF, Hanson AD (2002) Folate synthesis in plants: the first step of the pterin branch is mediated by a unique bimodular GTP cyclohydrolase I. Proceedings of the National Academy of Sciences **99:** 12489-12494

Basset GJ, Quinlivan EP, Gregory JF, Hanson AD (2005) Folate synthesis and metabolism in plants and prospects for biofortification. Crop Science **45:** 449-453

Basset GJ, Ravanel S, Quinlivan EP, White R, Giovannoni JJ, Rébeillé F, Nichols BP, Shinozaki K, Seki M, Gregory JF (2004) Folate synthesis in plants: the last step of the p-aminobenzoate branch is catalyzed by a plastidial

aminodeoxychorismate lyase. The Plant Journal **40:** 453-461

Baxter CJ, Carrari F, Bauke A, Overy S, Hill SA, Quick PW, Fernie AR, Sweetlove LJ (2005) Fruit carbohydrate metabolism in an introgression line of tomato with increased fruit soluble solids. Plant and Cell Physiology **46:** 425-437

Bedhomme M, Hoffmann M, McCarthy EA, Gambonnet B, Moran RG, Rébeillé F, Ravanel S (2005) Folate metabolism in plants an arabidopsis homolog of the mammalian mitochondrial folate transporter mediates folate import into chloroplasts. Journal of Biological Chemistry **280:** 34823-34831

Bekaert S, Storozhenko S, Mehrshahi P, Bennett MJ, Lambert W, Gregory JF, Schubert K, Hugenholtz J, Van Der Straeten D, Hanson AD (2007) Folate biofortification in food plants. Trends in plant science **13:** 28-35

Blancquaert D, Storozhenko S, Loizeau K, De Steur H, De Brouwer V, Viaene J, Ravanel S, Rébeillé F, Lambert W, Van Der Straeten D (2010) Folates and folic acid: from fundamental research toward sustainable health. Critical Reviews in Plant Science **29:** 14-35

Blancquaert D, Storozhenko S, Van Daele J, Stove C, Visser RG, Lambert W, Van Der Straeten D (2013) Enhancing pterin and para-aminobenzoate content is not sufficient to successfully biofortify potato tubers and Arabidopsis thaliana plants with folate. Journal of experimental botany **64:** 3899-3909

Bozzo GG, Basset GJ, Naponelli V, Noiriel A, Gregory JF, Hanson AD (2008) Characterization of the folate salvage enzyme p-aminobenzoylglutamate hydrolase in plants. Phytochemistry **69:** 29-37

Broun P (2004) Transcription factors as tools for metabolic engineering in plants. Current opinion in plant biology **7:** 202-209

Brown A (1992) Human impact on plant gene pools and sampling for their conservation. Oikos**:** 109-118

Butelli E, Titta L, Giorgio M, Mock H-P, Matros A, Peterek S, Schijlen EG, Hall RD, Bovy AG, Luo J (2008) Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nature biotechnology **26:** 1301-1308

Cakmak I, Pfeiffer WH, McClafferty B (2010) Review: biofortification of durum wheat with zinc and iron. Cereal Chemistry **87:** 10-20

Canene-Adams K, Campbell JK, Zaripheh S, Jeffery EH, Erdman JW (2005) The tomato as a functional food. The Journal of nutrition **135:** 1226-1230

Causse M, Duffe P, Gomez M, Buret M, Damidaux R, Zamir D, Gur A, Chevalier C, Lemaire-Chamley M, Rothan C (2004) A genetic map of candidate genes and QTLs involved in tomato fruit size and composition. Journal of experimental botany **55:** 1671-1685

Chan KC, Muschik GM, Issaq HJ, Garvey KJ, Generlette PL (1996) High-speed screening of polymerase chain reaction products by capillary electrophoresis. Analytical biochemistry **243:** 133-139

Chapman NH, Bonnet J, Grivet L, Lynn J, Graham N, Smith R, Sun G, Walley PG, Poole M, Causse M (2012) High-resolution mapping of a fruit firmness-related quantitative trait locus in tomato reveals epistatic interactions associated with a complex combinatorial locus. Plant Physiology **159:** 1644-1657

Chen L, Hao L, Parry MA, Phillips AL, Hu YG (2014) Progress in TILLING as a tool for functional genomics and improvement of crops. Journal of integrative plant biology **56:** 425-443

Cherest H, Thomas D, Surdin-Kerjan Y (2000) Polyglutamylation of Folate Coenzymes Is Necessary for Methionine Biosynthesis and Maintenance of Intact Mitochondrial Genome inSaccharomyces cerevisiae. Journal of Biological Chemistry **275:** 14056-14063

Chiang PK, Gordon RK, Tal J, Zeng G, Doctor B, Pardhasaradhi K, McCann PP (1996) S-Adenosylmethionine and methylation. The FASEB journal **10:** 471-480

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 physiology **126:** 480-484

Collakova E, Goyer A, Naponelli V, Krassovskaya I, Gregory JF, Hanson AD, Shachar-Hill Y (2008) Arabidopsis 10-formyl tetrahydrofolate deformylases are essential for photorespiration. The Plant Cell **20:** 1818-1832

Comai L, Young K, Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR (2004) Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. The Plant Journal **37:** 778-786

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

Cooper BA, Jonas E (1973) Superiority of simplified assay for folate with Lactobacillus casei ATCC 7469 over assay with chloramphenicol-adapted strain. Journal of clinical pathology **26:** 963-967

Cooper JL, Till BJ, Laport RG, Darlow MC, Kleffner JM, Jamai A, El-Mellouki T, Liu S, Ritchie R, Nielsen N (2008) TILLING to detect induced mutations in soybean. BMC plant biology **8:** 9

Cossins EA (2000) The fascinating world of folate and one-carbon metabolism. Botany **78:** 691

Cox K, Robertson D, Fites R (1999) Mapping and expression of a bifunctional thymidylate synthase, dihydrofolate reductase gene from maize. Plant molecular biology **41:** 733-739

Dabrowska W, Kazenko A, Laskowski M (1949) Concerning the Specificity of Chicken Pancreas Conjugase. Science (New York, NY) **110:** 95

De Brouwer V, Storozhenko S, Van De Steene JC, Wille SM, Stove CP, Van Der Straeten D, Lambert WE (2008) Optimisation and validation of a liquid chromatography–tandem mass spectrometry method for folates in rice. Journal of Chromatography A **1215:** 125-132

De Brouwer V, Zhang GF, Storozhenko S, Van Der Straeten D, Lambert WE (2007) pH stability of individual folates during critical sample preparation steps in prevision of the analysis of plant folates. Phytochemical analysis **18:** 496-508

de la Garza RD, Quinlivan EP, Klaus SM, Basset GJ, Gregory JF, Hanson AD (2004) Folate biofortification in tomatoes by engineering the pteridine branch of folate synthesis. Proceedings of the National Academy of Sciences of the United States of America **101:** 13720-13725

de La Garza RID, Gregory JF, Hanson AD (2007) Folate biofortification of tomato

fruit. Proceedings of the National Academy of Sciences **104:** 4218-4222

Delchier N, Herbig AL, Rychlik M, Renard CM (2016) Folates in Fruits and Vegetables: Contents, Processing, and Stability. Comprehensive Reviews in Food Science and Food Safety **15:** 506-528

Delchier N, Ringling C, Le Grandois J, Aoudé-Werner D, Galland R, Georgé S, Rychlik M, Renard CM (2013) Effects of industrial processing on folate content in green vegetables. Food chemistry **139:** 815-824

Diretto G, Al-Babili S, Tavazza R, Papacchioli V, Beyer P, Giuliano G (2007) Metabolic engineering of potato carotenoid content through tuber-specific overexpression of a bacterial mini-pathway. PLoS One **2:** e350

Drewke C, Leistner E (2001) Biosynthesis of vitamin B 6 and structurally related derivatives. Vitamins & Hormones **61:** 121-155

Elliott KJ, Butler WO, Dickinson CD, Konno Y, Vedvick TS, Fitzmaurice L, Mirkov TE (1993) Isolation and characterization of fruit vacuolar invertase genes from two tomato species and temporal differences in mRNA levels during fruit ripening. Plant molecular biology **21:** 515-524

Elsenhans B, Ahmad O, Rosenberg I (1984) Isolation and characterization of pteroylpolyglutamate hydrolase from rat intestinal mucosa. Journal of Biological Chemistry **259:** 6364-6368

Emmanuel E, Levy AA (2002) Tomato mutants as tools for functional genomics. Current opinion in plant biology **5:** 112-117

Endres M, Ahmadi M, Kruman I, Biniszkiewicz D, Meisel A, Gertz K (2005) Folate deficiency increases postischemic brain injury. Stroke **36:** 321-325

Eshed Y, Zamir D (1994a) A genomic library of Lycopersicon pennellii in L. esculentum: a tool for fine mapping of genes. Euphytica **79:** 175-179

Eshed Y, Zamir D (1994b) Introgressions from Lycopersicon pennellii can improve the soluble-solids yield of tomato hybrids. Theoretical and Applied Genetics **88:** 891-897

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

Eudes A, Kunji ER, Noiriel A, Klaus SM, Vickers TJ, Beverley SM, Gregory JF, Hanson AD (2010) Identification of transport-critical residues in a folate transporter from the folate-biopterin transporter (FBT) family. Journal of Biological Chemistry **285:** 2867-2875

Fajardo V, Alonso-Aperte E, Varela-Moreiras G (2015) Folate content in fresh-cut vegetable packed products by 96-well microtiter plate microbiological assay. Food chemistry **169:** 283-288

Fazili Z, Pfeiffer CM, Zhang M (2007) Comparison of serum folate species analyzed by LC-MS/MS with total folate measured by microbiologic assay and Bio-Rad radioassay. Clinical chemistry **53:** 781-784

Fernie AR, Tadmor Y, Zamir D (2006) Natural genetic variation for improving crop quality. Current opinion in plant biology **9:** 196-202

Fischer M, Bacher A (2006) Biosynthesis of vitamin B2 in plants. Physiologia Plantarum **126:** 304-318

Freisleben A, Schieberle P, Rychlik M (2002) Syntheses of labeled vitamers of folic acid to be used as internal standards in stable isotope dilution assays. Journal of agricultural and food chemistry **50:** 4760-4768

Frerichmann SL, Kirchhoff M, Müller AE, Scheidig AJ, Jung C, Kopisch-Obuch FJ (2013) EcoTILLING in Beta vulgaris reveals polymorphisms in the FLC-like gene BvFL1 that are associated with annuality and winter hardiness. BMC plant biology **13:** 1

Fridman E, Pleban T, Zamir D (2000) A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proceedings of the National Academy of Sciences **97:** 4718-4723

Friedman M (2002) Tomato glycoalkaloids: role in the plant and in the diet. Journal of agricultural and food chemistry **50:** 5751-5780

Friso S, Choi S-W (2005) Gene-nutrient interactions in one-carbon metabolism. Current drug metabolism **6:** 37-46

Fulop D, Ranjan A, Ofner I, Covington MF, Chitwood DH, West D, Ichihashi Y, Headland L, Zamir D, Maloof JN (2016) A new advanced backcross tomato population enables high resolution leaf QTL mapping and gene identification. bioRxiv**:** 040923

Gady AL, Hermans FW, Van de Wal MH, van Loo EN, Visser RG, Bachem CW (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

Galivan J, Ryan TJ, Chave K, Rhee M, Yao R, Yin D (2000) Glutamyl hydrolase: pharmacological role and enzymatic characterization. Pharmacology & therapeutics **85:** 207-215

Garritano S, Gemignani F, Voegele C, Nguyen-Dumont T, Le Calvez-Kelm F, De Silva D, Lesueur F, Landi S, Tavtigian SV (2009) Determining the effectiveness of High Resolution Melting analysis for SNP genotyping and mutation scanning at the TP53 locus. BMC genetics **10:** 1

Gilchrist EJ, Haughn GW, Ying CC, Otto SP, Zhuang J, Cheung D, Hamberger B, Aboutorabi F, Kalynyak T, Johnson L (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

Goyer A, Navarre DA (2007) Determination of folate concentrations in diverse potato germplasm using a trienzyme extraction and a microbiological assay. Journal of agricultural and food chemistry **55:** 3523-3528

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

Grossowicz N, Mandelbaum-Shavit F, Davidoff R, Aronovitch J (1962a) Microbiologic determination of folic acid derivatives in blood. Blood **20:** 609-616

Grossowicz N, Rachmilewitz M, Izak G, Zan S (1962b) Determination of folic acid metabolites in normal subjects and in patients with nutritional megaloblastic anemia. Experimental Biology and Medicine **109:** 770-773

Grossowicz N, Waxman S, Schreiber C (1981) Cryoprotected Lactobacillus casei: an approach to standardization of microbiological assay of folic acid in serum. Clinical chemistry **27:** 745-747

Groth M, Moissiard G, Wirtz M, Wang H, Garcia-Salinas C, Ramos-Parra PA, Bischof S, Feng S, Cokus SJ, John A (2016) MTHFD1 controls DNA methylation in Arabidopsis. Nature communications **7**

Gundry CN, Vandersteen JG, Reed GH, Pryor RJ, Chen J, Wittwer CT (2003) Amplicon melting analysis with labeled primers: a closed-tube method for differentiating homozygotes and heterozygotes. Clinical chemistry **49:** 396-406

Guo Y, Abernathy B, Zeng Y, Ozias-Akins P (2015) TILLING by sequencing to identify induced mutations in stress resistance genes of peanut (Arachis hypogaea). BMC genomics **16:** 1

Halsted CH (1989) The intestinal absorption of dietary folates in health and disease. Journal of the American College of Nutrition **8:** 650-658

Han J-Y, Tyler RT (2003) Determination of folate concentrations in pulses by a microbiological method employing trienzyme extraction. Journal of agricultural and food chemistry **51:** 5315-5318

Hanson AD, Gage DA, Shachar-Hill Y (2000) Plant one-carbon metabolism and its engineering. Trends in plant science **5:** 206-213

Hanson AD, Gregory III JF (2011) Folate biosynthesis, turnover, and transport in plants. Annual review of plant biology **62:** 105-125

Hanson AD, Roje S (2001) One-carbon metabolism in higher plants. Annual review of plant biology **52:** 119-137

Hawkes J, Villota R (1988) Folates in foods: reactivity, stability during processing, and nutritional implications. Critical Reviews in Food Science and Nutrition **28:** 439-538

Henderson GB, Huennekens F (1974) Transport of folate compounds into Lactobacillus casei. Archives of biochemistry and biophysics **164:** 722-728

Henderson GB, Zevely EM, Huennekens F (1976) Folate transport in Lactobacillus casei: solubilization and general properties of the binding protein. Biochemical and biophysical research communications **68:** 712-717

Henikoff S, Till BJ, Comai L (2004) TILLING. Traditional mutagenesis meets functional genomics. Plant physiology **135:** 630-636

Hercberg S, Galan P (1992) Nutritional anaemias. Baillière's clinical haematology **5:** 143-168

Holtan HE, Hake S (2003) Quantitative trait locus analysis of leaf dissection in tomato using Lycopersicon pennellii segmental introgression lines. Genetics **165:** 1541-1550

Horne DW, Krumdieck CL, Wagner C (1981) Properties of folic acid gamma-glutamyl hydrolase (conjugase) in rat bile and plasma. The Journal of nutrition **111:** 442-449

Horne DW, Patterson D (1988) Lactobacillus casei microbiological assay of folic acid derivatives in 96-well microtiter plates. Clinical chemistry **34:** 2357-2359

Hossain T, Rosenberg I, Selhub J, Kishore G, Beachy R, Schubert K (2004) Enhancement of folates in plants through metabolic engineering. Proceedings of the National Academy of Sciences of the United States of America **101:** 5158-5163

Howard DH (1956) The preservation of bacteria by freezing in glycerol broth. Journal of bacteriology **71:** 625

Hyun TH, Tamura T (2005) Trienzyme extraction in combination with microbiologic assay in food folate analysis: an updated review. Experimental Biology and Medicine **230:** 444-454

Imeson HC, Cossins EA (1991) Higher plant folylpolyglutamate synthetase. II. Some major catalytic properties of the enzyme from Pisum sativum L. Journal of plant physiology **138:** 483-488

Iniesta MD, Pérez-Conesa Do, García-Alonso J, Ros G, Periago MJs (2009) Folate content in tomato (Lycopersicon esculentum). Influence of cultivar, ripeness, year of harvest, and pasteurization and storage temperatures. Journal of agricultural and food chemistry **57:** 4739-4745

Irimia M, Roy SW (2014) Origin of spliceosomal introns and alternative splicing. Cold Spring Harbor perspectives in biology **6:** a016071

Iwatani Y, Arcot J, Shrestha AK (2003) Determination of folate contents in some Australian vegetables. Journal of Food Composition and Analysis **16:** 37-48

Iyer R, Tomar S (2009) Folate: a functional food constituent. Journal of food science **74:** R114-R122

Jha AB, Ashokkumar K, Diapari M, Ambrose SJ, Zhang H, Tar'an B, Bett KE, Vandenberg A, Warkentin TD, Purves RW (2015) Genetic diversity of folate profiles in seeds of common bean, lentil, chickpea and pea. Journal of Food Composition and Analysis **42:** 134-140

Jiang L, Liu Y, Sun H, Han Y, Li J, Li C, Guo W, Meng H, Li S, Fan Y (2013) The mitochondrial folylpolyglutamate synthetase gene is required for nitrogen utilization during early seedling development in Arabidopsis. Plant physiology **161:** 971-989

Just D, Garcia V, Fernandez L, Bres C, Mauxion J-P, Petit J, Jorly J, Assali J, Bournonville C, Ferrand C (2013) Micro-Tom mutants for functional analysis of target genes and discovery of new alleles in tomato. Plant Biotechnology **30:** 225-231

Kadaru SB, Yadav AS, Fjellstrom RG, Oard JH (2006) Alternative ecotilling protocol for rapid, cost-effective single-nucleotide polymorphism discovery and genotyping in rice (Oryza sativa L.). Plant Molecular Biology Reporter **24:** 3-22

Kader AA, Perkins-Veazie P, Lester GE (2004) Nutritional quality of fruits, nuts, and vegetables and their importance in human health. US Dept. Agric Handbook **66**

Katoh A, Hashimoto T (2004) Molecular biology of pyridine nucleotide and nicotine biosynthesis. Front Biosci **9:** 1577-1586

Kawakami K, Ooyama A, Ruszkiewicz A, Jin M, Watanabe G, Moore J, Oka T, Iacopetta B, Minamoto T (2008) Low expression of γ-glutamyl hydrolase mRNA in primary colorectal cancer with the CpG island methylator phenotype. British journal of cancer **98:** 1555-1561

Kazenko A, Laskowski M (1948) On the specificity of chicken pancreas conjugase (γ-glutamic acid carboxypeptidase). Journal of Biological Chemistry **173:** 217-221

Keagy P (1985) Folacin: Microbiological and animal assays. Methods of vitamin assay **4:** 445-463

Kharabian-Masouleh A, Waters DL, Reinke RF, Henry RJ (2011) Discovery of polymorphisms in starch-related genes in rice germplasm by amplification of pooled DNA and deeply parallel sequencing. Plant Biotechnology Journal **9:** 1074-1085

Kilambi HV, Kumar R, Sharma R, Sreelakshmi Y (2013) Chromoplast-specific carotenoid-associated protein appears to be important for enhanced accumulation of carotenoids in hp1 tomato fruits. Plant physiology **161:** 2085-2101

Klaus SM, Kunji ER, Bozzo GG, Noiriel A, de La Garza RD, Basset GJ, Ravanel S, Rébeillé F, Gregory JF, Hanson AD (2005) Higher plant plastids and cyanobacteria have folate carriers related to those of trypanosomatids. Journal of Biological Chemistry **280:** 38457-38463

Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson RK, Ding L (2009) VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics **25:** 2283-2285

Konings E (1998) A validated liquid chromatographic method for determining folates in vegetables, milk powder, liver, and flour. Journal of AOAC International **82:** 119-127

Kozlowski P, Krzyzosiak WJ (2001) Combined SSCP/duplex analysis by capillary electrophoresis for more efficient mutation detection. Nucleic acids research **29:** e71-e71

Kuypers AW, Willems PM, van der Schans MJ, Linssen PC, Wessels HM, de Bruijn CH, Everaerts FM, Mensink EJ (1993) Detection of point mutations in DNA using capillary electrophoresis in a polymer network. Journal of Chromatography B: Biomedical Sciences and Applications **621:** 149-156

Le H, Fung D, Trent R (1997) Applications of capillary electrophoresis in DNA mutation analysis of genetic disorders. Molecular Pathology **50:** 261

Lemley C, Yan S, Dole VS, Madhubala R, Cunningham ML, Beverley SM, Myler PJ, Stuart KD (1999) The Leishmania donovani LD1 locus gene ORFG encodes a biopterin transporter (BT1). Molecular and biochemical parasitology **104:** 93-105

Lerma-Ortiz C, Jeffryes JG, Cooper AJ, Niehaus TD, Thamm AM, Frelin O, Aunins T, Fiehn O, de Crécy-Lagard V, Henry CS (2016) 'Nothing of chemistry disappears in biology': the Top 30 damage-prone endogenous metabolites. Biochemical Society Transactions **44:** 961-971

Leung H, Wang H, Wu J, Naredo ME, Baraoidan M, Bordeos A, Madamba S, Carrillo G, Sangha J, Negussie Z (2005) Allelic and functional diversity of stress-tolerance genes in rice. Copyright International Rice Research Institute 2005**:** 73

Lewandowska I, Balińska M, Natorff R, Paszewski A (1996) Regulation of folate-dependent enzyme levels in Aspergillus nidulans: studies with regulatory mutants. Biochimica et Biophysica Acta (BBA)-General Subjects **1290:** 89-94

Li G-M, Presnell SR, Gu L (2003) Folate deficiency, mismatch repair-dependent apoptosis, and human disease. The Journal of nutritional biochemistry **14:** 568-575

Li L, Hill-Skinner S, Liu S, Beuchle D, Tang HM, Yeh CT, Nettleton D, Schnable PS (2015) The maize brown midrib4 (bm4) gene encodes a functional folylpolyglutamate synthase. The Plant Journal **81:** 493-504

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. The Plant Journal **27:** 235-242

Li-Li Z, Ying L, Song L, Cossins EA (1992) The polyglutamate nature of plant folates. Phytochemistry **31:** 2277-2282

Lin S, Rogiers S, Cossins EA (1993) γ-Glutamyl hydrolase from pea cotyledons. Phytochemistry **32:** 1109-1117

Lin T, Zhu G, Zhang J, Xu X, Yu Q, Zheng Z, Zhang Z, Lun Y, Li S, Wang X (2014) Genomic analyses provide insights into the history of tomato breeding. Nature genetics **46:** 1220-1226

Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M (2012) Comparison of next-generation sequencing systems. BioMed Research International **2012**

Liu W, Smith DI, Rechtzigel KJ, Thibodeau SN, James CD (1998) Denaturing high performance liquid chromatography (DHPLC) used in the detection of germline and somatic mutations. Nucleic acids research **26:** 1396-1400

Lochlainn SÓ, Amoah S, Graham NS, Alamer K, Rios JJ, Kurup S, Stoute A, Hammond JP, Østergaard L, King GJ (2011) High Resolution Melt (HRM) analysis is an efficient tool to genotype EMS mutants in complex crop genomes. Plant methods **7:** 1

Lucock M, Daskalakis I (2000) New perspectives on folate status: a differential role for the vitamin in cardiovascular disease, birth defects and. British journal ofBiomedical Science **57:** 2541260

Luo M, Piffanelli P, Rastelli L, Cella R (1993) Molecular cloning and analysis of a cDNA coding for the bifunctional dihydrofolate reductase-thymidylate synthase of Daucus carota. Plant molecular biology **22:** 427-435

Marroni F, Pinosio S, Di Centa E, Jurman I, Boerjan W, Felice N, Cattonaro F, Morgante M (2011) Large-scale detection of rare variants via pooled multiplexed next-generation sequencing: towards next-generation Ecotilling. The Plant Journal **67:** 736-745

Martin J, Landen Jr W, Soliman A, Eitenmiller R (1989) Application of a tri-enzyme extraction for total folate determination in foods. Journal-Association of Official Analytical Chemists **73:** 805-808

Matamoros MA, Loscos J, Coronado MJ, Ramos J, Sato S, Testillano PS, Tabata S, Becana M (2006) Biosynthesis of ascorbic acid in legume root nodules. Plant physiology **141:** 1068-1077

Matherly LH, Goldman ID (2003) Membrane transport of folates. Vitamins & Hormones **66:** 403-456

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

McGuire JJ, Hsieh P, Coward JK, Bertino JR (1980) Enzymatic synthesis of folylpolyglutamates. Characterization of the reaction and its products. Journal of Biological Chemistry **255:** 5776-5788

McIntyre CL, Jackson M, Cordeiro GM, Amouyal O, Hermann S, Aitken K, Eliott F, Henry R, Casu R, Bonnett G (2006) The identification and characterisation

of alleles of sucrose phosphate synthase gene family III in sugarcane. Molecular Breeding **18:** 39-50

Mehrshahi P, Gonzalez-Jorge S, Akhtar TA, Ward JL, Santoyo-Castelazo A, Marcus SE, Lara-Núñez A, Ravanel S, Hawkins ND, Beale MH (2010) Functional analysis of folate polyglutamylation and its essential role in plant metabolism and development. The Plant Journal **64:** 267-279

Mejlhede N, Kyjovska Z, Backes G, Burhenne K, Rasmussen SK, Jahoor A (2006) EcoTILLING for the identification of allelic variation in 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. The Plant Journal **38:** 861-872

Minoia S, Petrozza A, D'Onofrio O, Piron F, Mosca G, Sozio G, Cellini F, Bendahmane A, Carriero F (2010) A new mutant genetic resource for tomato crop improvement by TILLING technology. BMC research notes **3:** 1

Missirian V, Comai L, Filkov V (2011) Statistical mutation calling from sequenced overlapping DNA pools in TILLING experiments. BMC bioinformatics **12:** 1

Mohan V, Gupta S, Thomas S, Mickey H, Charakana C, Chauhan VS, Sharma K, Kumar R, Tyagi K, Sarma S (2016) Tomato Fruits Show Wide Phenomic Diversity but Fruit Developmental Genes Show Low Genomic Diversity. PLoS One **11:** e0152907

Moldt J, Pokorny R, Orth C, Linne U, Geisselbrecht Y, Marahiel MA, Essen L-O, Batschauer A (2009) Photoreduction of the folate cofactor in members of the photolyase family. Journal of Biological Chemistry **284:** 21670-21683

Morandini P (2013) Control limits for accumulation of plant metabolites: brute force is no substitute for understanding. Plant Biotechnology Journal **11:** 253-267

Mouillon J-M, Ravanel S, Douce R, Rébeillé F (2002) Folate synthesis in higher-plant mitochondria: coupling between the dihydropterin pyrophosphokinase and the dihydropteroate synthase activities. Biochemical Journal **363:** 313-319

Muir SR, Collins GJ, Robinson S, Hughes S, Bovy A, De Vos CR, van Tunen AJ, Verhoeyen ME (2001) Overexpression of petunia chalcone isomerase in tomato results in fruit containing increased levels of flavonols. Nature biotechnology **19:** 470-474

Naqvi S, Zhu C, Farre G, Ramessar K, Bassie L, Breitenbach J, Conesa DP, Ros G, Sandmann G, Capell T (2009) Transgenic multivitamin corn through biofortification of endosperm with three vitamins representing three distinct metabolic pathways. Proceedings of the National Academy of Sciences **106:** 7762-7767

Navarrete O, Van Daele J, Stove C, Lambert W, Van Der Straeten D, Storozhenko S (2012) A folate independent role for cytosolic HPPK/DHPS upon stress in Arabidopsis thaliana. Phytochemistry **73:** 23-33

Ndaw S, Bergaentzlé M, Aoudé-Werner D, Lahély S, Hasselmann C (2001) Determination of folates in foods by high-performance liquid chromatography with fluorescence detection after precolumn conversion to 5-methyltetrahydrofolates. Journal of Chromatography A **928:** 77-90

Negrao S, Almadanim C, Pires I, McNally K, Oliveira M (2011) Use of EcoTILLING to identify natural allelic variants of rice candidate genes involved in salinity tolerance. Plant Genetic Resources **9:** 300-304

Nelson BC, Pfeiffer CM, Margolis SA, Nelson CP (2004) Solid-phase extraction–electrospray ionization mass spectrometry for the quantification of folate in human plasma or serum. Analytical biochemistry **325:** 41-51

Newman EM, Tsai JF (1986) Microbiological analysis of 5-formyltetrahydrofolic acid and other folates using an automatic 96-well plate reader. Analytical biochemistry **154:** 509-515

Ng PC, Henikoff S (2006) Predicting the effects of amino acid substitutions on protein function. Annu. Rev. Genomics Hum. Genet. **7:** 61-80

Nielsen R, Paul JS, Albrechtsen A, Song YS (2011) Genotype and SNP calling from next-generation sequencing data. Nature Reviews Genetics **12:** 443-451

Nieto C, Piron F, Dalmais M, Marco CF, Moriones E, Gómez-Guillamón ML, Truniger V, Gómez P, Garcia-Mas J, Aranda MA (2007) EcoTILLING for the identification of allelic variants of melon eIF4E, a factor that controls virus susceptibility. BMC plant biology **7:** 1

Noiriel A, Naponelli V, Bozzo GG, Gregory JF, Hanson AD (2007) Folate salvage in plants: pterin aldehyde reduction is mediated by multiple non-specific aldehyde reductases. The Plant Journal **51:** 378-389

Nunes AC, Kalkmann DC, Aragao FJ (2009) Folate biofortification of lettuce by expression of a codon optimized chicken GTP cyclohydrolase I gene. Transgenic research **18:** 661-667

O'broin S, Kelleher B (1992) Microbiological assay on microtitre plates of folate in serum and red cells. Journal of clinical pathology **45:** 344-347

Ofner I, Lashbrooke J, Pleban T, Aharoni A, Zamir D (2016) Solanum pennellii backcross inbred lines (BILs) link small genomic bins with tomato traits. The Plant Journal

Okabe Y, Ariizumi T (2016) Mutant Resources and TILLING Platforms in Tomato Research. *In* Functional Genomics and Biotechnology in Solanaceae and Cucurbitaceae Crops. Springer, pp 75-91

Okabe Y, Ariizumi T, Ezura H (2013) Updating the micro-tom TILLING platform. Breeding science **63:** 42-48

Okabe Y, Asamizu E, Saito T, Matsukura C, Ariizumi T, Brès C, Rothan C, Mizoguchi T, Ezura H (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:** 1994-2005

Okinaka O, Iwai K (1970) biosynthesis of folic acid compounds in plants. III. DISTRIBUTION of the dihydropteroate-synthesizing enzyme in plants. Journal of vitaminology

Oleykowski CA, Mullins CRB, Godwin AK, Yeung AT (1998) Mutation detection using a novel plant endonuclease. Nucleic acids research **26:** 4597-4602

Orsomando G, Bozzo GG, Garza RD, Basset GJ, Quinlivan EP, Naponelli V, Rébeillé F, Ravanel S, Gregory JF, Hanson AD (2006) Evidence for folate-

salvage reactions in plants. The Plant Journal **46:** 426-435

Orsomando G, de la Garza RD, Green BJ, Peng M, Rea PA, Ryan TJ, Gregory JF, Hanson AD (2005) Plant γ-Glutamyl Hydrolases and Folate Polyglutamates CHARACTERIZATION, COMPARTMENTATION, AND CO-OCCURRENCE IN VACUOLES. Journal of Biological Chemistry **280:** 28877-28884

Osterhues A, Ali NS, Michels KB (2013) The role of folic acid fortification in neural tube defects: a review. Critical Reviews in Food Science and Nutrition **53:** 1180-1190

Pandrangi S, LaBorde LF (2004) Optimization of microbiological assay of folic acid and determination of folate content in spinach. International journal of food science & technology **39:** 525-532

Papakostas GI, Cassiello CF, Iovieno N (2012) Folates and S-adenosylmethionine for major depressive disorder. The Canadian Journal of Psychiatry **57:** 406-413

Pawlosky RJ, Flanagan VP, Doherty RF (2003) A mass spectrometric validated high-performance liquid chromatography procedure for the determination of folates in foods. Journal of agricultural and food chemistry **51:** 3726-3730

Perchiniak E, Lawrence SA, Kasten S, Woodard BA, Taylor SM, Moran RG (2007) Probing the mechanism of the hamster mitochondrial folate transporter by mutagenesis and homology modeling. Biochemistry **46:** 1557-1567

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

Pfeiffer CM, Rogers LM, Gregory JF (1997) Determination of folate in cereal-grain food products using trienzyme extraction and combined affinity and reversed-phase liquid chromatography. Journal of agricultural and food chemistry **45:** 407-413

Phillips KM, Ruggio DM, Ashraf-Khorassani M, Eitenmiller RR, Cho S, Lemar LE, Perry CR, Pehrsson PR, Holden JM (2010) Folic acid content of ready-to-eat cereals determined by liquid chromatography-mass spectrometry: comparison to product label and to values determined by microbiological assay. Cereal Chemistry **87:** 42-49

Piironen V, Edelmann M, Kariluoto S, Bedő Z (2008) Folate in wheat genotypes in the HEALTHGRAIN diversity screen. Journal of agricultural and food chemistry **56:** 9726-9731

Piron F, Nicolaï M, Minoïa S, Piednoir E, Moretti A, Salgues A, Zamir D, Caranta C, Bendahmane A (2010) An induced mutation in tomato eIF4E leads to immunity to two potyviruses. PLoS One **5:** e11313

Quadrana L, Almeida J, Asís R, Duffy T, Dominguez PG, Bermúdez L, Conti G, da Silva JVC, Peralta IE, Colot V (2014) Natural occurring epialleles determine vitamin E accumulation in tomato fruits. Nature communications **5**

Quinlivan EP, Gregory JF (2003) Effect of food fortification on folic acid intake in the United States. The American journal of clinical nutrition **77:** 221-225

Quinlivan EP, Hanson AD, Gregory JF (2006) The analysis of folate and its metabolic precursors in biological samples. Analytical biochemistry **348:** 163-184

Quinlivan EP, Roje S, Basset G, Shachar-Hill Y, Gregory JF, Hanson AD (2003)

The folate precursor p-aminobenzoate is reversibly converted to its glucose ester in the plant cytosol. Journal of Biological Chemistry **278:** 20731-20737

Rader JI, Weaver CM, Angyal G (1998) Use of a microbiological assay with tri-enzyme extraction for measurement of pre-fortification levels of folates in enriched cereal-grain products. Food chemistry **62:** 451-465

Raghavan C, Naredo MEB, Wang H, Atienza G, Liu B, Qiu F, McNally KL, Leung H (2007) Rapid method for detecting SNPs on agarose gels and its application in candidate gene mapping. Molecular Breeding **19:** 87-101

Rahman T, Chowdhury MMH, Islam MT, Akhtaruzzaman M (2015) Microbiological Assay of Folic Acid Content in Some Selected Bangladeshi Food Stuffs. International Journal of Biology **7:** 35

Raichaudhuri A, Peng M, Naponelli V, Chen S, Sánchez-Fernández R, Gu H, Gregory JF, Hanson AD, Rea PA (2009) Plant vacuolar ATP-binding cassette transporters that translocate folates and antifolates in vitro and contribute to antifolate tolerance in vivo. Journal of Biological Chemistry **284:** 8449-8460

Raiola A, Tenore GC, Barone A, Frusciante L, Rigano MM (2015) Vitamin E Content and Composition in Tomato Fruits: Beneficial Roles and Bio-Fortification. International journal of molecular sciences **16:** 29250-29264

Ramírez Rivera NG, García-Salinas C, Aragão FJ, Díaz de la Garza RI (2016) Metabolic engineering of folate and its precursors in Mexican common bean (Phaseolus vulgaris L.). Plant Biotechnology Journal

Ramos-Parra PA, García-Salinas C, Hernández-Brenes C, Díaz de la Garza RoI (2013) Folate levels and polyglutamylation profiles of papaya (Carica papaya cv. Maradol) during fruit development and ripening. Journal of agricultural and food chemistry **61:** 3949-3956

Ravanel S, Cherest H, Jabrin S, Grunwald D, Surdin-Kerjan Y, Douce R, Rébeillé F (2001) Tetrahydrofolate biosynthesis in plants: molecular and functional characterization of dihydrofolate synthetase and three isoforms of folylpolyglutamate synthetase in Arabidopsis thaliana. Proceedings of the National Academy of Sciences **98:** 15360-15365

Ravanel S, Gakière B, Job D, Douce R (1998) The specific features of methionine biosynthesis and metabolism in plants. Proceedings of the National Academy of Sciences **95:** 7805-7812

Rébeillé F, Macherel D, Mouillon JM, Garin J, Douce R (1997) Folate biosynthesis in higher plants: purification and molecular cloning of a bifunctional 6-hydroxymethyl-7, 8-dihydropterin pyrophosphokinase/7, 8-dihydropteroate synthase localized in mitochondria. The EMBO journal **16:** 947-957

Rébeillé F, Ravanel S, Jabrin S, Douce R, Storozhenko S, Van Der Straeten D (2006) Folates in plants: biosynthesis, distribution, and enhancement. Physiologia Plantarum **126:** 330-342

Reis-Filho JS (2009) Next-generation sequencing. Breast Cancer Research **11:** 1

Reyes-Hernández BJ, Srivastava AC, Ugartechea-Chirino Y, Shishkova S, Ramos-Parra PA, Lira-Ruan V, Díaz de la Garza RI, Dong G, Moon JC, Blancaflor EB (2014) The root indeterminacy-to-determinacy developmental switch is

operated through a folate-dependent pathway in Arabidopsis thaliana. New Phytologist **202:** 1223-1236

Rick CM, Uhlig JW, Jones AD (1994) High alpha-tomatine content in ripe fruit of Andean Lycopersicon esculentum var. cerasiforme: developmental and genetic aspects. Proceedings of the National Academy of Sciences **91:** 12877-12881

Rigola D, van Oeveren J, Janssen A, Bonné A, Schneiders H, van der Poel HJ, van Orsouw NJ, Hogers RC, de Both MT, van Eijk MJ (2009) High-throughput detection of induced mutations and natural variation using KeyPoint™ technology. PLoS One **4:** e4761

Ringling C, Rychlik M (2013) Analysis of seven folates in food by LC–MS/MS to improve accuracy of total folate data. European Food Research and Technology **236:** 17-28

Ron M, Dorrity MW, de Lucas M, Toal T, Hernandez RI, Little SA, Maloof JN, Kliebenstein DJ, Brady SM (2013) Identification of novel loci regulating interspecific variation in root morphology and cellular development in tomato. Plant physiology **162:** 755-768

Rousseaux MC, Jones CM, Adams D, Chetelat R, Bennett A, Powell A (2005) QTL analysis of fruit antioxidants in tomato using Lycopersicon pennellii introgression lines. Theoretical and Applied Genetics **111:** 1396-1408

Ruggeri S, Vahteristo LT, Aguzzi A, Finglas P, Carnovale E (1999) Determination of folate vitamers in food and in Italian reference diet by high-performance liquid chromatography. Journal of Chromatography A **855:** 237-245

Rychlik M, Englert K, Kapfer S, Kirchhoff E (2007) Folate contents of legumes determined by optimized enzyme treatment and stable isotope dilution assays. Journal of Food Composition and Analysis **20:** 411-419

Sagor G, Berberich T, Tanaka S, Nishiyama M, Kanayama Y, Kojima S, Muramoto K, Kusano T (2015) A novel strategy to produce sweeter tomato fruits with high sugar contents by fruit-specific expression of a single bZIP transcription factor gene. Plant Biotechnology Journal

Saito T, Ariizumi T, Okabe Y, Asamizu E, Hiwasa-Tanase K, Fukuda N, Mizoguchi T, Yamazaki Y, Aoki K, Ezura H (2011) TOMATOMA: a novel tomato mutant database distributing Micro-Tom mutant collections. Plant and Cell Physiology **52:** 283-296

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning, Vol 2. Cold spring harbor laboratory press New York

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 science **56:** 179-183

Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Bruedigam C, Kopka J (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nature biotechnology **24:** 447-454

Schneeberger K, Weigel D (2011) Fast-forward genetics enabled by new sequencing technologies. Trends in plant science **16:** 282-288

Schneider E, Ryan TJ (2006) Gamma-glutamyl hydrolase and drug resistance. Clinica chimica acta **374:** 25-32

Scott J, Weir D (1996) Homocysteine and cardiovascular disease. QJM **89:** 561-563

Scott J, Rébeillé F, Fletcher J (2000) Folic acid and folates: the feasibility for nutritional enhancement in plant foods. Journal of the Science of Food and Agriculture **80:** 795-824

Selhub J (1989) Determination of tissue folate composition by affinity chromatography followed by high-pressure ion pair liquid chromatography. Analytical biochemistry **182:** 84-93

Semel Y, Nissenbaum J, Menda N, Zinder M, Krieger U, Issman N, Pleban T, Lippman Z, Gur A, Zamir D (2006) Overdominant quantitative trait loci for yield and fitness in tomato. Proceedings of the National Academy of Sciences **103:** 12981-12986

Sen Gupta D, Thavarajah D, Knutson P, Thavarajah P, McGee RJ, Coyne CJ, Kumar S (2013) Lentils (Lens culinaris L.), a rich source of folates. Journal of agricultural and food chemistry **61:** 7794-7799

Seshadri S, Beiser A, Selhub J, Jacques PF, Rosenberg IH, D'Agostino RB, Wilson PW, Wolf PA (2002) Plasma homocysteine as a risk factor for dementia and Alzheimer's disease. New England Journal of Medicine **346:** 476-483

Shabalina SA, Spiridonov NA, Kashina A (2013) Sounds of silence: synonymous nucleotides as a key to biological regulation and complexity. Nucleic acids research **41:** 2073-2094

Shane B (1989) Folylpolyglutamate synthesis and role in the regulation of one-carbon metabolism. Vitamins & Hormones **45:** 263-335

Shane B (2011) Folate status assessment history: implications for measurement of biomarkers in NHANES. The American journal of clinical nutrition **94:** 337S-342S

Shane B, Stokstad E (1975) Transport and metabolism of folates by bacteria. Journal of Biological Chemistry **250:** 2243-2253

Sharma S, Tyagi K, Narasu ML, Sreelakshmi Y, Sharma R (2011) Mismatch Cleavage by CEL-I Endonuclease: A Tool for Rapid Detection of Homozygous and Heterozygous Mutants. IUP Journal of Genetics & Evolution **4**

Shi H, Xiong L, Stevenson B, Lu T, Zhu J-K (2002) The Arabidopsis salt overly sensitive 4 mutants uncover a critical role for vitamin B6 in plant salt tolerance. The Plant Cell **14:** 575-588

Shohag M, Wei Y-y, Yu N, Zhang J, Wang K, Patring J, He Z-l, Yang X-e (2011) Natural variation of folate content and composition in spinach (Spinacia oleracea) germplasm. Journal of agricultural and food chemistry **59:** 12520-12526

Silink M, Reddel R, Bethel M, Rowe P (1975) Gamma-glutamyl hydrolase conjugase). Purification and properties of the bovine hepatic enzyme. Journal of Biological Chemistry **250:** 5982-5994

Singer VL, Jones LJ, Yue ST, Haugland RP (1997) Characterization of PicoGreen reagent and development of a fluorescence-based solution assay for double-stranded DNA quantitation. Analytical biochemistry **249:** 228-238

Slade AJ, Knauf VC (2005) TILLING moves beyond functional genomics into crop improvement. Transgenic research **14:** 109-115

Smith AG, Croft MT, Moulin M, Webb ME (2007) Plants need their vitamins too. Current opinion in plant biology **10:** 266-275

Sood R, English MA, Jones M, Mullikin J, Wang D-M, Anderson M, Wu D, Chandrasekharappa SC, Yu J, Zhang J (2006) Methods for reverse genetic screening in zebrafish by resequencing and TILLING. Methods **39:** 220-227

Spray G (1964) Microbiological assay of folic acid activity in human serum. Journal of clinical pathology **17:** 660-665

Spronk A, Cossins E (1972) Folate derivatives of photosynthetic tissues. Phytochemistry **11:** 3157-3165

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:** 1

Srivastava AC, Chen F, Ray T, Pattathil S, Peña MJ, Avci U, Li H, Huhman DV, Backe J, Urbanowicz B (2015) Loss of function of folylpolyglutamate synthetase 1 reduces lignin content and improves cell wall digestibility in Arabidopsis. Biotechnology for biofuels **8:** 1

Srivastava AC, Ramos-Parra PA, Bedair M, Robledo-Hernández AL, Tang Y, Sumner LW, de la Garza RID, Blancaflor EB (2011) The folylpolyglutamate synthetase plastidial isoform is required for postembryonic root development in Arabidopsis. Plant physiology **155:** 1237-1251

Srivastava AC, Tang Y, Díaz de la Garza RI, Blancaflor EB (2011) The plastidial folylpolyglutamate synthetase and root apical meristem maintenance. Plant signaling & behavior **6:** 751-754

Staiger D, Brown JW (2013) Alternative splicing at the intersection of biological timing, development, and stress responses. The Plant Cell **25:** 3640-3656

Stanger O, Herrmann W, Pietrzik K, Fowler B, Geisel J, Dierkes J, Weger M (2004) Clinical use and rational management of homocysteine, folic acid, and B vitamins in cardiovascular and thrombotic diseases. Zeitschrift für Kardiologie **93:** 439-453

Stea TH, Johansson M, Jägerstad M, Frølich W (2007) Retention of folates in cooked, stored and reheated peas, broccoli and potatoes for use in modern large-scale service systems. Food chemistry **101:** 1095-1107

Steinberg S (1984) Mechanisms of folate homeostasis. American Journal of Physiology-Gastrointestinal and Liver Physiology **246:** G319-G324

Steinberg SE, Campbell CL, Hillman RS (1979) Kinetics of the normal folate enterohepatic cycle. Journal of Clinical Investigation **64:** 83

Stevens R, Page D, Gouble B, Garchery C, Zamir D, Causse M (2008) Tomato fruit ascorbic acid content is linked with monodehydroascorbate reductase activity and tolerance to chilling stress. Plant, cell & environment **31:** 1086-1096

Stokes P, Webb K (1999) Analysis of some folate monoglutamates by high-performance liquid chromatography–mass spectrometry. I. Journal of Chromatography A **864:** 59-67

Stokstad E (1943) Some properties of a growth factor for Lactobacillus casei. Journal of Biological Chemistry **149:** 573-574

Storozhenko S, De Brouwer V, Volckaert M, Navarrete O, Blancquaert D, Zhang

G-F, Lambert W, Van Der Straeten D (2007b) Folate fortification of rice by metabolic engineering. Nature biotechnology **25:** 1277-1279

Storozhenko S, Navarrete O, Ravanel S, De Brouwer V, Chaerle P, Zhang G-F, Bastien O, Lambert W, Rébeillé F, Van Der Straeten D (2007a) Cytosolic Hydroxymethyldihydropterin Pyrophosphokinase/Dihydropteroate Synthase from Arabidopsis thaliana A SPECIFIC ROLE IN EARLY DEVELOPMENT AND STRESS RESPONSE. Journal of Biological Chemistry **282:** 10749-10761

Strålsjö LM, Witthöft CM, Sjöholm IM, Jägerstad MI (2003) Folate content in strawberries (Fragaria× ananassa): effects of cultivar, ripeness, year of harvest, storage, and commercial processing. Journal of agricultural and food chemistry **51:** 128-133

Syed NH, Kalyna M, Marquez Y, Barta A, Brown JW (2012) Alternative splicing in plants–coming of age. Trends in plant science **17:** 616-623

Tieman DM, Zeigler M, Schmelz EA, Taylor MG, Bliss P, Kirst M, Klee HJ (2006) Identification of loci affecting flavour volatile emissions in tomato fruits. Journal of experimental botany **57:** 887-896

Till BJ, Colbert T, Codomo C, Enns L, Johnson J, Reynolds SH, Henikoff JG, Greene EA, Steine MN, Comai L (2006) High-throughput TILLING for Arabidopsis. Arabidopsis Protocols**:** 127-135

Till BJ, Reynolds SH, Weil C, Springer N, Burtner C, Young K, Bowers E, Codomo CA, Enns LC, Odden AR (2004) Discovery of induced point mutations in maize genes by TILLING. BMC plant biology **4:** 12

Tomsho JW, Moran RG, Coward JK (2008) Concentration-Dependent Processivity of Multiple Glutamate Ligations Catalyzed by Folylpoly-γ-glutamate Synthetase†. Biochemistry **47:** 9040-9050

Tsai H, Howell T, Nitcher R, Missirian V, Watson B, Ngo KJ, Lieberman M, Fass J, Uauy C, Tran RK (2011) Discovery of rare mutations in populations: TILLING by sequencing. Plant physiology **156:** 1257-1268

Tyagi K, Upadhyaya P, Sarma S, Tamboli V, Sreelakshmi Y, Sharma R (2015) High performance liquid chromatography coupled to mass spectrometry for profiling and quantitative analysis of folate monoglutamates in tomato. Food chemistry **179:** 76-84

Ulrich CM, Potter JD (2006) Folate supplementation: too much of a good thing? Cancer Epidemiology Biomarkers & Prevention **15:** 189-193

Vahteristo L, Lehikoinen K, Ollilainen V, Varo P (1997) Application of an HPLC assay for the determination of folate derivatives in some vegetables, fruits and berries consumed in Finland. Food chemistry **59:** 589-597

Van Daele J, Blancquaert D, Kiekens F, Van Der Straeten D, Lambert WE, Stove CP (2014) Folate profiling in potato (Solanum tuberosum) tubers by ultrahigh-performance liquid chromatography–tandem mass spectrometry. Journal of agricultural and food chemistry **62:** 3092-3100

Van Wilder V, De Brouwer V, Loizeau K, Gambonnet B, Albrieux C, Van Der Straeten D, Lambert WE, Douce R, Block MA, Rebeille F (2009) C1 metabolism and chlorophyll synthesis: the Mg-protoporphyrin IX methyltransferase activity is dependent on the folate status. New Phytologist **182:**

137-145

Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends in biotechnology **27:** 522-530

Varshney RK, Terauchi R, McCouch SR (2014) Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. PLoS Biol **12:** e1001883

Verwoerd TC, Dekker B, Hoekema A (1989) A small-scale procedure for the rapid isolation of plant RNAs. Nucleic acids research **17:** 2362

Virk B, Correia G, Dixon DP, Feyst I, Jia J, Oberleitner N, Briggs Z, Hodge E, Edwards R, Ward J (2012) Excessive folate synthesis limits lifespan in the C. elegans: E. coli aging model. BMC biology **10:** 1

Wagner C (2001) BIOCHEMICAL ROLE OF FOLATE IN CELLULAR METABOLISM*. Clinical Research and Regulatory Affairs **18:** 161-180

Wall KP, Pagratis M, Armstrong G, Balsbaugh JL, Verbeke E, Pearson CG, Hough LE (2016) Molecular determinants of tubulin's C-terminal tail conformational ensemble. ACS Chemical Biology

Waller JC, Akhtar TA, Lara-Núñez A, Gregory JF, McQuinn RP, Giovannoni JJ, Hanson AD (2010) Developmental and feedforward control of the expression of folate biosynthesis genes in tomato fruit. Molecular plant **3:** 66-77

Wang C, Riedl KM, Schwartz SJ (2010) A liquid chromatography–tandem mass spectrometric method for quantitative determination of native 5-methyltetrahydrofolate and its polyglutamyl derivatives in raw vegetables. Journal of Chromatography B **878:** 2949-2958

Wang C, Riedl KM, Schwartz SJ (2013) Fate of folates during vegetable juice processing—Deglutamylation and interconversion. Food research international **53:** 440-448

Wang J, Scofield D, Street NR, Ingvarsson PK (2015) Variant calling using NGS data in European aspen (Populus tremula). *In* Advances in the Understanding of Biological Sciences Using Next Generation Sequencing (NGS) Approaches. Springer, pp 43-61

Wang J, Sun J, Liu D, Yang W, Wang D, Tong Y, Zhang A (2008) Analysis of Pina and Pinb alleles in the micro-core collections of Chinese wheat germplasm by Ecotilling and identification of a novel Pinb allele. Journal of Cereal Science **48:** 836-842

Wani NA, Hamid A, Kaur J (2012) Alcohol-associated folate disturbances result in altered methylation of folate-regulating genes. Molecular and cellular biochemistry **363:** 157-166

Ward JL, Poutanen K, Gebruers K, Piironen V, Lampi A-M, Nyström L, Andersson AA, Boros D, Rakszegi M, Bedő Z (2008) The HEALTHGRAIN cereal diversity screen: concept, results, and prospects. Journal of agricultural and food chemistry **56:** 9699-9709

Webb ME, Smith AG, Abell C (2004) Biosynthesis of pantothenate. Natural product reports **21:** 695-721

Wilson S, Horne D (1982) Use of glycerol-cryoprotected Lactobacillus casei for microbiological assay of folic acid. Clinical chemistry **28:** 1198-1200

Winkler S, Schwabedissen A, Backasch D, Bökel C, Seidel C, Bönisch S, Fürthauer M, Kuhrs A, Cobreros L, Brand M (2005) Target-selected mutant screen by TILLING in Drosophila. Genome Research **15:** 718-723

Xia Y, Li R, Ning Z, Bai G, Siddique KH, Yan G, Baum M, Varshney RK, Guo P (2013) Single nucleotide polymorphisms in HSP17. 8 and their association with agronomic traits in barley. PLoS One **8:** e56816

Ye X, Al-Babili S, Klöti A, Zhang J, Lucca P, Beyer P, Potrykus I (2000) Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science **287:** 303-305

Yon M, Hyun TH (2003) Folate content of foods commonly consumed in Korea measured after trienzyme extraction. Nutrition Research **23:** 735-746

Yuan D, Chen J, Shen H, Yang W (2008) Genetics of flesh color and nucleotide sequence analysis of phytoene synthase gene 1 in a yellow-fruited tomato accession PI114490. Scientia horticulturae **118:** 20-24

Yun M-K, Wu Y, Li Z, Zhao Y, Waddell MB, Ferreira AM, Lee RE, Bashford D, White SW (2012) Catalysis and sulfa drug resistance in dihydropteroate synthase. Science **335:** 1110-1114

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

Zeng H, Chen Z-S, Belinsky MG, Rea PA, Kruh GD (2001) Transport of Methotrexate (MTX) and Folates by Multidrug Resistance Protein (MRP) 3 and MRP1 Effect of Polyglutamylation on MTX Transport. Cancer research **61:** 7225-7232

Zhang G-F, Maudens KE, Storozhenko S, Mortier KA, Van Der Straeten D, Lambert WE (2003) Determination of total folate in plant material by chemical conversion into para-aminobenzoic acid followed by high performance liquid chromatography combined with on-line postcolumn derivatization and fluorescence detection. Journal of agricultural and food chemistry **51:** 7872-7878

Zhang G-F, Storozhenko S, Van Der Straeten D, Lambert WE (2005) Investigation of the extraction behavior of the main monoglutamate folates from spinach by liquid chromatography–electrospray ionization tandem mass spectrometry. Journal of Chromatography A **1078:** 59-66

Zhao R, Matherly LH, Goldman ID (2009) Membrane transporters and folate homeostasis: intestinal absorption and transport into systemic compartments and tissues. Expert reviews in molecular medicine **11:** e4

Zhou H-R, Zhang F-F, Ma Z-Y, Huang H-W, Jiang L, Cai T, Zhu J-K, Zhang C, He X-J (2013) Folate polyglutamylation is involved in chromatin silencing by maintaining global DNA methylation and histone H3K9 dimethylation in Arabidopsis. The Plant Cell **25:** 2545-2559

Annexures

Annexure 1

List of tomato accessions used for folate estimation (microbiological assay and LC-MS method) and screening for polymorphism.

S No	Accession No.	Source	Accessions screened for folate estimation using Microbial assay	Accessions screened for folate estimation using LCMS	SNPs scored in genes of folate pathway
1	2-141	TGRC	N	N	*FPGSm*
2	A. Arka	IIVR	N	N	-
3	Agata-30	IIVR	N	N	-
4	ALT9797	IIVR	N	Y	-
5	Arka Vikas (Reference cultivar)	IIVR	Y	Y	-
6	Azad No.1	IIVR	Y	N	-
7	BL-1208	IIVR	N	N	*GCH1*& *GGH3*
8	BT-111-3-2-3	IIVR	N	Y	-
9	Castle Rock	IIVR	N	N	*FPGSm*
10	Cerasiformae	IIVR	Y	N	-
11	Cherry Red	IIVR	N	N	-
12	Chiku Grande	IIVR	Y	N	-
13	CLN-2998	IIVR	N	Y	*FPGSp*
14	DT-10	IIVR	N	Y	-
15	DVKT-1	IIVR	N	N	-
16	EC1129	NBPGR	Y	N	-
17	EC14073	NBPGR	N	N	*FPGSm*
18	EC1753	NBPGR	Y	N	-
19	EC2630	NBPGR	Y	N	-
20	EC 315478	NBPGR	N	N	-
21	EC 320571	NBPGR	N	N	-
22	EC32211	NBPGR	N	Y	-
23	EC 520053	NBPGR	N	N	-
24	EC007345	IIVR	N	N	-

25	EC007785	NBPGR	N	N	-
26	EC009046	IIVR	Y	N	-
27	EC16788	NBPGR	Y	Y	-
28	EC012689	NBPGR	N	N	-
29	EC10662	NBPGR	Y	Y	-
30	EC1087	NBPGR	Y	N	*GGH3*
31	EC11309	NBPGR	N	N	-
32	EC114375	NBPGR	Y	N	-
33	EC1177297	NBPGR	Y	Y	*GGH1*
34	EC1191	NBPGR	N	Y	*GGH3*
35	EC12689	NBPGR	Y	Y	*ADCS, ADCL2, FPGSp* & *GGH3*
36	EC12692	NBPGR	Y	Y	*ADCL2*
37	EC129604	NBPGR	Y	Y	*GGH3*
38	EC13736	NBPGR	Y	Y	*ADCS* & *ADCL1*
39	EC13904	NBPGR	N	Y	*FPGSp* & *GGH3*
40	EC14073	NBPGR	N	N	-
41	EC14078	NBPGR	Y	N	*FPGSm*
42	EC144336	NBPGR	N	Y	-
43	EC15127	NBPGR	Y	N	-
44	EC155	NBPGR	N	Y	-
45	EC161645	NBPGR	N	N	-
46	EC163598	NBPGR	N	N	*GGH3*
47	EC163605	NBPGR	Y	N	-
48	EC16368	NBPGR	Y	N	-
49	EC164465	NBPGR	N	N	-
50	EC164660	NBPGR	N	Y	-
51	EC164665	NBPGR	N	Y	-
52	EC16780	NBPGR	N	N	-
53	EC16786	NBPGR	Y	N	-
54	EC16788	NBPGR	N	N	-
55	EC16790	NBPGR	N	N	-

56	EC168283	NBPGR	N	N	-
57	EC168290	NBPGR	Y	Y	-
58	EC170047	NBPGR	N	Y	*ADCL2*
59	EC177371	NBPGR	Y	Y	-
60	EC1914	NBPGR	N	Y	*FPGSm* & *FPGSp*
61	EC193538	IIVR	N	Y	-
62	EC20636	NBPGR	N	N	*GGH3*
63	EC20639	NBPGR	N	N	*GGH3*
64	EC237288	NBPGR	N	N	-
65	EC241446	NBPGR	Y	Y	-
66	EC251581	NBPGR	N	N	-
67	EC251649	NBPGR	N	N	-
68	EC252	NBPGR	N	N	-
69	EC25265	NBPGR	N	Y	*ADCS*
70	EC25563	NBPGR	N	N	-
71	EC25772	NBPGR	N	N	-
72	EC26150	NBPGR	N	N	*FPGSm* & *GGH3*
73	EC26684	NBPGR	Y	N	*GGH2*
74	EC2673	NBPGR	Y	N	-
75	EC27251	NBPGR	N	N	-
76	EC273966	IIVR	N	N	-
77	EC27885	NBPGR	Y	N	-
78	EC2790	NBPGR	Y	Y	*GGH3*
79	EC279088	NBPGR	N	N	-
80	EC2791	NBPGR	N	N	-
81	EC27910	NBPGR	Y	Y	*ADCS, FPGSm, FPGSp* & *GGH2*
82	EC27960	NBPGR	N	N	-
83	EC27960	NBPGR	N	N	-
84	EC2798	NBPGR	N	N	*GGH2* & *GGH3*
85	EC27995	NBPGR	N	Y	*FPGSm*
86	EC2802	NBPGR	N	Y	-

87	EC28356	NBPGR	Y	N	-
88	EC2977	NBPGR	Y	N	*FPGSp*
89	EC2990	NBPGR	N	N	*ADCL1*
90	EC29914	NBPGR	Y	N	-
91	EC29933	NBPGR	Y	N	-
92	EC29969	NBPGR	N	Y	-
93	EC2997	IIVR	N	N	-
94	EC30303	NBPGR	Y	Y	*ADCL2, FPGSp* & *GGH3*
95	EC3176	NBPGR	Y	Y	-
96	EC31767	NBPGR	Y	Y	*ADCL2*
97	EC320571	NBPGR	Y	N	*GGH3*
98	EC320583	NBPGR	N	N	*GGH3*
99	EC321425	NBPGR	N	N	*GGH3*
100	EC3216	NBPGR	N	Y	*ADCS, FPGSp* & *GGH3*
101	EC32287	NBPGR	Y	N	-
102	EC32481	NBPGR	N	Y	-
103	EC32557	NBPGR	N	N	*GGH3*
104	EC3261	NBPGR	Y	N	-
105	EC326139	NBPGR	Y	N	*FPGSm*
106	EC338717	NBPGR	N	N	-
107	EC338725	NBPGR	N	Y	-
108	EC33878	NBPGR	N	N	-
109	EC339058	IIVR	Y	Y	-
110	EC339066	NBPGR	Y	N	-
111	EC3414425	IIVR	Y	Y	*ADCL1*
112	EC34477	NBPGR	N	N	*GGH3*
113	EC34480	NBPGR	N	N	*ADCS, ADCL2* & *GGH3*
114	EC35236	NBPGR	N	N	-
115	EC35240	NBPGR	Y	Y	-
116	EC-35240	NBPGR	N	N	-
117	EC35242	NBPGR	N	Y	-
118	EC35244	NBPGR	N	N	-

119	EC35272	NBPGR	N	Y	-
120	EC35293	NBPGR	N	N	-
121	EC35322	NBPGR	N	Y	-
122	EC35360	NBPGR	N	N	*GCH1*
123	EC357828	NBPGR	N	N	-
124	EC362933	NBPGR	N	N	-
125	EC362941	NBPGR	N	N	-
126	EC362948	NBPGR	N	Y	-
127	EC362949	NBPGR	N	N	*GGH2* & *GGH3*
128	EC362958	NBPGR	Y	N	-
129	EC363863	NBPGR	N	Y	-
130	EC363942	NBPGR	N	N	-
131	EC3668	NBPGR	Y	N	-
132	EC368832	NBPGR	Y	N	-
133	EC368883	NBPGR	Y	N	*ADCL1* & *GGH3*
134	EC368943	NBPGR	Y	N	-
135	EC369020	NBPGR	N	N	*GCH1* & *FPGSm*
136	EC370867	IIVR	Y	N	*FPGSm*
137	EC372	NBPGR	Y	N	*ADCL2*
138	EC373378	NBPGR	Y	N	-
139	EC381554	NBPGR	Y	Y	*ADCS*
140	EC383117	NBPGR	Y	Y	*ADCS*
141	EC398405	IIVR	N	N	*GCH1* & *GGH3*
142	EC398600	NBPGR	Y	N	*GCH1*
143	EC398614	NBPGR	Y	N	-
144	EC398684	NBPGR	N	N	-
145	EC398685	NBPGR	N	N	-
146	EC398687	NBPGR	N	Y	-
147	EC398688	NBPGR	N	N	-
148	EC398695	NBPGR	N	N	*GGH2*
149	EC398697	NBPGR	Y	N	-
150	EC398699	NBPGR	Y	N	*GGH3*

151	EC398701	NBPGR	Y	N	*GCH1*
152	EC398704	NBPGR	N	N	*GGH3*
153	EC398707	NBPGR	N	Y	*ADCS*
154	EC398710	NBPGR	Y	Y	-
155	EC398711	NBPGR	Y	N	-
156	EC398712	NBPGR	Y	N	-
157	EC398714	NBPGR	Y	N	-
158	EC398715	NBPGR	Y	N	-
159	EC398716	NBPGR	Y	N	-
160	EC398717	NBPGR	Y	N	-
161	EC429	NBPGR	Y	N	-
162	EC433607	NBPGR	Y	Y	*ADCS* & *FPGSm*
163	EC439542	NBPGR	N	N	-
164	EC443369	NBPGR	Y	N	-
165	EC4506	NBPGR	N	N	-
166	EC458213	NBPGR	N	N	-
167	EC470413	NBPGR	N	Y	-
168	EC487	NPGRB	N	N	-
169	EC490128	NBPGR	Y	N	*GGH3*
170	EC490130	NBPGR	Y	N	-
171	EC492	NBPGR	Y	N	-
172	EC494372	IIVR	N	Y	-
173	EC498372	NBPGR	N	Y	-
174	EC5050	IIVR	N	Y	-
175	EC501577	NBPGR	N	N	*GGH3*
176	EC520046	NBPGR	Y	Y	*GGH3*
177	EC 52055	NBPGR	Y	Y	-
178	EC520052	NBPGR	Y	N	-
179	EC520059	NBPGR	Y	Y	-
180	EC520065	IIVR	N	N	-
181	EC520075	NBPGR	Y	N	*GGH3*
182	EC520076	IIVR	N	N	*GCH1, FPGSm* & *GGH3*

183	EC520077	IIVR	Y	N	*FPGSm*
184	EC520078	IIVR	N	N	-
185	EC520079	IIVR	N	N	*GGH3*
186	EC521039	NBPGR	N	N	-
187	EC521048	NBPGR	Y	N	-
188	EC521049	NBPGR	Y	N	-
189	EC521061	NBPGR	Y	Y	-
190	EC521067	NBPGR	Y	N	-
191	EC521068	NBPGR	N	Y	-
192	EC521070	NBPGR	Y	Y	-
193	EC521077	NBPGR	Y	Y	-
194	EC521078	IIVR	Y	Y	-
195	EC521079	NBPGR	N	N	-
196	EC521080	IIVR	N	Y	-
197	EC521082	NBPGR	N	Y	-
199	EC521083	NBPGR	Y	N	*GGH2* & *GGH3*
200	EC521086	NBPGR	Y	Y	*ADCS*
201	EC526146	NBPGR	N	N	*GGH3*
202	EC528362	NBPGR	N	Y	-
203	EC528365	IIVR	Y	N	-
204	EC528366	NBPGR	Y	N	-
205	EC528367	IIVR	N	N	*GGH3*
206	EC528372	NBPGR	Y	N	-
207	EC528373	NBPGR	Y	N	-
208	EC528374	NBPGR	Y	N	-
209	EC528388	NBPGR	Y	Y	*ADCS* & *GGH3*
210	EC529081	NBPGR	N	N	*ADCL1*, *FPGSm* & *GGH3*
211	EC529083	NBPGR	Y	Y	*FPGSp* & *GGH2*
212	EC529085	NBPGR	N	N	*ADCL1*, *GGH3*
213	EC529086	NBPGR	N	N	-
214	EC531800	NBPGR	Y	Y	-
215	EC531801	NBPGR	N	Y	-

216	EC531802	NBPGR	Y	N	-
217	EC531805	NBPGR	Y	Y	*GGH1* & *GGH3*
218	EC5358139	IIVR	N	N	-
219	EC538139	NBPGR	Y	N	*GCH1* & *FPGSm*
220	EC538141	NBPGR	N	N	*GGH2*
221	EC538146	NBPGR	Y	Y	*GGH3*
222	EC538148	NBPGR	Y	N	-
223	EC538149	NBPGR	N	N	*ADCL1*
224	EC538153	NBPGR	N	N	-
225	EC538455	NBPGR	Y	Y	-
226	EC546727	NBPGR	Y	N	*GGH3*
227	EC562073	NBPGR	Y	N	-
228	EC5627	IIVR	Y	N	-
229	EC565216	NBPGR	N	N	-
230	EC57442	NBPGR	N	Y	-
231	EC5863	NBPGR	N	N	-
232	EC5888	NBPGR	N	N	-
233	EC6053	NBPGR	Y	N	-
234	EC6192	NBPGR	N	N	*GGH3*
235	EC6486	NBPGR	N	Y	-
236	EC6488	NBPGR	Y	N	-
237	EC6845	NBPGR	Y	N	-
238	EC7317	NBPGR	Y	Y	-
239	EC742	NBPGR	N	N	-
240	EC7785	NBPGR	Y	N	-
241	EC7912	NBPGR	Y	N	-
242	EC8372	IIVR	N	N	*GCH1*
243	EC8591	NBPGR	Y	N	*FPGSm*
244	EC8630	NBPGR	Y	N	-
245	EC8822	NBPGR	Y	Y	-
246	EC8936	NBPGR	Y	N	*GGH2* & *GGH3*
247	EC9046	NBPGR	Y	Y	-

248	F6030	IIVR	N	N	-
249	F6102	IIVR	Y	N	-
250	Feb.2	IIVR	N	N	*GCH1*
251	Feb.4	IIVR	Y	N	-
252	FLA7171	IIVR	N	N	-
253	Gujrat Tomato	IIVR	N	N	-
254	H-24	IIVR	N	Y	*GCH1* & *FPGSp*
255	H-86	IIVR	N	N	-
256	H-88-78-2	IIVR	Y	N	-
257	H-88-78-5	IIVR	N	Y	*FPGSp*
258	H-88-87	IIVR	N	Y	*FPGSm*
259	IC 469697	NBPGR	N	N	-
260	IC447706	NBPGR	Y	Y	-
261	IC447708	NBPGR	N	Y	-
262	IC469597	NBPGR	N	Y	-
263	IC469628	NBPGR	N	Y	-
264	IC469648	NBPGR	Y	Y	-
265	IC469653	NBPGR	Y	N	-
266	IC469682	NBPGR	Y	N	-
267	IC469714	NBPGR	N	Y	-
268	IIHR2201	IIVR	Y	Y	*FPGSp, GGH1* & *GGH3*
269	IIVR 121	IIVR	N	Y	-
270	IIVR2200	IIVR	Y	N	-
271	KT15	IIVR	Y	Y	-
272	LA0012	TGRC	N	Y	-
273	LA0180	TGRC	Y	N	-
274	LA0215	TGRC	N	N	-
275	LA0266	TGRC	Y	N	-
276	LA0274	TGRC	N	N	-
277	LA0276	TGRC	Y	N	-
278	LA0292	TGRC	Y	N	-
279	LA0505	TGRC	Y	N	-

280	LA0516	TGRC	N	N	-
281	LA0517	TGRC	N	Y	*FPGSm*
282	LA0533	TGRC	N	N	-
283	LA0744	TGRC	N	N	-
284	LA0806	TGRC	N	Y	-
285	LA0842	TGRC	Y	N	-
286	LA0854	TGRC	Y	N	-
287	LA1016	TGRC	Y	N	-
288	LA1021	TGRC	N	N	-
289	LA1088	TGRC	N	N	-
290	LA1089	TGRC	N	N	-
291	LA1090	TGRC	N	Y	-
292	LA1091	TGRC	N	N	-
293	LA1504	TGRC	Y	N	-
294	LA1506	TGRC	N	N	-
295	LA1795	TGRC	N	N	-
296	LA1996	TGRC	Y	N	-
297	LA2133	TGRC	N	N	-
298	LA2374	TGRC	N	N	-
299	LA2400	TGRC	N	N	-
300	LA2529	TGRC	Y	N	-
301	LA2713	TGRC	N	Y	-
302	LA2714	TGRC	N	N	-
303	LA2715	TGRC	Y	N	-
304	LA2818	TGRC	Y	N	-
305	LA2921	TGRC	Y	Y	-
306	LA2968	TGRC	Y	N	-
307	LA2999	TGRC	N	N	-
308	LA3012	TGRC	N	N	-
309	LA3024	TGRC	Y	N	-
310	LA3121	TGRC	N	Y	-
311	LA3144	TGRC	Y	N	-

312	LA3202	TGRC	Y	N	-
313	LA3203	TGRC	N	Y	-
314	LA3229	TGRC	Y	N	-
315	LA3231	TGRC	Y	N	-
316	LA3233	TGRC	Y	Y	-
317	LA3234	TGRC	Y	N	-
318	LA3237	TGRC	N	N	-
319	LA3243	TGRC	N	N	-
320	LA3246	TGRC	N	N	-
321	LA3247	TGRC	N	Y	-
322	LA3317	TGRC	N	N	-
323	LA3430	TGRC	Y	N	-
324	LA3465	TGRC	Y	Y	*GGH1*
325	LA3530	TGRC	Y	N	-
326	LA3532	TGRC	N	N	-
327	LA3534	TGRC	Y	N	-
328	LA3537	TGRC	N	N	-
329	LA3538	TGRC	Y	N	-
330	LA3539	TGRC	Y	Y	-
331	LA3551	TGRC	Y	N	-
332	LA3554	TGRC	Y	N	-
333	LA3579	TGRC	Y	N	-
334	LA3630	TGRC	Y	N	-
335	LA3754	TGRC	N	N	-
336	LA3770	TGRC	N	N	-
337	LA3903	TGRC	Y	N	-
339	LA4024	TGRC	Y	N	-
340	LA4104	TGRC	Y	N	*GGH3*
341	M.local	IIVR	Y	N	-
342	M-88-78-3	IIVR	Y	Y	*FPGSp*
343	Mount favest	IIVR	Y	N	-
344	Mutant	IIVR	N	N	-

345	N2298	BJJ	Y	N	-
346	N2369	BJJ	Y	N	-
347	N2389	BJJ	Y	Y	-
348	N2403	BJJ	Y	Y	-
349	N2411	BJJ	Y	Y	-
350	N2448	BJJ	Y	Y	-
351	N2465	BJJ	Y	Y	-
352	N2501	BJJ	Y	Y	*FPGSp*
353	N2664	BJJ	Y	Y	-
354	N2675	BJJ	Y	Y	-
355	N2764	BJJ	Y	Y	-
356	N2812	BJJ	Y	Y	-
357	N2829	BJJ	Y	N	-
358	N2832	BJJ	Y	Y	-
359	N4210	BJJ	N	N	-
360	N4218	BJJ	N	Y	-
361	N4241	BJJ	Y	N	-
362	Nandi	IIVR	Y	N	*GGH1*
363	P.Gaurav	IIVR	N	N	-
364	P.Pink	IIVR	Y	N	*FPGSp* & *GGH1*
365	P.Rohit	IIVR	N	N	-
366	PBC(Punjab Chuhara)	IIVR	N	N	-
367	PDT-3-1	IIVR	Y	N	-
368	PKM-1	IIVR	Y	Y	-
369	Pusa Rohini	IIVR	N	N	-
370	Pusa-Gaurav	IIVR	N	N	-
371	Rashmi	IIVR	N	N	-
372	S.local	IIVR	N	N	-
373	S-2-95-1-3-1	IIVR	N	Y	-
374	Sankranti	IIVR	N	N	*FPGSp*
375	Sel-14	IIVR	N	Y	-

376	Siberia	IIVR	Y	N	-
377	Superbug	IIVR	N	N	*GGH1*
378	T.Local	IIVR	Y	N	-
379	T-HL	IIVR	N	N	-
380	TLBR-12	IIVR	N	N	-
381	TLBR-2	IIVR	Y	N	*FPGSm*
382	TLBR-3	IIVR	Y	N	-
383	TLBR-4	IIVR	Y	N	-
384	Type-1	IIVR	Y	N	*FPGSm*
385	Vaibhav	IIVR	Y	N	*GGH3*
386	VLT-34	IIVR	N	N	-
387	WIR13717	IIVR	Y	N	*FPGSm*
388	WIR3768	IIVR	Y	N	*GGH2* & *GGH3*
389	WIR3928	IIVR	N	N	*FPGSm* & *GGH3*
390	WIR3957	IIVR	Y	Y	*FPGSm* & *FPGSp*
391	WIR4361	IIVR	N	Y	-

TGRC: Tomato Genetics Resource Center at University of California, Davis (www.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) (http://www.nbpgr.ernet.in:8080/PGRPortal/(S(o3cy12bkoz5s5e55w0dacf2a))/AdvancePassportSearch.aspx).

BSS: Bejo Sheetal Seeds Pvt. Ltd. Jalna, India (http://www.bejosheetalseeds.com/).

IIHR: Indian Institute of Horticulture Research, India (http://www.iihr.res.in/)

Y= Accessions used for the above mentioned analysis.

N =Accessions were not available for analysis.

- = No SNPs detected.

SNPs shown in genes other than *FPGS* and *GGH* isoforms were screened by other researchers of our lab.

Annexure 2

Number of accessions and codes in protein similarity tree-84 genomes

Gene name	Group name in tree	Accessions in that group
FPGSp	73	73
	74	74
	75	75
	46	46
	25	25
	H	Heinz,1,2,3,4,5,6,7,8,11,12,13,14,15,16,17,18,19,20,21,22,23,24,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,44,45,47,51,52,53,54,55,56,59,77,78,88,89,90,91,93,94,96,97,102,103,104,105
	60	60
	58	58
	57	57
	49	49
	67	67
	69	69
	63	63
	62	62
	64	64
	65	65
	43	43
	66	66,68,71,72
	70	70
FPGSm	4	4,13,18,20,22,23,37,40,43,78,88,96,102
	68	68,71
	63	63
	3	3,5,7,17,21,27,44,45,47,54,93,97
	55	55
	56	56
	57	57
	64	64
	105	105
	58	58
	65	65
	104	104
	75	75
	H	Heinz,1,2,6,8,11,12,14,15,16,19,24,26,28,29,30,31,32,33,34,35,36,38,39,41,77,89,90,94,103
	49	49
	46	46
	42	42
	25	25
	67	67
	66	66
	53	53
	51	51,52
	72	72
	70	70
	74	74
	73	73
	62	62
	59	59
	60	60
	69	69

Gene name	Group name in tree	Accessions in that group
DHFS	70	70
	71	71
	51	51,52
	25	25
	22	22,23,88
	H	Heinz,1,2,3,4,5,6,7,8,11,12,13,14,15,16,17,18,19,20,21,24,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,53,54,55,77,78,89,90,91,93,94,96,97,102,103,104,105
	56	56,57
	73	73
	49	49
	26	26
	75	75
	67	67,69
	60	60
	58	58,59
	62	62
	63	63
	64	64
	65	65
	66	66
	74	74
	68	68
GGH1	22	22,44,54
	67	67
	62	62
	15	15,42,46,103,105
	51	51
	52	52
	56	56
	63	63
	57	57
	49	49
	64	64
	75	75
	74	74
	H	Heinz,1,2,3,4,5,6,7,8,11,12,13,14,16,17,18,19,20,21,24,26,27,28,29,30,31,32,33,34,35,36,38,40,41,43,45,77,78,88,89,90,91,93,94,96,97,102
	39	39
	37	37,55
	25	25
	23	23,53,104
	66	66
	65	65
	68	68
	47	47
	70	70
	69	69
	73	73
	71	71,72
	58	58
	59	59
	60	60

Gene name	Group name in tree	Accessions in that group
GGH2	52	52
	51	51
	74	74
	25	25
	16	16,17,22,23,37,42,44,53,88,103,104
	H	Heinz,1,2,3,4,5,6,7,8,11,12,13,14,15,18,19,20,21,24,26,27,28,29,30,31,32,33,34,35,36,38,39,40,41,43,45,54,77,78,89,90,91,93,94,96,97,102
	73	73
	49	49
	47	47
	46	46,105
	63	63
	75	75
	66	66,68,71,72
	67	67,69,70
	56	56,57,59
	58	58
	60	60
	62	62
	64	64,65
	55	55
GGH3	53	53,104
	66	66,72
	58	58
	59	59
	67	67
	56	56,57
	68	68
	23	23
	6	6,8,14,21,27,40,45,88,90,93,97
	H	Heinz,1,2,3,4,5,7,11,12,13,15,16,17,18,19,20,22,24,26,28,29,30,31,32,33,34,35,36,38,39,41,42,43,44,46,47,54,77,78,89,91,94,96,102,103,105
	51	51,52
	49	49
	37	37
	25	25
	70	70
	69	69
	73	73
	71	71
	75	75
	74	74
	60	60
	62	62
	63	63
	64	64
	65	65

Annexure 2.1

Accession names and codes for 84 genomes

Nr	Accession ID	EUSOL accession ID	Accession Name	Reasons	Genebank
1	PV	several	Moneymaker		PBR
2	PV	several	Ailsa Craig	old cultivar	PBR
3	PV	EA06086	Garderners Delight	old cultivar	PBR
4	PV	EA00465	Rutgers	old cultivar	Craig Lehoullier
5		EA00325	GALINA	Heirloom, diversity	Craig Lehoullier
6		EA00448	PONDEROSA (revision 050215)	Heirloom, diversity	
7		EA00375	Katinka Cherry	Heirloom / phylogeny	Craig Lehoullier
8		EA00371	John's big orange	Heirloom / phylogeny	Craig Lehoullier
11	LYC 1365	EA02617	All Round	Agronomic traits	IPK Gatersleben
12	LYC 1969	EA02724	Sonato	Agronomic traits	IPK Gatersleben
13	LYC 3897	EA03701	Cross Country	Agronomic traits	IPK Gatersleben
14	LYC 3476	EA03362	Iidi	Plant architecture	IPK Gatersleben
15		TR00003	Momotaro (Tough Boy)	pink beef	Rijk Zwaan
16	CGN15464	EA01965	Rote Beere	brix	TGRC
17	LYC 3340	EA03306		brix	IPK Gatersleben
18		EA01155	DANA	fruit size	Sand Hill Preservation Center
19		EA01049	LARGE PINK	fruit weight	Sand Hill Preservation Center
20	LYC 3153	EA03221	Lycopersicon esculentum. Mill.	fruit weight	IPK Gatersleben
21	LYC 3155	EA03222	Bolivar' Lycopersicon esculentum. Mill. Convar. infiniens Lehm. Var. commune bail.	fruit weight	IPK Gatersleben
22	PI 129097	EA04710		fruit weight	USDA, ARS, Geneva
23	PI 272654	EA05170		fruit weight	USDA, ARS, Geneva
24		EA00990	JERSEY DEVIL	fruit size	Sand Hill Preservation Center
26		EA00157	Polish Joe	fruit locule count	
27	CGN20815	EA02054	Cal J TM VF	fruit locule count	TGRC
28	PI 303721	EA05581	The Dutchman	fruit locule count	GENEVA
29		EA00027	BLACK CHERRY		
30	V710092	EA01835	ANTO	collection date	
31	PC711092	EA01854	WINTER TIPE (NOR)	collection date	
32	PI 93302	EA04243	'Chang Li' Lycopersicon esculentum	collection date / collection iste	USDA, ARS, Geneva
33	SG 16	EA00892	Belmonte		
34		EA01088	TIFFEN MENNONITE	phylogeny	Sand Hill Preservation Center

35	PI 203232	EA04939	Wheatley's Frost Resistant	collection site	USDA, ARS, Geneva
36	PI 311117	EA05701		collection site	USDA, ARS, Geneva
37	PI 365925	EA05891		collection site	USDA, ARS, Geneva
38	PI 158760	EA04828	Chih-Mu-Tao-Se	collection site / phylogeny	USDA, ARS, Geneva
39	LA0113	EA00526		collection site / germplasm status	TGRC
40	LYC 1410	EA02655	ES 58 Heinz' Lycopersicon esculentum Mill. Convar. infiniens Lehm. Var. pluriloculare Lehm.	phylogeny	IPK Gatersleben
41	PI 169588	EA04861	Dolmalik	collection date / phylogeny	USDA, ARS, Geneva
42	LYC 2962	EA03107		phylogeny	IPK Gatersleben
45	LYC 2740	EA02960		Plant architecture, Request Keygen	IPK Gatersleben
77		TR00018	Large Red Cherry		Totally Tomatoes
78		EA00940	Porter		Tomato Growers Supply Company
88		TR00019	Bloody Butcher		Totally Tomatoes
89		EA01019	Brandywine		Tomato Growers Supply Company
90		TR00020	Dixy Golden Giant		Tomato Growers Supply Company
91		EA01037	Giant Belgium		Tomato Growers Supply Company
93		TR00021	Kentucky Beefsteak		Tomato Growers Supply Company
94		TR00022	Marmande VFA		Tomato Growers Supply Company
96		TR00023	Thessaloniki		Tomato Growers Supply Company
97		EA01640	Watermelon Beefsteak		Tomato Growers Supply Company
102	LA4133	TR00026			TGRC
103	LA1421	TR00027			TGRC
105	LA1479	TR00028			TGRC

Continued on the next page….

Wild relatives of tomato					
43	LYC 2910	EA03058	*S. pimpinellifolium (Jusl.) Mill.*		collection site
44	LYC 2798	EA02994	*S. pimpinellifolium (Jusl.) Mill. var. pimpinellifolium.*		fruit wieght
46	LA1584	EA00676	*S. pimpinellifolium*		Whitefly resistance
47	LA1578	EA00674	*S. pimpinellifolium*		Drought resistance
49	LA1278	TR00005	*S. peruvianum?*		Salinity / alkalinity resistance
51	LA2663	TR00007	*S. chiemliewskii*		Steep banks in rocks
52	LA2695	EA00759	*S. chiemliewskii*		
53	LA0483	EA00581	*S. cheesmaniae*		Agronomic traits
54	CGN15820	TR00024	*S. cheesmaniae*		F. oxysporum resistance
55	LA1401	EA00652	*S. cheesmaniae*		
56	LA2133	EA00729	*S. neorickii*		Agronomic traits
57	CGN24193	TR00025	*S. neorickii*		O. neolycopersici & B. cinerea resistance
58	LA2157	TR00008	*S. Arcanum*		C. michiganensis resistance
59	LA2172	TR00009	*S. Arcanum*		
60	LA1954	EA00713	*S. peruvianum*		
62	LA1983	TR00010	*S. huaylasense*		dry habitat
63	LA1365	TR00011	*S. huaylasense*		Steep dry banks
64	CGN15532	TR00012	*S. chilense*		
65	CGN15530	TR00013	*S. chilense*		
66	CGN15791	TR00014	*S. habrochaites f. glabratum*		
67	PI134418	TR00015	*S. habrochaites f. glabratum*		
68	CGN15792	TR00016	*S. habrochaites f. glabratum*		
69	LA1718	EA00699	*S. habrochaites f. glabratum*		
70	LA1777	EA00703	*S. habrochaites*		Agronomic traits
71	LA407	EA00558	*S. habrochaites*		C. michiganensis resistance
72	LYC4	TR00017	*S. habrochaites*		B. cinerea resistance & Agronomic traits
73	LYC 1831	EA02701	*S. pennellii*		steep rocky slope, LA1272
74	LA0716	EA00585	*S. pennellii*		Parent introgression lines
75	LA1364	TR00030	*S. huaylense*		
104	LA1044	TR00029	*S. galapagense*		
25	T 1248	EA03384	*S. corneliomulleri*LA0118		fruit size

Annexure 3

Number of accessions and codes in protein similarity tree-360 genomes

Gene name	Group name in tree	Accessions in that group
FPGSp	H	Heinz,ts1,ts2,1,10,100,101,102,103,104,105,106,107,108,109,11,110,111,112,113,114,115,116,117,118,119,12,120,121,122,123,124,125,126,127,128,129,13,130,131,132,133,134,135,136,137,138,139,14,140,141,142,143,144,145,146,147,149,150,151,152,153,154,155,156,157,159,16,160,161,162,163,164,165,166,167,168,169,17,170,171,172,173,174,175,176,177,178,179,18,180,181,182,183,184,185,186,187,188,189,19,190,191,192,193,194,195,196,197,198,199,2,20,200,201,202,203,204,205,206,207,208,209,21,210,211,212,213,214,215,217,218,219,22,220,221,222,224,225,226,227,228,229,23,230,231,232,233,234,235,236,237,238,239,24,240,241,242,243,244,245,246,247,248,249,25,250,251,252,253,254,255,256,257,258,259,26,260,261,262,263,264,265,266,267,268,269,27,270,271,272,273,274,275,276,277,278,28,280,281,282,283,284,285,286,287,288,289,29,290,291,292,293,294,295,296,297,298,299,3,300,302,303,304,305,306,308,309,31,310,311,312,313,314,315,316,317,318,319,32,320,321,35,36,37,38,39,4,40,400,401,409,41,410,411,412,413,414,415,416,417,418,419,42,420,,421,422,423,424,425,426,427,428,429,43,430,431,432,433,434,435,436,437,438,439,44,440,441,45,46,47,48,49,5,50,51,52,53,54,55,56,57,58,59,6,60,61,62,63,64,65,66,67,68,69,7,70,71,72,73,74,75,76,77,78,79,8,80,81,82,83,84,85,86,87,88,89,9,90,91,92,93,94,95,96,97,98,99
	148	148,15,158,223,301,33,34
	216	216,402,403,404
	30	30
	307	307
	407	407,408
FPGSm	H	Heinz,ts1,ts2,1,10,100,101,102,103,104,108,109,11,111,114,117,12,121,126,127,13,130,131,132,133,136,140,141,142,143,152,159,160,162,1163,167,168,169,170,173,175,176,178,179,180,183,185,191,192,193,194,195,197,198,2,201,203,204,205,206,210,212,213,215,218,225,226,230,232,238,239,241,242,243,246,249,251,252,253,255,259,27,270,271,272,273,275,276,277,28,280,282,283,284,285,286,289,29,290,292,293,294,297,30,302,303,305,306,308,310,311,312,313,314,315,316,36,400,401,409,41,42,43,44,46,47,49,5,51,52,55,58,60,63,66,7,70,72,,73,74,75,76,8,83,86,88,9,93,95
	105	105,112,119,120,129,139,154,158,165,172,177,188,189,200,219,221,229,240,247,250,258,261,281,295,31,34,35,38,426,432,48,50,61,64,65,67,69,81,99
	116	116,118,123,124,128,134,138,149,150,156,166,17,174,186,187,190,199,207,208,209,214,220,227,228,23,231,237,244,260,265,274,278,296,298,299,3,300,317,318,319,320,321,39,40,411,415,420,421,422,423,425,427,430,431,435,437,438,439,440,45,53,54,57,62,68,77,84,85,90,92,94,96,98
	106	106,196,82
	107	107,257,417
	110	110,122,125,287
	113	113,184,222
	115	115,164,171,202,262,424
	135	135,89
	137	137
	14	14,145,148,15,16,22,301,33,413,419,56
	144	144,147,151,18,19,21,211,234,24,245,254,266,291,32,37,4,410,412,414,416,429,434,441,78,79,91,97
	146	146
	153	153,235
	155	155,157,181
	161	161
	182	182,59,87
	20	20,436
	216	216
	217	217
	223	223
	224	224,428
	233	233
	236	236
	248	248
	25	25
	256	256
	267	267

Gene name	Group name in tree	Accessions in that group
FPGSm...	268	268
	288	288
	304	304
	307	307
	309	309
	402	402,408
	403	403
	404	404
	407	407
	418	418
	433	433
	6	6
	71	71
DHFS	H	Heinz,ts1,ts2,1,10,100,101,102,103,104,105,106,107,108,11,110,111,112,113,114,115,117,118,119,12,120,121,122,123,124,125,126,127,128,129,13,130,131,132,133,134,135,136,137,138,139,14,140,141,142,143,145,146,147,148,149,150,151,152,153,154,155,157,158,159,16,160,161,162,163,164,165,166,167,168,169,170,171,172,173,174,175,176,177,178,179,18,180,181,182,183,184,185,186,187,188,189,19,190,191,192,193,194,195,196,197,198,199,2,20,200,201,202,203,204,205,206,207,208,209,21,210,211,212,213,214,215,216,217,218,219,220,221,222,223,224,225,226,227,228,229,23,230,231,232,233,234,235,236,237,238,239,24,241,242,243,244,245,246,247,248,249,25,250,251,252,253,254,255,256,257,258,259,26,260,261,262,263,264,265,266,267,268,269,27,270,271,272,273,274,275,276,277,278,28,280,281,282,283,284,285,286,287,288,289,29,290,291,292,293,294,295,296,297,298,299,3,30,300,301,302,303,304,305,306,31,310,312,314,315,316,317,318,319,32,320,321,34,35,36,37,38,39,4,400,401,402,403,404,408,409,41,410,411,413,414,415,416,417,418,419,42,420,421,422,424,425,426,427,428,429,43,430,432,433,434,435,437,438,439,44,440,441,45,46,47,48,49,5,50,51,52,53,54,55,56,57,58,59,6,60,61,62,63,64,65,66,67,68,69,7,70,71,72,73,74,75,76,77,78,79,8,80,81,82,83,84,85,86,87,88,89,9,90,91,92,93,94,95,96,97,98,99
	109	109,116,144,156,17,22,33,412,423,431
	15	15,240,40,436
	307	307,308,309,311,313
	407	407
GGH1	H	Heinz,ts1,ts2,1,10,100,101,102,103,104,105,106,107,108,109,110,111,112,113,114,115,117,119,120,122,125,126,128,129,130,131,132,133,135,136,137,138,139,140,141,142,143,147,150,151,152,153,154,155,157,158,159,160,161,162,163,166,167,168,169,170,171,172,173,174,175,176,177,178,179,180,183,184,185,186,187,188,189,190,191,192,193,194,195,196,197,198,2,200,201,202,203,204,205,206,209,210,211,212,213,214,215,218,219,220,221,222,224,225,227,228,229,230,231,232,233,234,235,236,237,238,239,240,241,242,243,245,246,249,250,251,252,253,254,255,256,257,258,259,26,260,261,263,264,268,269,27,270,271,272,273,274,275,276,277,278,28,280,281,282,283,284,285,286,288,289,29,290,292,293,294,297,298,3,301,303,307,308,309,310,311,313,314,317,318,319,32,320,321,4,400,401,41,42,423,426,427,43,436,44,45,46,47,49,5,51,52,53,54,55,56,57,58,59,6,60,61,62,63,64,65,66,67,69,70,71,72,73,74,75,76,78,8,80,81,82,83,85,86,88,89,9,90,91,93,95,96,98,99
	11	11,123,23,244,248,262,287,302,304,312,316,37,39,40,409,410,411,412,414,415,419,420,434,438,440,441,7,79,92
	116	116,226
	118	118,199,207,217,30,31,36
	12	12,18,424,84
	121	121,124,164,182,265,291
	127	127
	13	13,247,48
	134	134,148,15,208,223,25,295,299,33,34,413,430,431,50,87
	14	14
	144	144,97
	145	145,16,267,300,417,421
	146	146
	149	149
	156	156,315
	165	165,181,35,432
	17	17
	19	19
	20	20,305,416,433
	21	21,306,435
	216	216

Gene name	Group name in tree	Accessions in that group
GGH1	22	22,24
	266	266
	296	296,422,425,439
	38	38
	402	402
	403	403
	404	404
	407	407
	408	408
	418	418
	428	428,429
	437	437,77
	68	68
	94	94
GGH2	H	Heinz,ts1,ts2,1,10,100,101,102,103,104,105,107,108,109,11,110,111,112,113,114,115,117,118,119,12,120,121,122,123,124,125,126,127,128,129,130,131,132,133,135,136,137,138,139,140,141,142,143,144,145,147,149,150,151,152,153,154,155,157,159,160,161,162,163,164,165,166,167,168,169,170,171,172,173,174,175,176,177,178,179,180,181,183,184,185,186,187,188,189,190,191,192,193,194,195,196,197,198,199,2,200,201,202,203,204,206,209,210,211,212,213,214,215,218,219,220,221,222,224,225,226,227,228,229,230,232,233,234,235,236,237,238,239,241,242,243,245,246,249,250,251,252,253,254,254,255,256,257,259,26,260,263,264,267,268,269,27,270,271,272,273,274,275,276,277,278,28,280,281,282,283,284,285,286,288,289,29,290,291,292,293,294,296,297,298,3,301,302,304,305,306,307,308,309,310,311,312,313,314,315,316,317,318,319,32,320,321,35,36,37,4,400,417,42,426,428,43,432,44,45,46,47,49,5,51,52,54,55,56,58,59,6,60,61,62,63,64,65,66,67,69,7,70,71,72,73,74,75,76,78,8,80,81,82,83,84,85,86,88,89,9,90,91,92,93,94,95,96,97,98,99
	106	106,13,134,14,148,15,158,16,17,18,205,207,208,21,217,223,23,231,24,240,247,25,258,261,262,266,295,299,30,300,303,31,33,34,38,39,407,409,41,411,412,413,416,418,419,420,422,425,429,430,431,434,436,438,439,440,441,48,50,53,57,68,77,79,87
	116	116,22,414,427
	146	146
	156	156,182,248,40,421,423
	19	19,244,415,424,437
	20	20,287,410,433,435
	216	216
	265	265
	402	402,403,404
	408	408
GGH3	H	ts1,ts2,1,100,101,104,105,107,108,109,110,113,114,115,119,120,122,124,128,129,130,131,135,136,138,140,141,142,147,149,150,152,153,154,155,157,158,160,162,166,167,171,172,175,177,178,180,181,182,183,184,185,187,188,189,190,191,192,193,194,195,196,197,198,2,200,202,206,213,214,218,219,221,222,224,226,227,228,229,232,233,235,236,237,238,240,243,245,246,249,250,251,252,253,254,255,256,257,259,26,260,261,263,264,268,269,27,270,271,273,274,275,277,278,28,280,281,282,283,284,286,288,289,29,290,291,293,294,297,298,3,301,303,310,313,317,318,319,32,320,321,35,4,41,42,423,426,427,428,43,432,436,44,45,46,49,53,56,57,6,61,63,64,65,66,67,69,70,71,72,74,75,78,8,80,81,82,83,87,88,9,90,91,92,93,95,98,99
	10	10,102,103,106,111,112,117,125,126,132,133,137,139,143,151,159,161,163,165,168,169,170,173,174,176,179,186,201,204,205,209,210,211,212,215,220,225,230,231,234,239,241,242,258,272,276,285,292,307,308,309,311,314,400,401,47,5,51,52,54,55,58,59,60,62,73,76,85,86,89,94
	11	11,305,306
	116	116,12,121,123,127,13,14,156,16,164,17,18,20,21,22,23,24,244,247,248,262,265,266,267,287,296,300,302,304,312,315,316,33,37,39,40,409,410,411,412,414,415,416,417,419,421,422,424,425,433,434,435,437,438,439,441,48,68,7,77,79,84,97
	118	118,144
	134	134
	145	145,146,148,15,223,295,299,34,403,413,418,420,430,50,96
	19	19
	199	199,207,208,217
	203	203
	216	216
	25	25
	30	30,31,36
	145	145

	38	38
	402	402,404,407,408
	429	429
	431	431
	440	440

Annexure 3.1

Accession names and codes for 360 genomes

Individual code	Group	TGRC	PI CGN#	EA #	Name	Botanical variety	Categories	Origin	Province/Site
TS-14	PIM	LA1547	-	-	Chota to EI Angel	S. pimpinellifolium	Wild species	Ecuador	Carchi
TS-15	PIM	LA2093	-	-		S. pimpinellifolium	Wild species	Ecuador	El Oro
TS-16	PIM	LA1246	PI365912	-		S. pimpinellifolium	Wild species	Ecuador	Loja
TS-17	PIM	LA0373	-	-	Culebras	S. pimpinellifolium	Wild species	Peru	Ancash
TS-18	PIM	LA1579	-	-		S. pimpinellifolium	Wild species	Peru	Lambayeque
TS-19	PIM	LA1589	PI407545	EA01467		S. pimpinellifolium	Wild species	Peru	La Libertad
TS-20	PIM	LA0442	-	-	Sechin	S. pimpinellifolium	Wild species	Peru	Ancash
TS-21	PIM	LA1375	PI365967	-		S. pimpinellifolium	Wild species	Peru	Lima
TS-22	PIM	LA1269	PI365957	-	Pisiquillo	S. pimpinellifolium	Wild species	Peru	Lima
TS-23	PIM	LA1521	-	-	EI Pinon, Asia	S. pimpinellifolium	Wild species	Peru	Lima
TS-24	PIM	-	-	-		S. pimpinellifolium	Wild species		
TS-50	PIM	LA0417	-	EA00565		S. pimpinellifolium	Wild species	Ecuador	Guayas
TS-77	PIM	LA1237	PI365910	-	Atacames	S. pimpinellifolium	Wild species	Ecuador	Esmeraldas
TS-79	PIM	LA1924	-	-	Piedras Gordas	S. pimpinellifolium	Wild species	Peru	Ica
TS-92	PIM	LA1582	PI407539	-	Punto Cuatro	S. pimpinellifolium	Wild species	Peru	Lambayeque
TS-123	PIM	LA0722	-	-	Trujillo	S. pimpinellifolium	Wild species	Peru	La Libertad
TS-124	PIM	LA1245	PI365911	-	Santa Rosa	S. pimpinellifolium	Wild species	Ecuador	El Oro
TS-144	PIM	LA0411	PI251319	-	Pichilingue	S. pimpinellifolium	Wild species	Ecuador	Los Rios
TS-145	PIM	LA1617	-	-	Tumbes south	S. pimpinellifolium	Wild species	Peru	Tumbes
TS-156	PIM	LA2181	-	-	Balsa Huaico	S. pimpinellifolium	Wild species	Peru	Cajamarca
TS-164	PIM	LA1584	PI407541	-	Jayanca to La Vina	S. pimpinellifolium	Wild species	Peru	Lambayeque
TS-182	PIM	LA2183	-	-	Corral Quemado	S. pimpinellifolium	Wild species	Peru	Amazonas
TS-222	PIM	-	-	-	Wva 700	S. pimpinellifolium	Wild species		
TS-244	PIM	LA1578	-	-	Santa Marta	S. pimpinellifolium	Wild species	Peru	La Libertad
TS-265	PIM	LA0400	-	-	Hacienda Buenos Aires	S. pimpinellifolium	Wild species	Peru	Piura
TS-266	PIM	LA1478	-	-	Santo Tome	S. pimpinellifolium	Wild species	Peru	Piura
TS-267*	PIM	LA2660	-	-	San Ignacia de Moxos	S. pimpinellifolium	Wild species	Bolivia	Beni
TS-291	PIM	LA1589	PI407545	EA01467	Viru to Galumga	S. pimpinellifolium	Wild species	Peru	La Libertad
TS-410	PIM	-	PI370093	-		S. pimpinellifolium	Wild species	Canada	
TS-411	PIM	LA0480	-	-		S. pimpinellifolium	Wild species	Peru	Ica
TS-412	PIM	LA0722	-	-		S. pimpinellifolium	Wild species	Peru	La Libertad
TS-413	PIM	LA1242	-	-		S. pimpinellifolium	Wild species	Ecuador	Guayas
TS-414	PIM	LA1341	PI379020	-		S. pimpinellifolium	Wild species	Peru	Lima
TS-415	PIM	LA1596	PI407552	-		S. pimpinellifolium	Wild species	Peru	Ancash
TS-416	PIM	LA1847	-	-		S. pimpinellifolium	Wild species		
TS-417	PIM	LA1933	-	-		S. pimpinellifolium	Wild species	Peru	Arequipa
TS-418	PIM	LA2147	-	-		S. pimpinellifolium	Wild species	Peru	Cajamarca
TS-419	PIM	LA2173	-	-		S. pimpinellifolium	Wild species	Peru	Cajamarca
TS-420	PIM	LA2184	-	-		S. pimpinellifolium	Wild species	Peru	Amazonas

TS-421	PIM	LA2187	-	-		S. pimpinellifolium	Wild species	Peru	Amazonas
TS-422	PIM	LA2425	-	-		S. pimpinellifolium	Wild species		
TS-424	PIM	-	PI126947	-		S. pimpinellifolium	Wild species	Peru	
TS-425	PIM	-	PI126925	-		S. pimpinellifolium	Wild species	Peru	
TS-429	PIM	-	PI126954	-		S. pimpinellifolium	Wild species	Peru	
TS-432	PIM	-	PI270449	-		S. pimpinellifolium	Wild species	Mexico	
TS-433	PIM	-	PI370093	-		S. pimpinellifolium	Wild species	Canada	
TS-434	PIM	LA1591	PI407547	-	SAL1871	S. pimpinellifolium	Wild species	Peru	La Libertad
TS-435	PIM	LA1595	PI407551	-	SAL1875	S. pimpinellifolium	Wild species	Peru	Ancash
TS-437	PIM	LA1578	-	-	SAL1858	S. pimpinellifolium	Wild species	Peru	La Libertad
TS-438	PIM	-	-	-	CN7542	S. pimpinellifolium	Wild species		
TS-439	PIM	LA2656	PI503524	-		S. pimpinellifolium	Wild species	Peru	Tumbes
TS-440	PIM	LA2857	-	-		S. pimpinellifolium	Wild species	Ecuador	Galapagos Islands
TS-441	PIM	LA4431	-	-		S. pimpinellifolium	Wild species		
TS-13	CER	-	-	-	Pime-Bruce	S. pimpinellifolium	Wild species		
TS-25	CER	-	-	-	Clémentine	S. lycopersicum var cerasiforme	cultivar		
TS-26	CER	-	-	-		S. lycopersicum var cerasiforme	cultivar	Argentina	
TS-27	CER	-	-	-	N135 Green Gage	S. lycopersicum var cerasiforme	cultivar	Deutchland	
TS-28	CER	-	-	-	N 347 Yablochnyi	S. lycopersicum var cerasiforme		Mexico	
TS-29	CER	-	-	-		S. lycopersicum var cerasiforme	cultivar	India	
TS-30	CER	LA1204	-	-		S. lycopersicum var cerasiforme	Wild species	Guatemala	Quetzaltenango
TS-31	CER	LA1464	-	-		S. lycopersicum var cerasiforme	Wild species	Honduras	El Progreso, Yoro
TS-32	CER	LA0172	-	-	Santa Cruz	S. lycopersicum var cerasiforme	Latin American cultivar	Bolivia	Santa Cruz
TS-33	CER	LA2137	-	-		S. lycopersicum var cerasiforme	Wild species	Ecuador	Morona-Santiago
TS-34	CER	LA2675	-	-		S. lycopersicum var cerasiforme	Wild species	Peru	Puno
TS-35	CER	-	-	-	tomate Richters	S. lycopersicum var cerasiforme	uncertain		
TS-36	CER	-	PI187002-1	-		S. lycopersicum var cerasiforme	cultivar	Guatemala	
TS-37	CER	-	-	-	L. pimpinellifolium atypique, site 10 (F300045)	S. lycopersicum var cerasiforme	Wild Species	Peru	
TS-38	CER	-	PI129088	-		S. lycopersicum var cerasiforme	cultivar	Colombia	
TS-39	CER	-	-	-	Cerise Gold	S. lycopersicum var cerasiforme	cultivar		
TS-40	CER	-	-	-	Cerise VFNT	S. lycopersicum var cerasiforme	cultivar		
TS-53	CER	LA2095	-	-		S. lycopersicum var cerasiforme	Wild species	Ecuador	Loja
TS-54	CER	-	-	-	cerise rose	S. lycopersicum var cerasiforme	cultivar		
TS-56	CER	LA1320	PI365923	-		S. lycopersicum var cerasiforme	Wild species	Peru	Apurimac
TS-57	CER	LA1307	PI378998	-		S. lycopersicum var cerasiforme	Wild species	Peru	Ayacucho
TS-61	CER	LA2670	-	-	Huayvaruni	S. lycopersicum var cerasiforme	Wild species	Peru	Puno
TS-62	CER	-	-	EA02959				Argentina	
TS-63	CER	-	-	-					
TS-64	CER	-	-	EA03525			Cocktail tomato	Russia	
TS-65	CER	LA1482	-	-		S. lycopersicum var cerasiforme	Wild species	Malaysia	

TS-66	CER	LA1388	PI379009	-	H707	S. lycopersicum var cerasiforme	Wild species	Peru	Junin
TS-67	CER	-	-	EA02304			Cocktail tomato	EL SALVADOR	
TS-70	CER	-	-	-	Pyriforme	S. lycopersicum var cerasiforme	cultivar		
TS-71	CER	-	-	-	Ohmiya SunCerise	S. lycopersicum var cerasiforme	cultivar		
TS-72	CER	-	-	-	Principe Borghese		Vintage cultivar	Italy	
TS-75	CER	-	-	-	Poire jaune	S. lycopersicum var cerasiforme	cultivar		
TS-83*	CER	-	-	EA01953		S. lycopersicum var cerasiforme	cultivar	ZAIRE	
TS-84	CER	-	-	-					
TS-87	CER	LA1701	-	-	Trujillo	S. lycopersicum var cerasiforme	Latin American cultivar; Wild species	Peru	La Libertad
TS-91	CER	-	-	-	N 2759 Enano	S. lycopersicum var cerasiforme	cultivar	Argentina	
TS-94	CER	-	-	-	Farthest North	S. lycopersicum var cerasiforme	cultivar	France	
TS-96	CER	LA1456	-	-		S. lycopersicum var cerasiforme	Wild species	Mexico	Vera Cruz
TS-97	CER	LA0154	-	-	Tiny tim	S. lycopersicum var cerasiforme		France	
TS-98	CER	LA4355	-	-	Gold Nugget	S. lycopersicum var cerasiforme	cultivar	United States	
TS-99	CER	-	-	-	Celsior	S. lycopersicum var cerasiforme			
TS-105	CER	-	-	EA01448		S. lycopersicum var cerasiforme	Wild species	Costa Rica	
TS-106	CER	-	-	-		S. lycopersicum var cerasiforme	cultivar	Costa Rica	
TS-107	CER	LA1542	-	-	Turrialba	S. lycopersicum var cerasiforme	Wild species	Costa Rica	Costa Rica
TS-109	CER	-	-	EA02660			Cocktail tomato		
TS-116	CER	-	-	-	N1565	S. lycopersicum var cerasiforme	cultivar	Peru	
TS-118	CER	LA0292	-	-	Santa Cruz, Galapagos	S. lycopersicum var cerasiforme	Latin American cultivar;Wild species	Ecuador	Galapagos Islands
TS-119*	CER	-	-	-	Wva 106	S. lycopersicum var cerasiforme	Wild species	France	
TS-120	CER	-	-	EA01356				unknown	
TS-129	CER	LA2845	-	-	Moyobamba	S. lycopersicum var cerasiforme	Latin American cultivar; Wild species	Peru	San Martin
TS-131*	CER	LA1162	-	-	Cuba Plum	S. lycopersicum var cerasiforme	cultivar	Cuba	un
TS-134	CER	LA1429	-	-	La Estancilla	S. lycopersicum var cerasiforme	Wild species	Ecuador	Manabi
TS-138	CER	-	-	-	Malintka 101		cultivar	Italy	
TS-148	CER	LA1323	PI365924	-	Pfacchayoc	S. lycopersicum var cerasiforme	Wild species	Peru	Cusco
TS-149	CER	LA1425	PI379060	-	Villa Hermosa	S. lycopersicum var cerasiforme	Latin American cultivar	Colombia	Cauca
TS-154	CER	LA1623	-	-	Muna	S. lycopersicum var cerasiforme	Wild species	Mexico	Yucatan
TS-158	CER	LA2626	-	EA01541	PE-67	S. lycopersicum var cerasiforme	Wild species	Peru	Cusco
TS-165	CER	LA1218	-	-	Veracruz	S. lycopersicum var cerasiforme	Latin American cultivar	Mexico	Vera Cruz
TS-181	CER	LA1457	-	EA01438		S. lycopersicum var cerasiforme	Wild species	Mexico	
TS-187	CER	-	-	-	Monplaisir	S. lycopersicum var cerasiforme	cultivar		
TS-189	CER	-	-	-	Nagcarlang		cultivar	Philippines	
TS-202	CER	LA2402	-	-		S. lycopersicum var cerasiforme	Latin American cultivar	Brazil	Santa Catarina
TS-205	CER	-	-	-	8 bis	S. lycopersicum var cerasiforme	cultivar		
TS-209	CER	-	-	-	Da serbo		landrace	Italy	
TS-213	CER	-	-	EA01784				Italy	

TS-216	CER	-	-	-	Phyra	S. lycopersicum var cerasiforme	cultivar		
TS-219	CER	-	-	-	Mobalcon		landrace	Greece	
TS-221	CER	LA1569	-	-	Jalapa	S. lycopersicum var cerasiforme	Wild species	Mexico	Vera Cruz
TS-223	CER	LA2675	-	EA01557		S. lycopersicum var cerasiforme	Wild species	Peru	Puno
TS-227	CER	-	-	-	Atom	S. lycopersicum var cerasiforme	cultivar		
TS-229	CER	LA1620	-	-		S. lycopersicum var cerasiforme	Wild species	Brazil	Bahia
TS-230	CER	-	-	-	Cerise Ildi	S. lycopersicum var cerasiforme	cultivar		
TS-231	CER	-	-	-	L 285	S. lycopersicum var cerasiforme	cultivar		
TS-233	CER	LA1218	-	EA00602		S. lycopersicum var cerasiforme	Latin American cultivar	Mexico	Vera Cruz
TS-238	CER	LA1228	PI379047	-	Macas	S. lycopersicum var cerasiforme	Wild species	Ecuador	Morona-Santiago
TS-240	CER	LA3136	-	-	Arroyo Rico	S. lycopersicum var cerasiforme	Wild species	Cuba	Holguin
TS-243	CER	-	-	EA04228				Mexico	
TS-247	CER	-	-	-	N 795 Pescio	S. lycopersicum var cerasiforme	cultivar	Peru	
TS-248	CER	-	-	-					
TS-250	CER	-	-	EA03539				Russia	
TS-252	CER	-	-	EA01802			cocktail tomato	Italy	
TS-254	CER	-	-	EA02979					
TS-257	CER	-	-	-	Marpha N2	S. lycopersicum var cerasiforme	cultivar	France	
TS-258	CER	-	-	-	Minibel	S. lycopersicum var cerasiforme	cultivar		
TS-260	CER	-	-	-	Cerise Orange dUzès	S. lycopersicum var cerasiforme	cultivar	France	
TS-262	CER	LA1247	PI379058	EA01403		S. lycopersicum var cerasiforme	Wild species	Ecuador	Loja
TS-271	CER	-	-	-	Linosa		landrace	Italy	
TS-273	CER	LA2640	-	EA01545	PE-63	S. lycopersicum var cerasiforme	Wild species	Peru	Apurimac
TS-280	CER	-	-	-	Cerise du sud ouest N 2	S. lycopersicum var cerasiforme	Wild species		
TS-281	CER	LA1286	PI365920	-	San Martin de Pangoa	S. lycopersicum var cerasiforme	Wild species	Peru	Junin
TS-283	CER	-	-	-	Cisterno	S. lycopersicum var cerasiforme	cultivar		
TS-284	CER	-	-	-	Cerazjnho		landrace	Brasil	
TS-286	CER	-	-	-	Allungato piccolo		landrace	Brasil	
TS-287	CER	-	-	EA00915			Cocktail tomato	Unknown	
TS-289*	CER	LA1244	PI379044	EA00610		S. lycopersicum var cerasiforme	Latin American cultivar	Ecuador	Guayas
TS-290	CER	-	-	-	N 2257 Dikorastushii...	S. lycopersicum var cerasiforme	cultivar	Peru	
TS-294	CER	-	-	-	N 933	S. lycopersicum var cerasiforme	cultivar	Uzbekistan	
TS-295*	CER	LA0417	-	-	Puna	S. lycopersicum var cerasiforme	Latin American cultivar; Wild species	Ecuador	Guayas
TS-298	CER	LA2308	-	-	San Francisco	S. lycopersicum var cerasiforme	Wild species	Peru	San Martin
TS-299	CER	LA2131	-	-	Bomboiza	S. lycopersicum var cerasiforme	Wild species	Ecuador	Zamora-Chinchipe
TS-300	CER	LA1231	PI379049	-	Tena	S. lycopersicum var cerasiforme	Wild species	Ecuador	Napo
TS-301	CER	LA2688	-	-	Santa Cruz near Shintuyo	S. lycopersicum var cerasiforme	Wild species	Peru	Madre de Dios
TS-302	CER	LA1543	-	-	Upper Parana	S. lycopersicum var cerasiforme	Wild species	Brazil	
TS-303	CER	LA1461	-	-	Los Banos	S. lycopersicum var cerasiforme	Wild species	Phillipines	Los Banos
TS-304*	CER	LA2307	-	-	Tarapoto	S. lycopersicum var cerasiforme	Latin American cultivar	Peru	San Martin

TS-423*	CER	LA2840	-	-		S. lycopersicum var cerasiforme	Wild species	Peru	Amazonas
TS-426*	CER	LA1263	PI365918	-	SAL345	S. lycopersicum var cerasiforme	cultivar	Ecuador	Guayas
TS-427*	CER	-	PI126933	-		S. lycopersicum var cerasiforme	cultivar	Peru	
TS-428*	CER	-	PI126953	-		S. lycopersicum var cerasiforme	cultivar	Peru	
TS-430*	CER	-	PI127807	-	BL 587-S	S. lycopersicum var cerasiforme	cultivar	Peru	
TS-431*	CER	-	PI127807	-	BL 587	S. lycopersicum var cerasiforme	cultivar	Peru	
TS-436*	CER	-	-	-	BL 359	S. lycopersicum var cerasiforme	cultivar		
TS-1	BIG	LA0490	-	-	VF-36	S. lycopersicum	Vintage Processing	United States	
TS-2	BIG	LA2706	-	-	Moneymaker	S. lycopersicum	Vintage Fresh Market	England	
TS-3#	BIG	LA3475	-	-	M-82	S. lycopersicum	Modern Processing	Israel	
TS-4	BIG	LA3856	-	-	Hawaii 7998	S. lycopersicum	Inbreed line		
TS-5	BIG	LA2711	-	-	Edkawi	S. lycopersicum	Vintage Fresh Market	Egypt	
TS-6	BIG	LA3008	-	-	San Marzano	S. lycopersicum	Vintage Processing	Italy	
TS-7	BIG	LA3911	-	-	Micro-Tom	S. lycopersicum	Modern Fresh Market	United States	
TS-8#	BIG	LA4024	-	-	E-6203	S. lycopersicum	Modern Processing	United States	
TS-9	BIG	LA2838A	-	-	Ailsa Craig	S. lycopersicum	Vintage Fresh Market	England	
TS-10	BIG	LA0502	-	-	Marglobe	S. lycopersicum	Vintage Fresh Market	United States	
TS-11	BIG	-	-	-	KR2	S. lycopersicum	Modern Fresh Market		
TS-12	BIG	-	-	-	yoku improvement	S. lycopersicum	Modern Fresh Market		
TS-41	BIG	-	-	EA02435		S. lycopersicum	Cocktail tomato	EL SALVADOR	
TS-42	BIG	-	PI345565	EA05808		S. lycopersicum	Processing tomato	Russia	
TS-43	BIG	-	-	EA00840	Moneymaker		Fresh Market		
TS-44	BIG	-	-	-	A pera abruzzese		Landrace	Italy	
TS-45#	BIG	-	PI303718	EA05578		S. lycopersicum	Processing tomato	United States	
TS-46	BIG	-	-	EA01237			Processing tomato	unknown	
TS-47	BIG	-	-	EA01960		S. lycopersicum	Processing tomato	Russia	
TS-48	BIG	LA0146	-	-	Mexico City	S. lycopersicum	Landrace/Latin American cultivar	Mexico	Mercado
TS-49	BIG	LA3238	-	-	Earliana	S. lycopersicum	Vintage Processing	United States	
TS-51	BIG	-	-	-					
TS-52	BIG	-	-	-	05-4126 (97-49-2)		cultivar	unknown	
TS-55	BIG	-	-	EA00448				Unknown	
TS-58	BIG	-	-	EA03577		S. lycopersicum	Processing tomato	Russia	
TS-59#	BIG	-	-	EA02898			Processing tomato	Bulgaria	
TS-60	BIG	LA2009	-	-	New Yorker	S. lycopersicum	Vintage Fresh Market	United States	
TS-68	BIG	LA0395	-	-	Chiclayo	S. lycopersicum	Latin American cultivar	Peru	Lambayeque
TS-69	BIG	LA1459	-	-	Huachinango	S. lycopersicum	Latin American cultivar	Mexico	
TS-73†	BIG	-	-	-	Quarantino			Italy	
TS-74	BIG	-	-	-	N 739	S. lycopersicum	Fresh Market	Peru	
TS-76	BIG	-	-	EA01230			Processing tomato	Unknown	
TS-78#	BIG	-	-	EA02895		S. lycopersicum	Processing tomato	Russia	
TS-80	BIG	-	-	EA01020			Processing tomato	unknown	

TS-81	BIG	-	-	EA02761			Processing tomato	Russia	
TS-82	BIG	-	-	EA03274			Processing tomato	Spain	
TS-85#	BIG	-	-	-					
TS-86	BIG	-	-	EA01684				Italy	
TS-88	BIG	-	-	EA01804		S. lycopersicum	Cocktail tomato	Italy	
TS-89	BIG	-	-	EA01185			Processing tomato	Unknown	
TS-90	BIG	-	-	EA02753		S. lycopersicum	Cocktail tomato	Russia	
TS-93	BIG	-	-	EA01002			Processing tomato	Unknown	
TS-95	BIG	-	-	-	Moneymaker	S. lycopersicum	Fresh Market		
TS-100	BIG	-	-	EA03456		S. lycopersicum	Processing	Italy	
TS-101	BIG	-	-	EA00369		S. lycopersicum		Unknown	
TS-102	BIG	-	-	EA03673		S. lycopersicum	Processing tomato	Russia	
TS-103	BIG	-	-	EA00389		S. lycopersicum		Unknown	
TS-104	BIG	-	-	EA01756		S. lycopersicum	Processing tomato	Italy	
TS-108#	BIG	-	-	EA01989	Puno I	S. lycopersicum	Processing tomato	Peru	
TS-110	BIG	-	PI93302	EA04243		S. lycopersicum		CHINA	
TS-111	BIG	-	-	EA01270			Processing tomato	Unknown	
TS-112	BIG	-	-	EA03083		S. lycopersicum	Processing tomato	Russia	
TS-113	BIG	-	-	EA01198			Processing tomato	Unknown	
TS-114	BIG	-	-	EA01982		S. lycopersicum	Processing tomato	Russia	
TS-115	BIG	-	-	EA03426			Processing tomato	unknown	
TS-117	BIG	-	-	-	Scatolone di bolsena		Landrace	Italy	
TS-121	BIG	LA3846	-	-	NC EBR-6	S. lycopersicum	Modern Fresh Market	United States	
TS-122	BIG	LA1090	-	-	Rutgers	S. lycopersicum	Vintage Fresh Market	United States	
TS-125	BIG	-	-	EA00422			Processing tomato	Unknown	
TS-126	BIG	-	-	EA01903		S. lycopersicum		United States	
TS-127	BIG	LA0113	-	-	Hacienda Calera	S. lycopersicum	Landrace/Latin American cultivar	Peru	La Libertad
TS-128	BIG	LA0012	-	-	Pearson	S. lycopersicum	Vintage Processing	United States	
TS-130	BIG	LA2413	-	-	Severianin	S. lycopersicum	Modern Fresh Market	Russia	
TS-132	BIG	LA3903	-	-	Primabel	S. lycopersicum	Vintage Fresh Market	France	
TS-133#	BIG	LA3528	-	-	Peto95-43	S. lycopersicum	Modern Processing	United States	
TS-135	BIG	LA0466	PI 258469	-	Hacienda Rosario	S. lycopersicum	Landrace/Latin American cultivar	Chile	Arica and Parinacota
TS-136	BIG	-	-	-	Vito			Italy	
TS-137	BIG	-	-	-	Spagnoletta		Landrace	Italy	
TS-139	BIG	-	-	-	Red Setter		Vintage cultivar	USA	
TS-140	BIG	-	-	-	149-77		cultivar	Italy	
TS-141	BIG	-	-	-	Saladette		cultivar	Italy	
TS-142	BIG	-	-	-	Roma		Vintage cultivar	Italy	
TS-143	BIG	LA4025	-	-	Florida 7547	S. lycopersicum	Modern Fresh Market	United States	
TS-147#	BIG	-	-	-					
TS-150	BIG	LA2285	-	-	Tarapoto	S. lycopersicum	Landrace/Latin American cultivar	Peru	San Martin
TS-151	BIG	LA2399	-	-	T-5	S. lycopersicum	Modern Fresh Market	United States	
TS-152	BIG	LA1021	-	-	Santa Cruz B	S. lycopersicum	Landrace/Latin American cultivar	Brazil	Sao Paulo

TS-153	BIG	LA1544	-	-	Xol Languna	S. lycopersicum	Landrace/Latin American cultivar	Mexico	Campeche
TS-155	BIG	LA0533	-	-	Condine Red	S. lycopersicum	Vintage Fresh Market,Monogenic	Germany	
TS-157	BIG	-	-	EA03648		S. lycopersicum	Processing tomato	Russia	
TS-159	BIG	-	-	EA03028			Processing tomato	Russia	
TS-160	BIG	-	-	EA03533		S. lycopersicum	Processing tomato	Russia	
TS-161	BIG	-	-	EA02586		S. lycopersicum	Processing tomato	Russia	
TS-162#	BIG	-	-	EA03463		S. lycopersicum	Processing tomato	Azerbaijan	
TS-163	BIG	LA1504	-	-	Marmande	S. lycopersicum	Vintage Fresh Market	France	
TS-166	BIG	LA0404	-	-	Piura	S. lycopersicum	Landrace/Latin American cultivar	Peru	Piura
TS-167	BIG	LA0147	-	-	Tegucigalpa	S. lycopersicum	Landrace/Latin American cultivar	Honduras	Tegucigalpa mercado
TS-168#	BIG	-	-	-	Da appendere		Landrace	Italy	
TS-169	BIG	-	-	-	Cuor di bue di Albenga		Landrace	Italy	
TS-170	BIG	-	-	-	Cuor di bue		Landrace	Italy	
TS-171#	BIG	LA1706	-	-	UC-82	S. lycopersicum	Modern Processing	United States	
TS-172	BIG	-	PI280060	EA05480		S. lycopersicum	Processing tomato	Canada	
TS-173	BIG	-	-	EA03611		S. lycopersicum		Russia	
TS-174	BIG	-	-	EA00304			Processing tomato	Unknown	
TS-175	BIG	-	-	EA03586		S. lycopersicum	Processing tomato	Russia	
TS-176	BIG	-	-	EA02669			Processing tomato	Tunisia	
TS-177	BIG	-	-	EA01155			Processing tomato	Unknown	
TS-178	BIG	-	PI513036	EA06485		S. lycopersicum	Processing tomato	Spain	
TS-179	BIG	-	-	EA01027			Processing tomato	Unknown	
TS-180	BIG	-	-	EA02728			Processing tomato	Russia	
TS-183	BIG	-	-	EA02764			Processing tomato	Germany	
TS-184†	BIG	LA2283	-	-	Tarapoto	S. lycopersicum		Peru	San Martin
TS-185	BIG	LA4347	-	-	B-L-35	S. lycopersicum	Vintage Fresh Market	Spain	Mallorca
TS-186#	BIG	LA3214	-	-	Rowpac	S. lycopersicum	Modern Processing	United States	
TS-188	BIG	LA1506	-	-	Stone	S. lycopersicum	Vintage Fresh Market	United States	
TS-190#	BIG	-	-	-	Santa Chiara		cultivar	Italy	
TS-191	BIG	-	-	-	Francescano		Landrace	Italy	
TS-192	BIG	-	-	-	Severianin		Vintage cultivar	Russia	
TS-193	BIG	-	-	-	Pantano dArdea		Landrace	Italy	
TS-194	BIG	-	-	-					
TS-195	BIG	LA0516	-	-	Ace	S. lycopersicum	Vintage Fresh Market	United States	
TS-196	BIG	-	-	EA00240	N020212		Processing tomato	unknown	
TS-197	BIG	-	-	-	Libanese		Landrace	Libanon	
TS-198	BIG	-	-	EA00512					
TS-200	BIG	LA3320	-	-	Hot set	S. lycopersicum	cultivar	Italy	
TS-201	BIG	LA1210	-	-	San Salvador	S. lycopersicum	Landrace/Latin American cultivar	El Salvador	San Salvador
TS-203	BIG	-	-	-	Bell pepper-like		Landrace	Italy	
TS-204	BIG	LA3840	-	-	Florida 7060	S. lycopersicum	Modern Fresh Market	United States	
TS-206	BIG	LA0089	-	-	Prince Borghese	S. lycopersicum	Vintage Fresh Market	Italy	

TS-210	BIG	LA3625	-	-	NC 265-1 (93)-3-3	S. lycopersicum	Modern Fresh Market	United States	
TS-211	BIG	LA4354	-	-	NC 84173	S. lycopersicum	Modern Fresh Market	United States	
TS-212	BIG	LA3242	-	-	Flora-Dade	S. lycopersicum	Modern Fresh Market	United States	
TS-214	BIG	-	-	-	Panama		Landrace	Panama	
TS-215	BIG	-	-	-	Vrbikanske Nizke		cultivar	Bulgaria/Italy	
TS-218	BIG	-	-	-	Santa Clara 5800		cultivar	Brasil	
TS-220	BIG	-	-	-	Barnaulski Konservnyi		cultivar	Russia	
TS-224	BIG	LA0410	PI 258474	-	Guayaquil	S. lycopersicum	Landrace/Latin American cultivar	Ecuador	Guayas
TS-225	BIG	-	PI330336	EA05747		S. lycopersicum	Processing tomato	Japan	
TS-226	BIG	-	-	-	Microtom	S. lycopersicum	cultivar	France	
TS-228#	BIG	-	-	-	M-82	S. lycopersicum	cultivar		
TS-232	BIG	-	-	EA00951			Processing	unknown	
TS-234	BIG	-	-	EA01371			Processing tomato	Unknown	
TS-235	BIG	-	-	EA00892			Processing tomato	Guatemala	
TS-236	BIG	-	-	EA02732			Processing tomato	Russian federation	
TS-237	BIG	LA3243	-	-	Platense	S. lycopersicum	Vintage Fresh Market	Argentina	
TS-239	BIG	LA3845	-	-	NC EBR-5	S. lycopersicum	Modern Fresh Market	United States	
TS-241	BIG	LA0126	-	-	Quito	S. lycopersicum	Latin American cultivar	Ecuador	Pichincha
TS-242	BIG	LA0134C	-	-	Ayacucho	S. lycopersicum	Latin American cultivar	Peru	Ayacucho
TS-245#	BIG	-	-	EA03126		S. lycopersicum	Processing tomato	CZECH REPUBLIC	
TS-246	BIG	-	-	EA00983			Processing tomato	Unknown	
TS-249	BIG	LA1462	-	-	Merida	S. lycopersicum	Landrace/Latin American cultivar	Mexico	
TS-251	BIG	-	PI647249	EA04001		S. lycopersicum		Spain	
TS-253#	BIG	LA4345	-	-	Heinz 1706-BG	S. lycopersicum	Modern Processing	United States	
TS-255	BIG	-	-	EA03002			Processing tomato	Russian federation	
TS-256	BIG	LA2260		EA00744		S. lycopersicum	Latin American cultivar	Peru	San Martin
TS-259	BIG	-	-	EA01712			Processing tomato	Italy	
TS-261*	BIG	LA1511	-	EA01444		S. lycopersicum	Wild species	Brazil	Minas Gerais
TS-263#	BIG	LA3343	-	-	Rio Grande	S. lycopersicum	Processing tomato	United States	
TS-264	BIG	LA0025	-	-	King Humbert #1	S. lycopersicum	Vintage Fresh Market	Italy	
TS-268	BIG	-	-	EA01915		S. lycopersicum	cultivar	Unknown	
TS-269	BIG	-	-	-	Canestrino		Landrace	Italy	
TS-270	BIG	-	-	EA03174				Bulgaria	
TS-272#	BIG	-	-	EA06878			Processing tomato	unknown	
TS-274#	BIG	-	-	EA03613			Cocktail/Processing tomato	Russian federation	
TS-275	BIG	-	-	EA01049			Processing tomato	unknown	
TS-276	BIG	-	-	EA03650			Cocktail/Processing tomato	Moldova	
TS-277#	BIG	LA3144	-	-	Hunt100	S. lycopersicum	Modern Processing	United States	
TS-278	BIG	LA0517	-	-	Early Santa Clara	S. lycopersicum	Vintage Processing	United States	
TS-282*	BIG	-	-	-	Ucra2		Landrace	Ukraine	
TS-285	BIG	-	PI303752	EA05612		S. lycopersicum	Processing tomato	United States	
TS-288	BIG	-	-	EA04236		S. lycopersicum	Processing tomato	CHINA	

TS-292#	BIG	-	-	EA06902		S. lycopersicum	Processing tomato	unknown	
TS-293	BIG	-	-	EA03439		S. lycopersicum	Cocktail/Processing tomato	unknown	
TS-296	BIG	-	-	-	Droplet	S. lycopersicum			
TS-297	BIG	-	PI291344	EA05550		S. lycopersicum		China	
TS-400	BIG	-	-	-		S. lycopersicum	inbred line	China	
TS-401	BIG	-	-	-		S. lycopersicum	inbred line	China	
TS-409	BIG	-	PI124161	-		S. lycopersicum	Landrace	Guatemala	
TS-305	F1	-	-	-		S. lycopersicum	Fresh Market		
TS-306	F1	-	-	-		S. lycopersicum	Fresh Market		
TS-307	F1	-	-	-		S. lycopersicum	Fresh Market		
TS-308	F1	-	-	-		S. lycopersicum	Fresh Market		
TS-309	F1	-	-	-		S. lycopersicum	Fresh Market		
TS-310	F1	-	-	-		S. lycopersicum	Fresh Market		
TS-311	F1	-	-	-		S. lycopersicum	Fresh Market		
TS-312	F1	-	-	-		S. lycopersicum	Fresh Market		
TS-313	F1	-	-	-		S. lycopersicum	Fresh Market		
TS-314	F1	-	-	-		S. lycopersicum	Fresh Market		
TS-315	F1	-	-	-		S. lycopersicum	Fresh Market		
TS-316	F1	-	-	-		S. lycopersicum	Fresh Market		
TS-317	F1	-	-	-		S. lycopersicum	Processing		
TS-318	F1	-	-	-		S. lycopersicum	Processing		
TS-319	F1	-	-	-		S. lycopersicum	Processing		
TS-320	F1	-	-	-		S. lycopersicum	Processing		
TS-321	F1	-	-	-		S. lycopersicum	Processing		
TS-146	wild	LA2133	-	-		S. neorickii		Peru	
TS-208	wild	LA0528	-	-	Santa Cruz: Academy Bay	S. galapagense		Ecuador	
TS-199	wild	LA0746	-	EA00587		S. cheesmaniae		Ecuador	
TS-207	wild	LA1037	-	-		S. cheesmaniae		Ecuador	Galapagos Islands
TS-217	wild	LA0429	-	EA00568		S. cheesmaniae		Ecuador	Galapagos Islands
TS-402	wild	-	PI126935	-		S. peruvianum		Peru	
TS-403	wild	-	PI128650	-		S. peruvianum		Chile	
TS-404	wild	-	PI128657	-		S. peruvianum		Peru	
TS-408	wild	LA1969	-	-		S. chilense		Peru	
TS-407	wild	-	PI247087	-		S. habrochaites		Ecuador	

HARNESSING NATURAL DIVERSITY IN TOMATO TO IMPROVE FOLATE LEVELS USING METABOLIC PROFILING

ORIGINALITY REPORT

%12	%7	%11	%3
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Tyagi, Kamal, Pallawi Upadhyaya, Supriya Sarma, Vajir Tamboli, Yellamaraju Sreelakshmi, and Rameshwar Sharma. "High performance liquid chromatography coupled to mass spectrometry for profiling and quantitative analysis of folate monoglutamates in tomato", Food Chemistry, 2015. Publication	%1
2	www.plantmethods.com Internet Source	%1
3	Mohan, Vijee, Soni Gupta, Sherinmol Thomas, Hanjabam Mickey, Chaitanya Charakana, Vineeta Singh Chauhan, Kapil Sharma, Rakesh Kumar, Kamal Tyagi, Supriya Sarma, Suresh Kumar Gupta, Himabindu Vasuki Kilambi, Sapana Nongmaithem, Alka Kumari, Prateek Gupta, Yellamaraju Sreelakshmi, and Rameshwar Sharma. "Tomato Fruits Show Wide Phenomic Diversity but Fruit Developmental Genes Show Low Genomic Diversity", PLoS	<%1

ONE, 2016.
Publication

4	agronomy.wisc.edu Internet Source	<%1
5	GUPTA, SURESH KUMAR, SULABHA SHARMA, PARANKUSAM SANTISREE, HIMABINDU VASUKI KILAMBI, KLAUS APPENROTH, YELLAMARAJU SREELAKSHMI, and RAMESHWAR SHARMA. "Complex and shifting interactions of phytochromes regulate fruit development in tomato : Phytochrome modulation of tomato ripening", Plant Cell & Environment, 2014. Publication	<%1
6	Arcot, J.. "Folate: methods of analysis", Trends in Food Science & Technology, 200506/07 Publication	<%1
7	R. D. de la Garza. "Folate biofortification in tomatoes by engineering the pteridine branch of folate synthesis", Proceedings of the National Academy of Sciences, 09/09/2004 Publication	<%1
8	Dieter Blancquaert. "Folates and Folic Acid: From Fundamental Research Toward Sustainable Health", Critical Reviews in Plant Sciences, 01/2010 Publication	<%1

9	www2.ifr.bbsrc.ac.uk Internet Source	<%1
10	Iniesta, M. Dolores, Darío Pérez-Conesa, Javier García-Alonso, Gaspar Ros, and M. Jesús Periago. "Folate Content in Tomato (Lycopersicon esculentum). Influence of Cultivar, Ripeness, Year of Harvest, and Pasteurization and Storage Temperatures", Journal of Agricultural and Food Chemistry, 2009. Publication	<%1
11	Aflitos, Saulo, Elio Schijlen, Hans de Jong, Dick de Ridder, Sandra Smit, Richard Finkers, Jun Wang, Gengyun Zhang, Ning Li, Likai Mao, Freek Bakker, Rob Dirks, Timo Breit, Barbara Gravendeel, Henk Huits, Darush Struss, Ruth Swanson-Wagner, Hans van Leeuwen, Roeland C.H.J. van Ham, Laia Fito, Laëtitia Guignier, Myrna Sevilla, Philippe Ellul, Eric Ganko, Arvind Kapur, Emannuel Reclus, Bernard de Geus, Henri van de Geest, Bas te Lintel Hekkert, Jan van Haarst, Lars Smits, Andries Koops, Gabino Sanchez-Perez, Adriaan W. van Heusden, Richard Visser, Zhiwu Quan, Jiumeng Min, Li Liao, Xiaoli Wang, Guangbiao Wang, Zhen Yue, Xinhua Yang, Na Xu, Eric Schranz, Erik Smets, Rutger Vos, Johan Rauwerda, Remco Ursem, Cees Schuit, Mike Kerns, Jan van den Berg, Wim Vriezen, Antoine Janssen,	<%1

Food Chemistry 179 (2015) 76–84

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

High performance liquid chromatography coupled to mass spectrometry for profiling and quantitative analysis of folate monoglutamates in tomato

Kamal Tyagi, Pallawi Upadhyaya, Supriya Sarma, Vajir Tamboli, Yellamaraju Sreelakshmi, Rameshwar Sharma *

Repository of Tomato Genomics Resources, Department of Plant Sciences, School of Life Sciences, University of Hyderabad, Hyderabad 500046, India

ARTICLE INFO

Article history:
Received 21 October 2014
Received in revised form 20 January 2015
Accepted 23 January 2015
Available online 31 January 2015

Keywords:
Tomato
Folates
HPLC–MS/MS
Folate extraction

ABSTRACT

Folates are essential micronutrients for animals as they play a major role in one carbon metabolism. Animals are unable to synthesize folates and obtain them from plant derived food. In the present study, a high performance liquid chromatography coupled to mass spectrometric (HPLC–MS/MS) method was developed for the high throughput screening and quantitative analysis of folate monoglutamates in tomato fruits. For folate extraction, several parameters were optimized including extraction conditions, pH range, amount of tri-enzyme and boiling time. After processing the extract was purified using ultra-filtration with 10 kDa membrane filter. The ultra-filtered extract was chromatographed on a RP Luna C18 column using gradient elution program. The method was validated by determining linearity, sensitivity and recovery. This method was successfully applied to folate estimation in spinach, capsicum, and garden pea and demonstrated that this method offers a versatile approach for accurate and fast determination of different folate monoglutamates in vegetables.

1. Introduction

Folate (folic acid) is one of the B groups of vitamins that are essential for the human health. The dietary insufficiency of folate can lead to several health related disorders such as megaloblastic anemia (Gough, Read, McCarthy, & Waters, 1963), exacerbation of cardiovascular disease (Kolb & Petrie, 2013) and some types of cancer (Blount et al., 1997). It is also required for the development of a healthy fetus and its deficiency affects formation of fetus' spinal cord and brain (Blom, Shaw, den Heijer, & Finnell, 2006). Folate cannot be synthesized by animals and therefore it is solely obtained from diet.

Plants are the major source of dietary folates. Green leafy vegetables, legumes and some fruits are among the richest sources of folate. In several countries the cereal based food products are mandatorily fortified with folic acid to prevent folate-related disorders. In addition, transgenic approaches have been used to increase the biosynthesis of folates in tomato fruits, potato tubers and rice grains to remove dietary constraints of low folate levels in diet (Blancquaert, De Steur, Gellynck, & Van Der Straeten, 2014).

Folate is comprised of a pterin moiety attached by a methylene bridge to para-amino benzoic acid, which is coupled to one or more glutamyl residues. *In vivo* folates exist as tetrahydrofolate (THF) and its derivatives (5-methyl, methylene, methenyl, or 10-formyl) which vary in oxidation states, single carbon substituents, and with a variable number of glutamyl residues (Rébeillé et al., 2006), which are collectively called – folate or vitamin B9. THF plays a key role in one-carbon transfer reactions in all living organisms participating in diverse metabolic reactions such as amino acid metabolism, pantothenate synthesis, purines and thymidylate synthesis etc.

For quantitative determination of total folates, microbiological assay is the most commonly used method which is also recommended by Association of Official Analytical Communities International (AOAC, 2000). The main weakness of this method is that it cannot distinguish diverse forms of folate present in food samples. Above limitation has been overcome by using chromatography-based methods allowing separation of different vitamers. The folate determination has been carried out using high-performance liquid chromatography (HPLC) coupled with UV (Pfeiffer, Rogers, & Gregory, 1997), electrochemical (Bagley & Selhub, 2000), and fluorescence detection (Ndaw, Bergaentzlé, Aoudé-Werner,

* Corresponding author. Tel.: +91 40 23010514; fax: +91 40 23010120.
E-mail addresses: tyagi.kamal6672@gmail.com (K. Tyagi), pravas43@gmail.com (P. Upadhyaya), supu.megha@gmail.com (S. Sarma), vajirchem@gmail.com (V. Tamboli), syellamaraju@gmail.com (Y. Sreelakshmi), rameshwar.sharma@gmail.com (R. Sharma).

http://dx.doi.org/10.1016/j.foodchem.2015.01.110
0308-8146/