Nirala Ramchiary

Identification and expression analysis of genes involved in vitamins biosynthesis and fruit ripening in Capsicum species

Thesis Submission

Thesis Submission

JNU

Document Details

Submission ID
trn:oid:::1:3347564740

Submission Date
Sep 22, 2025, 3:11 PM GMT+5:30

Download Date
Jan 6, 2026, 12:43 PM GMT+5:30

File Name
Thesis_-trunitin.pdf

File Size
8.0 MB

216 Pages

44,529 Words

233,700 Characters

6% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Quoted Text

Match Groups

170 Not Cited or Quoted 5%
Matches with neither in-text citation nor quotation marks

30 Missing Quotations 1%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

2% Internet sources
6% Publications
1% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **170** Not Cited or Quoted 5%
  Matches with neither in-text citation nor quotation marks
- **30** Missing Quotations 1%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

2% Internet sources
6% Publications
1% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Publication	
Meenakshi Dubey, Khushbu Islam, Abdul Rawoof, Neha Tiwari, Vandana Jaiswal, ...		3%
2	Publication	
Meenakshi Dubey, Vandana Jaiswal, Abdul Rawoof, Ajay Kumar et al. "Identificati...		<1%
3	Student papers	
Jawaharlal Nehru University (JNU)		<1%
4	Internet	
www.researchgate.net		<1%
5	Internet	
www.ncbi.nlm.nih.gov		<1%
6	Internet	
www.frontiersin.org		<1%
7	Publication	
Khushbu Islam, John Momo, Abdul Rawoof, Archa Vijay, V. K. Anusree, Ajay Kuma...		<1%
8	Internet	
www.alice.cnptia.embrapa.br		<1%
9	Internet	
www.jnu.ac.in		<1%
10	Publication	
Priyanka Gogoi, Akib Ali, Aditya Sarkar, Pankaj Bharali, Ratul Saikia. "Rhizobacteri...		<1%

11	Internet	
assets-eu.researchsquare.com		<1%
12	Internet	
autosdriveamerica.org		<1%
13	Publication	
Patel, Hemilkumar Shaileshbhai. "Design, Characterization and Biological Signific...		<1%
14	Publication	
"Advances in Plant Breeding Strategies: Vegetable Crops", Springer Science and B...		<1%
15	Publication	
John Momo, Ajay Kumar, Khushbu Islam, Ilyas Ahmad, Abdul Rawoof, Nirala Ram...		<1%
16	Internet	
acikbilim.yok.gov.tr		<1%
17	Internet	
download.bibis.ir		<1%
18	Publication	
Ilyas Ahmad, Abdul Rawoof, Priyanka, Khushbu Islam, John Momo, Thattantavid...		<1%
19	Internet	
ccsuniversity.ac.in		<1%
20	Publication	
Plant Breeding in the Omics Era, 2015.		<1%
21	Publication	
Sushil Satish Chhapekar, Vijaya Brahma, Abdul Rawoof, Nitin Kumar et al. "Transc...		<1%
22	Internet	
sabraojournal.org		<1%
23	Internet	
www.amity.edu		<1%
24	Publication	
Ajay Kumar, Sushil Kumar, Thattantavide Anju, Nirala Ramchiary. " Genetic, Epige...		<1%

25 Internet
dspace.uohyd.ac.in <1%

26 Internet
www.carnivoreconservation.org <1%

27 Internet
dducollegedu.ac.in <1%

28 Internet
usermanual.wiki <1%

29 Publication
Ilyas Ahmad, Abdul Rawoof, Khushbu Islam, John Momo, Nirala Ramchiary. "Iden... <1%

30 Publication
Pooja Moni Baruah, Debasish B. Krishnatreya, Kuntala Sarma Bordoloi, Sarvajeet ... <1%

31 Publication
Elena Rosa Martínez. "Characterization and Enhancement of Genetic Resources f... <1%

32 Publication
"The Capsicum Genome", Springer Nature, 2019 <1%

33 Student papers
University of Hong Kong <1%

34 Internet
bibliotecadigital.exactas.uba.ar <1%

35 Student papers
Higher Education Commission Pakistan <1%

36 Internet
www.science.gov <1%

37 Publication
Sun, Shuangxi. "Carbon Based Materials Synthesis and Characterization for 3D In... <1%

38 Internet
d-nb.info <1%

39 Internet
iris.unito.it <1%

40 Internet
journals.plos.org <1%

41 Internet
kb.psu.ac.th:8080 <1%

42 Internet
krishikosh.egranth.ac.in <1%

43 Internet
www.mmumullana.org <1%

44 Internet
www.tandfonline.com <1%

45 Publication
Miah, Md. Shipon. "Making Apples Blush: Understanding How the Combined Use ... <1%

46 Publication
Szymańska, Renata, Beatrycze Nowicka, Michał Gabruk, Sława Glińska, Sylwia Mic... <1%

47 Publication
Zhengkun Qiu, Ren Li, Shuaibin Zhang, Ketao Wang, Meng Xu, Jiayang Li, Yongch... <1%

48 Internet
archiv.ub.uni-heidelberg.de <1%

49 Internet
www.coursehero.com <1%

50 Internet
www.mdpi.com <1%

51 Internet
www.researchsquare.com <1%

Abstract

Chili peppers (*Capsicum* spp., family Solanaceae) are among the world's most economically important vegetable crops, valued for their rich nutritional and medicinal properties. These diploid plants (2n = 24; ~3.5 Gb genome) are believed to have originated in tropical Central and South America around 7500 BCE. Of the more than 38 reported species, five are widely cultivated: *C. annuum*, *C. chinense*, *C. frutescens*, *C. baccatum*, and *C. pubescens*. India, the world's leading producer (1.7 million tons annually; Ministry of Agriculture & Farmers' Welfare, 2022; FAO, 2022; USDA FAS, 2023), harbors remarkable genetic diversity, particularly in the Northeast, which is home to unique landraces such as the fiery hot *Bhut jolokia* (*C. chinense*). The genus is a biochemical powerhouse, rich in vitamins (both water- and fat-soluble), antioxidants (carotenoids, flavonoids, polyphenols), and capsaicinoids, which collectively contribute to numerous health benefits including anti-cancer, anti-microbial, and anti-inflammatory effects, as well as industrial applications such as food coloring and pest control.

Although closely related to tomato- a well-established model for fruit crops the molecular mechanisms underlying important traits in *Capsicum*, such as fruit development, ripening, and the biosynthesis of essential vitamins (e.g., ascorbate and tocopherol), remain poorly understood. This knowledge gap has hindered targeted breeding efforts to improve desired fruit traits, quality and nutritional value in *Capsicum*.

This study employed an integrative genomics approach to address these gaps. Orthology-based mining identified key genes involved in fruit development, ripening, and vitamin biosynthesis (Vitamin C and E) pathways in *C. annuum*, *C. baccatum*, and *C. chinense genomes*. Their expression profiles were analyzed using RNA-Seq data and quantitative real-time PCR (qRT-PCR), and correlated with phenotypic data, including vitamin quantification by HPLC at three distinct fruit developmental stages in contrasting genotypes. Furthermore, a set of functional molecular markers (SSRs and InDels) were developed from the candidate genes associated with fruit development, ripening, and Vitamin C and E biosynthesis.

Our investigation yielded several important findings. We identified 32 orthologs associated with fruit development and ripening, of which 12 showed significant differential expressions, further validated by qRT-PCR, six out 12 genes e.g., *MADS-RIN, SGR1, ETR4, LeSPL, XTH5, MADS-protein1,)* shows almost similar expression in both qRT and transcriptome-*MADS-RIN gene* showed consistent, extreme upregulation (>500-fold) in ripening fruit across all species.*SGR1* displayed consistently high expression in fruit, with the highest levels in *C. frutescens* in both datasets. *ETR4, LeSPL, XTH5, MADS-protein1*: Both methods captured the same pattern of higher expression in early stages (flower, early fruit, breaker) and a decline upon maturity. Out of above some genes found with opposite/discordant expression, 2 genes *GLK2* and hydroquinone glucosyltransferase, and *NSGT1* significant differences in both qRT and transcriptome- *GLK2* qRT-PCR showed upregulation in *C. annuum*

but downregulation in others, while the transcriptome suggested downregulation in all three species. Hydroquinone glucosyltransferase showed opposing regulatory trends between the two methods. *NSGT1* expression patterns differed significantly between the platforms. *TAGL1, FUL1, FUL2*: While both methods agreed on high expression, they showed minor differences in the precise level and tissue-specificity of expression.

Significantly, the *MADS-RIN* ortholog *(LOC107847473)* exhibited a striking >500-fold upregulation in mature fruit, underscoring its central regulatory role. Expression analyses of ethylene receptors revealed species-specific ripening behaviors, with *C. frutescens* displaying climacteric-like characteristics. These results demonstrate strong concordance between transcriptome profiling and qRT-PCR validation across developmental stages.

Analysis of vitamin content across different fruit developmental stages in contrasting genotypes revealed the highest accumulation of Vitamin C in *Capsicum* species, in *C. chinense*, as in Cc-9 (38.41 mg/g) in *C. frutescens* as in Cf-7 (29.19 mg/g), and in *C. annuum in* Ca-8 (41.77 mg/g). while highest Vitamin E levels were recorded in *C. chinense* in Cc-10 (109.39 mg/g), in *C. frutescens* in Cf-9 (112.44 mg/g), and in *C. annuum in Ca-5 (*111.43 mg/g). On average, *C. chinense* exhibited the highest Vitamin C content across stages, (25.33 mg/g), followed by *C. annuum* (24.25 mg/g) and *C. frutescens* (22.09 mg/g). Genotypes such as from *C. chinense* (Cc-1,Cc-2, Cc-3, Cc-5, Cc-6 and Cc-9), from *C. annuum* (Ca-3, Ca-4,Ca-7 and Ca-8) and from *C. frutescens* (Cf-4, Cf-5, Cf-7 and Cf-8) for Vitamin C, and genotypes from *C. chinense* (Cc-2, Cc-3, Cc-9, and Cc-10), from *C. annuum* (Ca-1, Ca-2 ,Ca-f and Ca-6) and from *C.*

frutescens (Cf-2, Cf-4, Cf-7, Cf-9 and Cf-10) for Vitamin E, represent valuable genetic resources for breeding programs aimed at developing *Capsicum* varieties enriched in Vitamins C and E content.

A comprehensive genome-wide analysis identified 29 and 81, 44 and 85, and 36 and 70 putative genes involved in Vitamin C and Vitamin E biosynthesis/degradation in the genomes of *C. annuum*, *C. baccatum*, and *C. chinense*, respectively. Among these, several genes associated with Vitamins C and E were found to be differentially expressed in RNA-Seq data generated from three fruit developmental stages of contrasting vitamin-rich species (*C. annuum*, *C. chinense*, and *C. frutescens*), including key regulators for Vitamin C such as *GMP* Downregulated during fruit development in all species, *GME* shows same expression pattern in both qRT and transcriptome, *AKR2 & GPI*: Both downregulated in all three species, *GGP* expression higher in *C. annuum* (qRT) vs higher in early fruit (transcriptome), *MIOX*: high in *C. chinense* (qRT) vs low in all species (transcriptome), *AKR38* high in *C. annuum* (qRT) vs no expression in transcriptome, GPP upregulated in maturity in qRT and downregulated transcriptome, *GalDH* higher expression in qRT and very low expression in transcriptome and *PMI* gene expression higher in C. frutescens qRT and in transcriptome higher in *C. annuum*. Similarly for Vitamin E key regulators genes *HPPD & IPI* upregulated during fruit development in both qRT and transcriptome, *VTE3 & VTE4*: Nearly identical expression patterns in both methods, *TAT* shows species-specific expression in qRT and opposite pattern in transcriptome, *TYRA* expression high in *C. frutescens* qRT and low/no detection in transcriptome, *VTE1*

found in qRT and low/no expression in transcriptome, *VTE2* expressed in all samples in qRT vs minimal expression transcriptome and *VTE5* gene expressed across tissues in qRT vs no expression in transcriptome. qRT-PCR validation confirmed the transcriptome-based expression patterns in Vitamin C with Similar expression 4 genes were found (*GMP, GME, AKR2, GPI, AKR1*) in both qRT and transcriptome and Opposite/Discordant expression in 6 genes (*GGP, MIOX, AKR38, GPP, GalDH, PMI*) and in Vitamin E with similar expression 5 genes (*GGDR, HPPD, IPI, VTE3, VTE4*) and in Opposite/Discordant expression 5 genes (*TAT, TYRA, VTE1, VTE2, VTE5*) was found in both qRT and transcriptome. Correlation analysis between gene expression and fruit vitamin content revealed significant associations, suggesting that these genes represent strong candidates for functional validation through approaches such as overexpression or genome editing.

To facilitate future genetic and breeding studies, we developed 49 gene-based SSR markers associated with fruit development and ripening traits, along with 24 SSR and 37 InDel markers linked to vitamin traits. These markers, after validation, could be deployed in breeding programs targeting improved fruit quality and nutritional content.

This is the first comprehensive study to report genome-wide identification and expression profiling of genes involved in fruit development, ripening, and vitamin biosynthesis in *Capsicum*. The identified candidate genes, together with the novel SSR and InDel markers, represent valuable genomic resources that can accelerate molecular breeding of biofortified *Capsicum* varieties with enhanced vitamin content. These

v

findings provide a foundational framework for researchers working on fruit crop improvement, particularly in the molecular dissection and manipulation of key nutritional and fruit traits.

DEDICATED TO

MY FAMILY

ACKNOWLEDGEMENTS

I am profoundly grateful to the Divine, the source of all wisdom and strength, whose unwavering compassion and grace supported me throughout my Ph.D. journey. His boundless blessings and guidance transformed this path into an enriching experience and brought me to this stage. This work stands not merely as a scholarly pursuit but as a testament to His blessings, for which I remain eternally thankful.

16 *I would like to express my sincere gratitude to my supervisors, Prof. Jai Gopal Sharma and Prof. Nirala Ramchiary, for their guidance, understanding, and patience throughout my research journey. Your positive encouragement and warm support were essential in completing this thesis. I am especially grateful for your insightful feedback, which helped me sharpen my thinking and research skills. It has been both an honour and a pleasure to learn from you both. My deepest Regards to both of you Sir.*

37 *Besides my supervisors, I would like to thank my SRC and DRC committee members for their insightful comments, suggestions, encouragement, and hard questions, which greatly helped refine my research work.*

41 *I would like to express my deepest appreciation and gratitude to the Department of Biotechnology, Delhi Technological University, and the School of Life Sciences, JNU. I sincerely thank the Department of Biotechnology Deans (current and former) for providing me with the best facilities to complete my research work. I also gratefully acknowledge the generous financial support from DST WOS-A, Department of Science and Technology.*

I also convey my gratitude to all the respected teachers and administrative staff of the 25 *Department of Biotechnology, DTU, and the School of Life Sciences, JNU, for their* 45 *kind help and support throughout my research period. I would also like to thank all the* 25 *CIF members of SLS and AIRF at JNU for assisting with my experiments. I am thankful to the staff and workers of the School of Life Sciences, JNU, and Department of*

Biotechnology, DTU, for providing a clean and hygienic working environment.

I am deeply grateful to all the esteemed individuals who provided me with courage and taught me the value of patience through their long experiences during my Ph.D. I extend my sincere thanks to Prof. D.P. Singh, Prof. R.K. Behel, and Prof. A.N. Singh. My connection to these esteemed mentors was made possible through my dear uncle, Mr. Sunil Tiwari, who always stood by me in my ups and downs. To him, I extend my heartfelt thanks and affection.

I am deeply grateful to Dr. Vandana Jaiswal for her guidance and unwavering support from the beginning until the completion of my Ph.D. journey. I will always cherish her encouragement, invaluable suggestions, and constant help through my ups and downs, patiently solving my doubts and problems despite geographical distance. Her willingness to listen and provide thoughtful guidance during my most challenging times means more than words can express, and for that, I will remain forever thankful.

I would like to express my heartfelt gratitude to Prof. S. Gourinath, who introduced me to a wonderful mentor and provided a positive environment to understand and conduct research. He has played a huge role in my research journey. Thank you, Sir.

My heartfelt thanks to Dr. Rashmi Gaur for her help and support at the beginning of my research journey. Also, my sincere thanks to Dr. Ajay Kumar, who was always available to support me whenever I faced difficulties during my Ph.D.

I am also very grateful appreciation to Dr. Ilyas Ahmed and Dr. Abdul Rawoof, who worked hard and supported me throughout my Ph.D. journey, assisting and at times guiding me during the completion of my research. Thank you very much for always being available to help solve my problems.

I would like to extend my heartfelt thanks and deep appreciation to dear, Dr. Khushbu Islam, for her invaluable suggestions and constant support throughout my research journey. I will forever cherish the moments she stood by me during difficult times, as

well as the joyful times we spent together, from the lab to the field and across the JNU campus. Khushbu, you will always have a special place in my heart.

My appreciation and special thanks go to Mr. Souparna and Ms. Rinku for their timely support during my Ph.D. submission. I also extend warm thanks to my other lab members in the Genomics and Systems Biology Research Lab, JNU—Dr. John Momo, Vishes, Ms. Garima, Ms. Mochoyia, Ms. Urmi, Ms. Anamika, Ms. Priyanka, Ms. Mona, Dr. Swati, Dr. Gitanjali, Mr. Nitin Kumar, Ms. Mandira, and Shyam and my seniors Dr. Amita, Dr. Sushil, Dr. Sarpras, Mr. Vineet, and Dr. Mukesh Nitin for their help throughout this journey. I also thank my lab staff, Mrs. Saraswati Devi and Mr. Gajanand Tiwari, for their support in both lab and field work.

I would like to extend my sincere gratitude to all my beloved friends who supported me through thick and thin. Most importantly, I am deeply grateful to my dearest and loving friend, Dr. Neha Tiwari, for her unwavering support and invaluable suggestions. You were always with me through the toughest moments; you saw me laugh in happy times and cry in misery. Thank you feels too small for everything you have done for me.

Lastly, my family deserves endless gratitude for being the greatest source of inspiration. I owe a great deal to my mother, Mrs. Veena Dubey, and my father, Mr. Prakash Dubey, who have given me unwavering support, love, and guidance in life. They shaped my vision and taught me the values that truly matter in life. I also extend my warm thanks and gratitude to my parents-in-law, Mrs. Sushma Pandey and Mr. R.S. Pandey, for their blessings and support throughout this journey. I would also like to thank Mr. Aditya and Mrs. Pallavi. I am deeply thankful to my loving sisters, Poornima and Nidhi, who stood by me every step of my journey and helped me at every stage of life. I feel very fortunate to have you both in my life. My appreciation and thanks also go to Mr. Sandeep and Mr. Mayank for their invaluable support and care for my parents in my absence.

I extend my heartfelt appreciation to every individual who has played a role, big or

small, in assisting me along this path. Their contributions have been indispensable and deeply appreciated.

Last but not least, I express my deepest gratitude and love to most important facilitator of this journey, my better half, Mr. Akash Pandey. Without his unwavering support, this journey would not have been possible. My heart will always remain indebted to his selfless love, infinite blessings, eternal sacrifice, unquestionable support, and constant guidance.

The successful completion of my doctoral program is truly the fruit of the prayers and blessings of many people.

Meenakshi Dubey

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

Abbreviations

ABA	Abscisic Acid
ACO	*ACC Oxidase*
ACS	*ACC Synthase*
AOX	Alternative Oxidase
BHT	Butylated Hydroxytoluene
BLAST	Basic Local Alignment Search Tool
BF	Breaker Fruit
bp	Base Pair
Ca	*Capsicum annuum*
Cb	*Capsicum baccatum*
Cc	*Capsicum chinense*
CDS	Coding Sequence
Cf	*Capsicum frutescens*
CNR	Colorless Non-Ripening
CTAB	Cetyltrimethylammonium Bromide
CTR1	Constitutive Triple Response 1
CSS	Conserved Syntenic Segment
DPA	Days Post Anthesis
EF	Early Fruit

EIL	*Ethylene Insensitive-Like*
EIN2	*Ethylene Insensitive 2*
ERF	*Ethylene Response Factor*
EST	Expressed Sequence Tag
ETR	Ethylene Receptor
ETR4	*Ethylene Receptor 2-Like*
FDR	False Discovery Rate
FP	Forward Primer
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
FUL1/FUL2	*Fruitful 1/Fruitful 2 (MADS-box genes)*
GLK	*GOLDEN2-LIKE Transcription Factor*
GLK2	*Golden 2-Like Protein*
HPLC	High-Performance Liquid Chromatography
InDel	Insertion-Deletion
MADS-RIN	*MADS-box Transcription Factor (Ripening Inhibitor)*
Mbp	Million Base Pairs
MF	Mature Fruit
MPA	Metaphosphoric Acid
NCBI	National Center for Biotechnology Information
NBPGR	National Bureau of Plant Genetic Resources
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction

PIC	Polymorphism Information Content
qRT-PCR	Quantitative Real-Time Polymerase Chain Reaction
ROS	Reactive Oxygen Species
RNS	Reactive Nitrogen Species
RP	Reverse Primer
SEP	SEPALLATA (MADS-box transcription factor)
SGR1	Senescence-Inducible Chloroplast Stay-Green Protein 1
SHU	Scoville Heat Units
Sl	Solanum lycopersicum (Tomato)
SSR	Simple Sequence Repeat
TAGL1	*Tomato AGAMOUS-LIKE1*
TFs	Transcription Factors
TSS	Transcription Start Site
VitC	Vitamin C (Ascorbic Acid)
VitE	Vitamin E (Tocopherol)
XTH5	*Xyloglucan Endotransglucosylase-Hydrolase 5*
DHAR	*Dehydroascorbate Reductase*
GME	*GDP-Mannose-3,5-Epimerase*
GGP	*L-Galactono-1,4-Lactone Dehydrogenase*
MDHAR	*Monodehydroascorbate Reductase*
VTE1-VTE5	*Vitamin E Biosynthesis Genes (Tocopherol Cyclase, Homogentisate Phytyltransferase, etc.)*

ANOVA	Analysis of Variance
HSD	Honestly Significant Difference (Tukey's Test)
r	Pearson Correlation Coefficient
TM	Melting Temperature

Symbols

°C	Degree Celsius
µg	Microgram
µg/g	Micrograms per Gram
µl	Microliter
µM	Micromolar
bp	Base Pair
Kb	Kilobase
mg/g	Milligram per Gram
mL	Milliliter
mM	Millimolar
nm	Nanometer
rpm	Revolutions Per Minute
v/v	Volume per Volume
±	Standard Deviation

XXX

CHAPTER 1

INTRODUCTION & REVIEW OF LITERATURE

CHAPTER 1 INTRODUCTION & REVIEW OF LITERATURE

1.1 Introduction, origin, and domestication of *Capsicum*

The genus *Capsicum*, commonly known as chili peppers, is a member of the Solanaceae (nightshade) family that includes other economically important crops like tomato, potato, and eggplant (Carrizo García et al., 2016). It is a globally significant genus comprising approximately 38 species (Barboza et al., 2022) with five primary species being domesticated for cultivation: *C. annuum* (e.g., bell peppers, jalapeños), *C. chinense* (e.g., Habanero, Bhut Jolokia), *C. frutescens* (e.g., Tabasco, Bird's Eye chili), *C. baccatum* (e.g., Aji peppers), and *C. pubescens* (e.g., Rocoto). These species are cultivated for a wide array of purposes, including culinary applications as spices and vegetables, medicinal uses due to their bioactive compounds, and for the development of self-defense sprays (e.g., pepper spray, (Ghosh et al., 2016; Zimmer et al., 2012). The characteristic pungency (spiciness) of peppers is attributed to a group of alkaloids known as capsaicinoids, of which capsaicin and dihydrocapsaicin are major components.

Capsicum is native to the Americas. Its center of origin is believed to be in the Bolivia-Brazil-Peru region, with a secondary center of diversity in Mexico and Central America (Pickersgill, 2007). Archaeological evidence, such as starch grain fossils, traces its presence back to as early as 7500–6000 BC in tropical regions of the Americas (Kraft et al., 2014; Ramchiary et al., 2014).

Capsicum is considered one of the earliest domesticated plants in the Americas, with independent domestication events occurring for different species:

- *C. annuum* was likely domesticated in east-central Mexico around 6,000 years ago (Kraft et al., 2014).
- *C. chinense* and *C. frutescens* were domesticated in northern South America (e.g., Peru, Ecuador) and lower Central America (Perry et al., 2007).
- *C. baccatum* was domesticated in the southern Andes (Peru, Bolivia, Chile, (Pickersgill, 2007).
- *C. pubescens* was domesticated in the highlands of the Andes (Perry et al., 2007).

The primary trait selected for domestication was likely pungency (capsaicinoid production), which acted as a natural defense against mammalian frugivores and fungal infestation, while birds (which are immune to capsaicin and are key seed dispersers) were attracted to the brightly colored fruits (Tewksbury et al., 2008).

Following the Columbian Exchange after 1492, *Capsicum* was rapidly introduced to the rest of the world by Spanish and Portuguese traders (Andrews, 1995). It was quickly adopted into cuisines and traditional medicine systems across Europe, Africa, and Asia due to its ability to grow in diverse climates and its value as a substitute for black pepper (Piper nigrum), which was expensive and controlled by trade monopolies (Bosland & Votava, 2012). India became a secondary center of diversification, developing a vast array of local landraces of chili (Bhagowati & Changkija, 2009).

1.2 Economic Importance and Production of *Capsicum*

Capsicum is a crop of significant economic and agricultural value due to its wide range of uses, nutritional benefits, and agronomic diversity. It possesses numerous commercially important traits such as high carotenoid content, capsaicin and capsanthin accumulation, fruit number, size, shape, and resistance to several diseases, especially studied in *C. annuum* genotypes. Among the different species, *C. chinense* (e.g. Bhut Jolokia or Ghost pepper) and *C. frutescens*, are known for their high and medium pungency. Bhut Jolokia, one of the world's hottest chilies, contains high levels of capsaicinoids, making it an ideal model for studying capsaicinoids biosynthesis. The heat, or pungency, attributable to capsaicinoids, offers unique sensory experiences, driving their use in both food and condiment industries (Maga & Todd, 1975). These qualities not only make products more appealing but also offer valuable genetic material that can be transferred to elite germplasm to improve yield, pest resistance, and overall crop performance.

Capsicum fruits are also valued globally for their culinary versatility, vibrant colors, and nutritional significance. They are consumed fresh, dried, or processed into powders and pastes, enhancing the flavor and appeal of various dishes. Additionally, *Capsicum* is prized as an ornamental plant, thanks to its colorful foliage, diverse fruit shapes, and aesthetic flowering patterns. Its demand is particularly high in tropical and subtropical regions such as Africa, Latin America, and parts of Asia. The wide variety of cultivars from mild culinary types like *C. annuum* to intensely pungent types like *C. chinense* and *C. frutescens* ensures its relevance across cuisines, cultures, and commercial markets.

Almost approximately 36 million tons of chilies/peppers are produced annually, with India

emerging as the largest producer, consumer, and exporter of chili peppers. In the 2023–2024, India produced 1.98 million metric tons of chili across 8.09 lakh hectares, with an average yield of 3,273 kg per hectare (FAOSTAT 2024). Andhra Pradesh led the production, contributing 44% of the national output, followed by Telangana, Madhya Pradesh, Karnataka, and Odisha. The Guntur district in Andhra Pradesh is renowned as Asia's leading hub for red chili production and exports, supplying major international markets, including Sri Lanka, Bangladesh, the Middle East, South Korea, the U.K., the USA, and Latin America. Indian chilies are globally recognized for their rich color and varying pungency levels, qualities that sustain their high demand in the global spice trade (Spice board, Govt of India, 2023-2024) India's dominance in chili production from 2020 to 2024, as noted by the Agricultural Market Intelligence Centre (PJTSAU), further highlights *Capsicum*'s critical economic and agricultural role.

1.3 Nutritional and Medicinal Aspects of *Capsicum*

Capsicum fruits are nutrient-rich and widely valued for their health-promoting properties. They contain high levels of antioxidants, capsaicinoids, flavonoids, and essential vitamins such as C and E, which help combat oxidative stress and reduce the risk of degenerative diseases (Chiaiese et al., 2019.; Ioannidi et al., 2009.; Isabelle et al., 2010.; Li et al., 2019; Perla et al., 2016; Tang et al., 2009.; Wahyuni et al., 2011). Variations in fruit color such as green, yellow, orange, chocolate, and purple black (Martínez et al., 2005; Matthäus et al., 2009), result from pigments like capsanthin, β-carotene, lutein, and zeaxanthin, which further contribute to nutritional value. Several studies have shown that Vitamin C and E content varies with fruit maturity, genotypes, and drying methods (Karatas et al., 2021; Korkutata & Kavaz, 2015; Nerdy, 2018; Osuna-García

et al., 1998; Perla et al., 2016).

Medicinally, *Capsicum* has been used in traditional systems for treating various ailments such as asthma, gastrointestinal disorders, fever, toothache, arthritis, and headaches (Bosland, 1996; Ghosh et al., 2016; Meghvansi et al., 2010). Bioactive compounds like capsaicin exhibit analgesic, anti-inflammatory, antimicrobial, and anticancer properties (C.-H. Lin et al., 2013; Meral et al., 2014; Mori et al., 2006; Vendrely et al., 2017). Capsaicin also aids in weight management, improving metabolic health and thermoregulation (Rastogi et al., 2024). Cultivars like *C. chinense* (Bhut Jolokia) and *C. frutescens* are extensively used in traditional medicine for conditions ranging from respiratory issues to joint pain (Bhagowati & Changkija, 2009; Kamatenesi-Mugisha & Oryem-Origa, 2005; Meghvansi et al., 2010).

Capsicum is recognized in pharmacology for its wide-ranging health benefits, from treating colds and skin conditions to combating chronic disorders like cancer, hypertension, and arthritis (Govindarajan & Sathyanarayana, 1991; Palevitch & Craker, 1995; Wahyuni et al., 2013). Its bioactive compounds serve as free radical scavengers, boost immunity, and exhibit anti-obesity and neuroprotective effects (Ertani et al., 2014; Joo & Lee, 2010; Sarpras et al., 2019; Srinivasan, 2016).. Indigenous knowledge from Northeast India has highlighted its effectiveness in treating conditions such as night blindness, gastritis, and spondylitis (Dutta et al., 2017). Importantly, the discovery of capsaicin’s role in temperature perception was foundational to David Julius's Nobel Prize in 2021, emphasizing its scientific and medical significance (Latorre & Díaz-Franulic, 2022; Tobita et al., 2021).

1.4 Fruit Development and Ripening

The journey from a flower to a ripe, edible fruit is a complex and highly regulated process involving distinct phases of development followed by ripening. While often discussed together, development and ripening are separate biological programs. Fruit development encompasses all stages from pollination to the point where the fruit reaches its full size and physiological maturity. This phase is primarily focused on growth and morphogenesis, key stages of development:

(a) Fruit Set (Ovary Development): This is the initiation phase, triggered by the successful pollination and fertilization of the ovule. If fertilization fails, the flower abscission. Hormones like auxins and gibberellins produced by the developing seeds are critical signals that prevent abscission and promote ovary wall (pericarp) growth to form the fruit (Joo & Lee, 2010; Prasanna et al., 2007).

(b) Cell Division Period: Immediately after fruit set, the pericarp undergoes a phase of rapid mitotic activity. The number of cells increases exponentially, determining the final potential for fruit size and setting the future tissue structure. This stage is relatively short but crucial (Zhang et al., 2006) (Cong et al., 2020).

(c) Cell Expansion Period: This is the major phase of fruit growth, where cells enlarge dramatically without dividing. Growth is driven by water uptake, vacuole expansion, and the accumulation of biomass such as carbohydrates (mainly starch), organic acids, and minerals. Hormones like gibberellins and auxins play a key role in coordinating this expansion (Chen et

al., 2020; McAtee et al., 2013).

(d) Maturation (Full-sized fruit): The fruit reaches its maximum size and dry weight. It is now physiologically mature. At this stage, the fruit is often hard, green (chlorophyll-rich), acidic, and starchy. The seeds inside have completed their development and are viable. For non-climacteric fruits, this stage may be followed directly by ripening, while climacteric fruits remain in a quiescent state until triggered by ethylene.

The Regulation of development regulated by key hormones like early phases are dominated by auxins, gibberellins, and cytokinins, which promote cell division, cell expansion, and nutrient allocation, making the fruit a strong "sink" for photosynthates (Prasanna et al., 2007) (R. Kumar et al., 2013). **Fig 1.1**

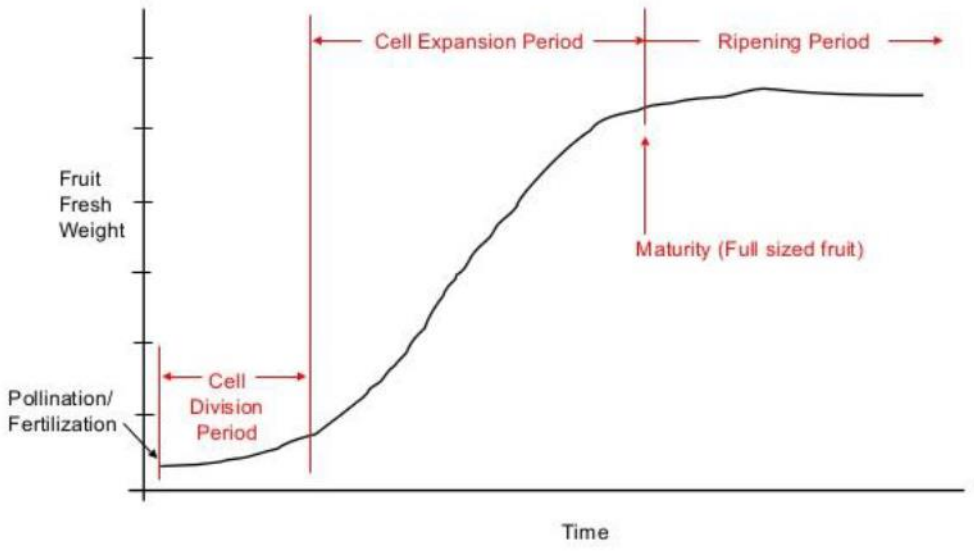


Fig. 1.1 Pattern of fruit development and ripening (Adapted from Sharma et al. 2020)

Ripening is the terminal phase of fruit development. It is not a degenerative process but a highly

coordinated genetic program that transforms a mature, often inedible fruit into a soft, sweet, flavorful, and aromatic organ to attract seed dispersers. Key Biochemical and Physiological Changes:

(a) Color Change (Pigmentation): Chlorophyll is degraded, unmasking existing carotenoids and allowing for the de novo synthesis of new pigments: lycopene and β-carotene (red, orange) and anthocyanins (purple, blue) (Barsan et al., 2010).

(b) Texture Softening: The rigid cell walls are broken down by a suite of enzymes including Polygalacturonase (PG), Pectin Methylesterase (PME), and Cellulase, leading to the dissolution of the middle lamella and cell wall loosening (Brummell, 2006; S. Wang et al., 2018).

(c) Flavor and Aroma Development:

- Sweetness: Starch is hydrolyzed into simple sugars like glucose, fructose, and sucrose.
- Acidity: Organic acids (e.g., malic, citric acid) are metabolized and used as respiratory substrates, reducing sourness.

Aroma: A complex blend of volatile compounds (esters, aldehydes, alcohols, terpenoids) is synthesized, giving the fruit its characteristic scent (Klee & Tieman, 2018).

(d) Nutritional Enhancement: Levels of vitamins (e.g., ascorbate/Vitamin C, carotenoids/Vitamin A precursors) and antioxidant compounds (e.g., flavonoids, phenolics) often increase significantly during ripening

The ripening program is governed by two principal hormonal pathways:

(i) Climacteric Fruits-Ripening is initiated and driven by a massive burst of ethylene production

and a concurrent spike in respiration; ethylene is the primary trigger for ripening; this process is autocatalytic (ethylene stimulates its own production). In this process governed by a cascade-*ACS/ACO* (biosynthesis) → *ETRs* (receptors) → *CTR1* → *EIN2/EIN3* (signaling) → ERF TFs (response) (X. Li et al., 2019; S. F. Yang & Hoffman, 1984) and regulates by MADS-box TFs (e.g. *RIN, TAGL1, FUL1/2* in tomato) are critical (Dong et al., 2013; Vrebalov et al., 2002). e.g. Tomato, Banana, Apple, Mango and *Capsicum* (Pepper).

(ii) Non-Climacteric Fruits-Ripening is not dependent on ethylene and occurs without a respiratory climacteric. Abscisic Acid (ABA) is a key regulator. Other hormones like auxin may also play roles in ripening, ethylene signaling components are present but do not initiate the full ripening program. Ripening is more gradual by involving other TFs, but some MADS-box genes (e.g., *FaSHP* in strawberry) are also important.

The ripening physiology of pepper exhibits significant complexity, transcending the categorization between climacteric and non-climacteric classifications. While often classified as non-climacteric, many hot pepper varieties (*Capsicum* spp.) show a partial or delayed climacteric rise in ethylene and respiration, suggesting a species-specific regulatory mechanism that blends both pathways (Aizat et al., 2013; Hou et al., 2018). The ortholog of the tomato *RIN* gene (*CaMADS-RIN*) has been identified in pepper and is a key positive regulator of ripening (Dong et al., 2014) **Fig 3.2.**

1.4.1 Diversity analysis

The Solanaceae family comprises a selection of the world's top economically important crops, including tomato (*Solanum lycopersicum*), pepper (*C. annuum*), and potato (*Solanum tuberosum*). These species display extraordinary diversity in fruit development and ripening traits, spanning morphological, biochemical, and genetic dimensions (Knapp et al., 2023). Such variation arises from distinct physiological mechanisms, such as pigment biosynthesis (e.g., anthocyanins, carotenoids, and chlorophylls), structural differentiation in fruit size and texture, and the molecular networks that regulate these processes (Liu et al., 2020). Ripening strategies diverge widely across the family from ethylene-dependent climacteric ripening in tomato to abscisic acid-mediated non-climacteric programs in pepper (Dubey et al., 2019; Klee & Giovannoni, 2011; Osorio et al., 2013). Characterized by striking morphological plasticity, the Solanaceae (nightshade family) defies simple taxonomic categorization, as few morphological traits are consistently conserved across all its members (Hunziker, 2001).

Evolutionary and functional studies have uncovered both conserved and lineage-specific pathways governing fruit maturation. Key transcription factors (e.g., *RIN, CNR*, and *NOR* in tomato) and epigenetics regulate these developmental transitions (A. Kumar et al., 2021; Seymour et al., 2013). Further complexity stems from environmental adaptations and domestication-driven selection, which have sculpted the phenotypic and metabolic landscapes of modern cultivars (Escobar Rodríguez et al., 2021). Deciphering this diversity is not only fundamental to plant biology but also critical for enhancing crop resilience, nutritional value, and postharvest performance.

The genotypic and phenotypic variation underlying Solanaceae fruit development and ripening by integrating comparative genomics, transcriptomics, and metabolomics, offering insights into the improvement of these vital crops. Fruit color transformation serves as a primary visual marker of ripening. *Capsicum* species demonstrate extensive variation in ripening pigmentation, ranging from yellow and orange to red and brown phenotypes. (Borovsky & Paran, 2008; Hou et al., 2018). *Capsicum* represents an economically and biologically important genus, exhibiting substantial variation in fruit traits, including fruit development (shapes, sizes) and ripening patterns (A. Kumar et al., 2021).

A Previous study by (Sarpras et al., 2016). evaluation of *Capsicum* germplasm and reported a wide intra- and inter-specific diversity in fruit morphology, fruiting habit, and metabolite content across 136 *Capsicum* genotypes from Northeast India belonging to, C. *annuum* (56 accessions), *C. chinense* (63 accessions of Bhut jolokia), and *C. frutescens* (17 accessions). The germplasm exhibited a significant spectrum of fruit traits: *C. annuum* displayed the greatest morphological diversity, with fruit shapes ranging from long, elongate, ovate, and round to pumpkin-shaped in various sizes, alongside both upright and pendant fruiting habits and single to bunch-type clustering. In contrast, *C. frutescens* were the most uniform, characterized by very small, elongated fruits, exclusively upright habit, and single clustering. *C. chinense* accessions exhibited pendant habits with either single or bunch-type fruiting and predominantly ovate to elongated fruits. Metabolically, *C. chinense* accessions possessed the highest pungency (SHU) and antioxidant activity (DPPH assay), followed by *C. frutescens*, with *C. annuum* showing the lowest values for these traits. This characterization of *Capsicum* spp. shows Northeast India as a significant reservoir of genetic diversity.

1.4.2 Molecular aspects/gene identification

The development/ripening of fleshy fruits is a complex phenomenon and affected by plant hormone signaling. Based on ethylene signaling and respiration, fruit ripening is divided into two categories- climacteric (ethylene dependent) and non-climacteric (ethylene independent) (J. J. Giovannoni, 2004; Hou et al., 2018). There are three important components of ethylene biosynthesis viz. *1-aminocyclopropane-1-carboxylic acid* (*ACC*), *ACC synthases* (*ACSs*) and *ACC oxidases* (*ACOs*) (S. F. Yang & Hoffman, 1984). Ethylene is perceived by ethylene receptors (*ETRs*) and then interacts with *Constitutive triple response 1* (*CTR1*) kinase gene to activate *ethylene insensitive 2* (*EIN2*) and starts transcriptional cascade including *EIN3*/ *EIN3-like* (*EIL*) and ethylene responsive factors. Climacteric fruit ripening, involving ethylene, is well characterized in tomato using several ripening mutants like never ripe *(nr), colorless non-ripening (cnr), ripening inhibitors (rin),* and *non-ripening (nor)*. However, molecular mechanism of non-climacteric fruit ripening is not well understood. It has been suggested that abscisic acid (ABA) may play important role in non-climacteric fruit ripening. Some key genes of ABA biosynthesis such as *FaNCED1, FaBG3, FaPYR1/FaABAR* etc. have been shown to be differentially expressed during fruit ripening in strawberries (model for non-climacteric fruit, (Han et al., 2015; Jia et al., 2013; Kadomura-Ishikawa et al., 2015; Medina-Puche et al., 2016). A new group of genes called *SEP* genes (*SlMADS1*) was reported to play an important role in the ripening of both climacteric and non-climacteric fruits (Dong et al., 2013, 2014). Furthermore, *SlMADS-RIN* (*RIPENING INHIBITOR*, typical *SEP* gene) is a well-characterized fruit ripening gene reported in tomato which is responsible for softening, carotenoid accumulation, ethylene production and perception in fruit. *SlMADS-RIN* also interacts with other

fruit ripening genes like *Tomato AGAMOUS-LIKE* (*TAGL1*), *FRUITFULL 1 (FUL1)*, and *FRUITFULL 2* (*FUL2*). Ortholog of *SlMADS-RIN* in *Capsicum* has been cloned, and its potential role in fruit ripening in *Capsicum* has been reported (Dong et al., 2014)

1.4.3 Sequencing/transcriptome analysis

The emergence of Next-Generation Sequencing (NGS) technologies has marked a revolutionary shift in biological research, providing an unparalleled ability to explore the genome, epigenome, and transcriptome of organisms on an extensive scale. (Z. Wang et al., 2009). Among these, transcriptome sequencing, or RNA-sequencing (RNA-seq), has emerged as a particularly powerful tool. It enables high-throughput, quantitative profiling of the entire transcriptome, capturing the dynamic expression of coding and non-coding RNAs with a high degree of sensitivity and accuracy. This technology effectively superseded older microarray-based methods by offering key advantages, including the discovery of novel transcripts without prior genome annotation, the detection of alternative splicing events, and the ability to identify sequence variations like single nucleotide polymorphisms (SNPs) within expressed regions (Conesa et al., 2016; T. Wang et al., 2009).

This vast genomic resource drives discoveries across numerous species, facilitating the identification of novel genes, the development of functional molecular markers (e.g., SNPs, SSRs), and the comprehensive expression profiling of global transcriptomes under various conditions (Chhapekar et al., 2020; Jaiswal et al., 2022; Strickler et al., 2012). Crucially, the

power of RNA-seq extends to non-model organisms, enabling genomic studies in species without a reference genome through de novo transcriptome assembly (Grabherr et al., 2011).

RNA-seq has been extensively deployed to unravel the complex genetic networks governing critical agronomic traits, including growth, development, stress response, and fruit quality. The process of fruit development and ripening is a classic and highly coordinated genetic program involving significant changes in morphology, biochemistry, and physiology. It encompasses phases of fruit set, growth, maturation, and ripening, each characterized by distinct transcriptional cascades regulating processes such as cell division, metabolite accumulation, pigment biosynthesis, cell wall modification, and aroma production (Klee & Giovannoni, 2011; Seymour et al., 2013).

By providing a comprehensive snapshot of gene expression dynamics at specific stages of fruit development and ripening, RNA seq technology enables a detailed analysis of expression patterns, offering valuable insights into the dynamic regulatory networks that coordinate ripening. This includes stage-specific transcriptome analysis of cellular growth and expansion, pigment biosynthesis and color development, cell wall modification and texture softening, volatile compound synthesis and aroma formation, carbohydrate metabolism, and sugar accumulation. This approach is fundamental to understanding how transcriptional regulation drives complex physiological transformations during fruit development and ripening.

Our understanding of these regulatory networks has been significantly advanced by key studies. The tomato *Ripening Inhibitor (RIN),* a *MADS-box* transcription factor (TF), was identified as a master regulator of ripening, with its knockout mutants showing a complete inhibition of the

process (Dubey et al., 2019; Osorio et al., 2013; Vrebalov et al., 2002). RNA-seq analyses revealed that *RIN* directly modulates ethylene biosynthesis genes, including *1-aminocyclopropane-1-carboxylic acid synthase (ACS)* and *oxidase (ACO)*, which are crucial for climacteric ripening (Martel et al., 2011). This regulatory role appears conserved; in apple (*Malus domestica*), silencing *MdMADS8* sharply reduces the expression of ethylene pathway genes *MdACS1* and *MdACO1* (Ireland et al., 2014), while *MdMA DS2* influences fruit firmness (Cevik et al., 2010). Beyond *MADS-box genes, NAC* transcription factors have emerged as upstream regulators, potentially acting earlier in the regulatory cascade (J. Giovannoni et al., 2017). alongside other TFs like *AP2/ERF, WRKY*, and *bZIP* (Xu et al., 2012).

Recent technological advances are further refining our spatial and mechanistic understanding. Single-cell RNA-seq and spatial transcriptomics in tomato and pepper have begun to uncover distinct regulatory networks in different fruit tissues (pericarp, placenta, and seeds), highlighting how tissue-specific gene expression influences ripening dynamics and flavor compound accumulation (Kumar et al., 2021). Furthermore, integrative transcriptome-methylome analyses suggest that DNA methylation provides an additional layer of regulation by modulating ethylene response genes (Gallusci et al., 2023; Jaiswal et al., 2022; Rawoof et al., 2020). The future of the field lies in leveraging long-read sequencing (PacBio, Nanopore) and multi-omics integration to build a complete, multi-layered model of fruit ripening regulation.

Despite the availability of genomic resources, the comparative transcriptomic landscape across different developmental stages of fruit in closely related *Capsicum species* with distinct fruit phenotypes remains underexplored. A detailed comparative analysis can reveal conserved and

species-specific regulatory mechanisms, providing deeper insights into the evolution of fruit traits and identifying master regulators of development and secondary metabolism, such as capsaicinoid biosynthesis (Fuentes et al., 2012; Kim SeungIll et al., 2014), We utilized available RNA-seq data generated in our lab (Chhapekar et al., 2020), which includes three developmental stages of fruits, Early Fruit (EF), Breaker Fruit (BF), and Mature Fruit (MF) as well as Fully Opened Flower tissue from three distinct *Capsicum* species: *C. annuum* (PRJNA505972), *C. chinense* (PRJNA327797), and *C. frutescens* (PRJNA327800). This approach allows for a high-resolution investigation into the transcriptional dynamics underpinning fruit development/ripening, mirroring strategies that have successfully identified key regulatory networks in other species (Giovannoni et al., 2017; Xu et al., 2012).

1.5 Vitamins

Vitamins are essential micronutrients that humans must obtain from plant-based foods; the human body cannot synthesize vitamins in sufficient quantities (Macknight et al., 2017). These vital compounds are classified into two main groups: water-soluble (B-complex vitamins and Vitamin C) and fat-soluble (Vitamins A, D, E, and K). Plants produce these essential nutrients through complex biochemical pathways that span multiple cellular compartments such as plastids, mitochondria, and the cytosol, using carbohydrate-derived precursors as building blocks (Macknight et al., 2017). The consequences of vitamin deficiencies can be severe, driving efforts to enhance vitamin content in crops through biofortification strategies (Naidu, 2003; Walingo, 2005).

Among these essential nutrients, Vitamin C (ascorbic acid, AsA) stands out for its crucial roles in human health. Synthesized in plants primarily through the L-galactose pathway, which begins with D-glucose phosphorylation and culminates in AsA production, this potent antioxidant helps neutralize harmful free radicals (Levine, 1986). Beyond its antioxidant capacity, Vitamin C strengthens immune function, accelerates wound healing, and may help prevent serious conditions ranging from cancer to COVID-19 (Carr et al., 2020.; Levine, 1986).

Vitamin E, a group of eight lipid-soluble compounds including tocopherols and tocotrienols. While tocopherols predominate in plant leaves and dicot seeds, tocotrienols are more abundant in monocot endosperms (Hussain et al., 2019). Despite their wide distribution in plants, Vitamin E deficiency remains surprisingly common, with only about 21% of the global population maintaining optimal serum α-tocopherol levels (Guevara et al., 2021; Hussain et al., 2019; Manosso et al., 2022). This vitamin's health benefits are extensive, offering protection against cardiovascular diseases, cancer, and neurodegenerative disorders like Alzheimer's disease, while also demonstrating anti-inflammatory and neuroprotective properties (Gugliandolo et al., 2017; G. Y. Lee et al., 2018; Lloret et al., 2019; Mutalip et al., 2018; Rizvi et al., 2014). These Vitamin C and E together form a critical defense system against oxidative damage caused by reactive oxygen and nitrogen species (Guevara et al., 2021).

Capsicum species (chili peppers), members of the economically and nutritionally significant Solanaceae family, serve as exceptional natural reservoirs of essential vitamins (Bhandari et al., 2013.; Osuna-García et al., 1998). These crops demonstrate variability in their vitamin content, which is influenced by multiple factors, including genotype, developmental stage, and tissue type (Lidiková et al., 2021). The vitamin composition of *Capsicum* fruits exhibits dynamic

accumulation patterns; Vitamin C concentrations typically increase during fruit maturation, while Vitamin E show distinct tissue-specific distribution, with γ-tocopherol predominating in seeds and α-tocopherol accumulating preferentially in the pericarp (Bhandari et al., 2013.; Osuna-García et al., 1998).

1.5.1 Vitamin biosynthesis

The biosynthetic pathways producing vitamins in plants constitute an evolutionary innovation in secondary metabolism. These complex biochemical networks yield specialized metabolites that, while not essential for plant viability, have acquired importance as micronutrients for human health (Macknight et al., 2017). This nutritional symbiosis between plants and humans, as plants derive vitamins support fundamental aspects of human diet and well-being. In contrast to many animals, through evolutionary gene loss, humans became dependent on dietary vitamins, retaining only partial endogenous synthesis (e.g., Vitamin D/K) while maintaining vitamin-requiring metabolic pathways, making dietary plant sources crucial for meeting our micronutrient needs (Naidu, 2003). Therefore, vitamin deficiencies have played a pivotal role in shaping human history and remain a major catalyst for health initiatives and biofortification strategies across the world (Walingo, 2005). Within plant cells, vitamin biosynthesis unfolds along intricate metabolic networks that traverse multiple organelles, converting simple carbohydrate building blocks into structurally and functionally complex molecules (Macknight et al., 2017).

Vitamin C (ascorbic acid) biosynthesis is a classic example, requiring at least eight enzymatic

steps **(Fig 1.2)**. The primary pathway, known as the L-galactose pathway, initiates with D-glucose and proceeds through a cascade of genes *VTC1 (GDP-mannose pyrophosphorylase), GME (GDP-mannose 3', 5'-epimerase), VTC2/VTC5 (GDP-L-galactose phosphorylase), GDH (L-galactose dehydrogenase)*, and *GLDH (L-galactono-1,4-lactone dehydrogenase)*, with the terminal reaction occurring in the mitochondria (Ioannidi et al., 2009; Macknight et al., 2017.; Tyapkina et al., 2019). Mutations or natural variation in these genes can lead to substantial differences in ascorbate content, as shown in diverse *Capsicum* species and model plants (Almeida et al., 2011; Wildman et al., 2016). Recent transcriptomic analyses further reveal dynamic regulation of these genes across development and in response to environmental cues, reflecting the complexity and plasticity of Vitamin C accumulation (Lima-Silva et al., 2012).

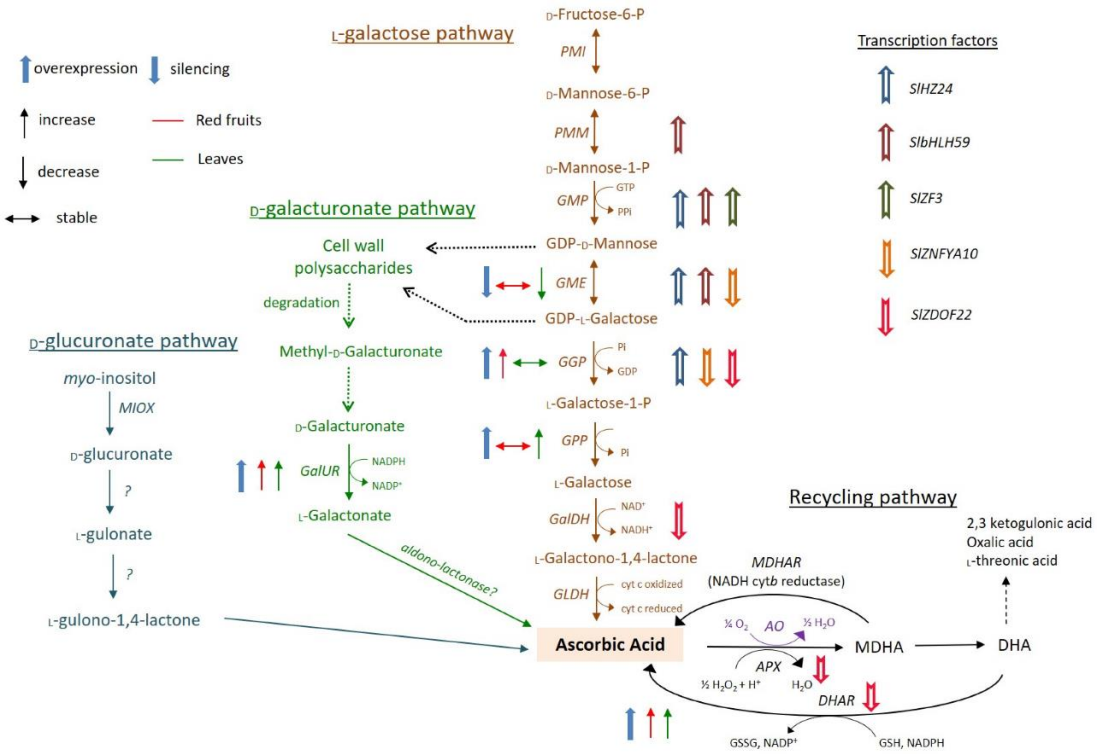


Fig. 1.2 The Vitamin C biosynthesis pathway illustrating important enzymatic steps with genes. The pathway has been taken from (Mellidou et al., 2021).

Vitamin E biosynthesis is similarly complex and integrates products from two distinct metabolic routes: the shikimate pathway (providing aromatic precursors) and the methylerythritol phosphate (MEP) pathway (yielding isoprenoid side chains). The condensation of *homogentisic acid (HGA) generated via tyrosine aminotransferase (TAT)* and *hydroxyphenylpyruvate dioxygenase (HPPD)* with *phytyl pyrophosphate (PPP)* or *geranylgeranyl pyrophosphate (GGPP)*, both derived from the MEP pathway, forms the backbone of tocopherol (Vitamin E) and tocotrienol synthesis (Meena et al., 2025; Mène-Saffrané & DellaPenna, 2010) Critical enzymatic steps are mediated by genes such as *VTE2 (homogentisate phytyltransferase), VTE3 (MPBQ methyltransferase), VTE1 (tocopherol cyclase)*, and *VTE4 (γ-tocopherol*

methyltransferase) (Almeida et al., 2011, 2015; Fritsche, Wang, Antioxidants, et al., 2017; Quadrana et al., 2012), each contributing to the biosynthetic flow and diversity of Vitamin E isoform (Niu et al., 2022; Liu et al., 2017). Mutations or overexpression of these genes can cause substantial shifts in tocopherol profiles and content, as demonstrated in *Arabidopsis*, soybean, and sweet corn studies (Li et al. 2025) (**Fig 1.3**).

The regulation of vitamin biosynthetic pathways is highly dynamic, regulated by developmental signals, tissue specificity, environmental factors, and stress responses (Lima-Silva et al., 2012; Osuna-García et al., 1998). In case of Vitamin C typically surges during fruit ripening in *Capsicum*, while Vitamin E isoforms display pronounced changes across tissues and developmental stages (Almeida et al., 2011; Osuna-García et al., 1998). Natural genetic variation imparts an additional layer of complexity. Single nucleotide polymorphisms and allelic diversity within biosynthetic genes can result in striking differences in vitamin content, providing both a challenge and an opportunity for enhancement of nutritional value (Wildman et al., 2016). Recent advances in omics technologies, genome editing, and transcriptome profiling have propelled forward our ability to uncover the regulatory networks and candidate genes underlying vitamin biosynthesis (Osuna-García et al., 1998) (F. Liu et al., 2017). These pathways and genes are well-characterized in model Solanaceous crops such as tomatoes; extrapolating this information to related species like *Capsicum* offers new opportunities for crop improvement through molecular breeding. A recent study showed, Vitamin C biosynthetic gene expression in specific cultivars of *Capsicum annuum L.*- 'Puparuriello' (PEP1) and 'Cazzone Rosso' (PEP10), but comprehensive molecular studies across the *Capsicum* genus are still lacking (Chiaiese et al., 2019) (Chiaiese et al., 2019) and Vitamin E.

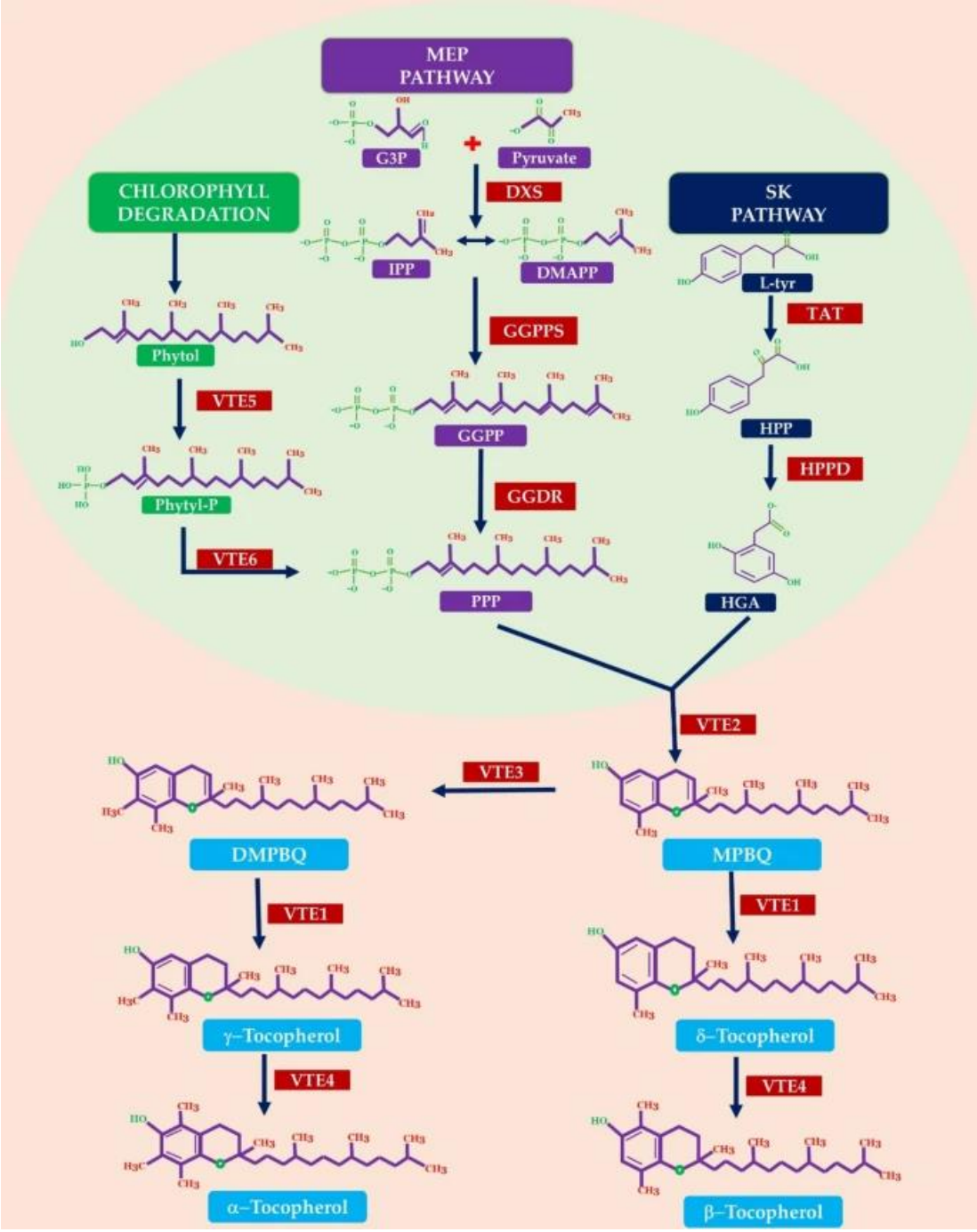


Fig. 1.3 The Vitamin E biosynthesis pathway illustrates important enzymatic steps with genes. The pathway has been taken from (Meena et al., 2025).

1.5.2 Diversity analysis

The Solanaceae family reveals diversity in vitamin composition, with *Capsicum* species representing particularly rich sources of essential micronutrients. Extensive phytochemical analyses demonstrate that peppers contain exceptional levels of Vitamin C (up to 400 mg/100g FW) and diverse carotenoids (provitamin A), along with significant quantities of Vitamin E (α-tocopherol) and B-complex vitamins (Chassy et al., 2006; Wahyuni et al., 2013). Such nutritional variability represents the combined inheritance of evolutionary biology and agricultural development working in across generations, with Vitamin C content varying 16-fold (11.9-195.8 mg/100 g) and Vitamin E levels differing 10-fold among accessions (Moon et al., 2023; Osuna-García et al., 1998). The genetic architecture underlying this variation includes polymorphisms in key biosynthetic genes - *PSY1, CCS* for carotenoids and VitC; *GalUR, GGP* for VitC and regulatory elements that have been shaped by domestication (Badejo et al., 2009; X. Li et al., 2024).

Comparative analyses reveal distinct vitamin accumulation patterns across Solanaceae crops. While tomato domestication favored lycopene over provitamin A carotenoids (Z. Liu et al., 2022), peppers maintained broader nutritional diversity through selection for both pigmentation and nutrient content.

Wild *Capsicum* species like *C. baccatum* accumulate exceptionally high Vitamin C (>400 mg/100 g FW) and unique tocopherol profiles, providing valuable genetic resources for breeding (Rodríguez et al., 2021) and few reports shows Vitamin C content was found higher and diverse in different genotypes of *Capsicum* (Conesa et al., 2016; Hamed et al., 2019; Kouassi et al.,

2012; Martínez-Ispizua et al., 2021; Mennella et al., 2017). According to Olatunji et al., 2019 few varieties of *Capsicum* show higher concentrations of Vitamin E. Vit E composition shows variation, with α-tocopherol predominating in pepper seeds while γ-tocopherol is more abundant in fruits, reflecting differential expression of *HPT (homogentisate phytyltransferase)* and *HPPD (4-hydroxyphenylpyruvate dioxygenase)* genes. Environmental factors further modulate vitamin content, as light exposure enhances Vitamin C but reduces folate levels, while nitrogen fertilization boosts B-Vitamin concentrations (Escobar Rodríguez et al., 2021; Goyer & Navarre, 2009).

The vitamin diversity in Solanaceae crops arises from several key genetic factors. Biosynthetic gene copy number variations, such as those affecting *PSY1* in carotenoid production, create fundamental differences in vitamin accumulation across species and cultivars (Liu et al., 2022). Similarly, loss of function mutations like *CCD4* in white potatoes explain the absence of certain vitamins in specific varieties (Campbell et al., 2010). Modern genomic tools have enabled precise identification of these genetic determinants, with GWAS studies revealing critical QTLs for instance, the chromosome 9 locus controlling ascorbate levels in peppers(Lee et al., 2020). Advanced genome editing techniques, particularly CRISPR-Cas9 targeting genes like *LCYE* in tomatoes, have demonstrated how single-gene modifications can significantly enhance β-carotene content (D'Ambrosio et al., 2018). Environmental factors interact strongly with genetic potential to shape vitamin profiles. Light exposure presents a particularly interesting dual effect of VitC production in tomato fruits; it simultaneously degrades folate in potato tubers (Goyer & Navarre, 2009) Agronomic practices modulate these relationships; shading reduces ascorbate in peppers, whereas nitrogen

fertilization boosts folate levels in potatoes. Breeding programs have successfully harnessed this knowledge, incorporating wild alleles like high β-carotene traits from *S. habrochaites* into cultivated varieties through targeted introgression (Sestili et al., 2019).

Analytical approaches are transforming understanding of vitamin diversity. Metabolite GWAS has proven a particularly powerful approach for linking vitamin traits to their genetic foundation, as demonstrated by the identification of ascorbate QTLs in pepper (M. Kim et al., 2020). Multi-omics integration provides even deeper insights, uncovering post-transcriptional regulatory mechanisms such as microRNA-mediated control of *GGP* expression in tomato (Mellidou et al., 2012). These approaches range from large scale association studies to molecular level investigations through analysis of the complex networks related to vitamin accumulation in Solanaceae crops and *Capsicum* species.

1.5.3 Identification of genes through genomics approaches

The molecular characterization of ascorbate and tocopherol biosynthesis in *Capsicum* species depends on two fundamental approaches: identification of biosynthetic genes and analyzing their transcriptional regulation via RNA sequencing/transcriptome (Chhapekar et al., 2020; Dubey et al., 2025). To investigate and improve the vitamin content in plants, particularly in economically important species like *Capsicum*, largely depends on molecular markers. These genetic keys allow researchers to identify, map, and select genes controlling complex metabolic traits. While Simple Sequence Repeats (SSRs) have been an important tool, the field is rapidly evolving towards more sophisticated, high-throughput technologies. SSRs have been extensively used in initial studies to map quantitative trait loci (QTLs) linked to vitamin biosynthesis pathways. For

instance, researchers have identified SSRs linked to key genes in the ascorbate (L-galactose) pathway, such as *GDP-L-galactose phosphorylase (GGP* or *VTC2*), a major rate-limiting enzyme (Bulley et al., 2009). *Galactono-1,4-lactone dehydrogenase (GalLDH)*, and *L-galactose-1-phosphate phosphatase* (*GPP* or *VTC4)*. Similarly, SSR markers have been developed for genes in the tocopherol pathway, such as *Homogentisate phytyltransferase (HPT)* and *p-Hydroxyphenylpyruvate dioxygenase (HPPD)* (Savadi et al., 2020).

However, the limited density of SSR maps often restricts their resolution, making it difficult to pinpoint the exact causal genes. This limitation is now being overcome by next-generation sequencing technologies like Genotyping-by-Sequencing (GBS), which can generate thousands of genome-wide markers (Elshire et al., 2011; Poland & Rife, 2012). The high-density genetic maps generated from GBS data enable genome-wide association studies (GWAS) and high-resolution QTL mapping, allowing for the precise identification of genomic regions controlling vitamin accumulation with far greater precision than was possible with SSRs. SSRs have limitations, Development Cost for Novel Species: Identifying and validating polymorphic SSR loci requires constructing genomic libraries and Sanger sequencing, which is time-consuming and expensive for non-model organisms.

GBS is a reduced-representation sequencing method that allows for the simultaneous discovery and genotyping of thousands of single nucleotide polymorphisms (SNPs) across the genome (Elshire et al., 2011; Poland & Rife, 2012). GBS can be used to sequence a diverse panel of pepper genotypes with varying Vitamin E content. By correlating SNP patterns with tocopherol concentration, we can identify novel genes or regulatory elements involved in the pathway that

were previously unknown.

Vitamin C (L-ascorbic acid, AsA) biosynthesis in plants involves a suite of enzymes encoded by diverse gene families, with several biosynthetic routes described the mainly D-mannose/L-galactose pathway. Recent genome and transcriptome wide analyses in *Capsicum annuum* (pepper) have systematically identified key genes encoding enzymes in this pathway. Specifically, 14 genes associated with the core D-mannose/L-galactose pathway, such as *PMM, GalLDH, GalDH, GMP, GME, GGP, GPP*, and *PMI*, were annotated in pepper, alongside additional genes for alternative AsA synthesis routes including *GalUR, GulLO*, and *MIOX* (ÉS de Aguiar et al., 2023). These genes are distributed across the pepper genome, suggesting a complexity in regulatory architecture and potential gene duplication events. Similarly, transcriptome profiling in tomato fruit reveals extensive characterization of the ascorbate pathway genes, with 46 structural genes putatively involved and high positive correlations identified between their transcript abundance and specific transcription factors (Ye et al., 2015). Particularly, transcription factors such as *MYB, NAC, AUX/IAA,* and *Dof* proteins show strong regulatory relationships with ascorbic acid biosynthetic enzymes, indicating multi-layered control over Vitamin C accumulation, especially during fruit development and ripening (Ye et al., 2015).

Additional transcriptomic studies in other crops, such as tea (Li et al., 2017), cucumber (Ren et al., 2024), and jujube (Lu et al., 2022), confirm the centrality of conserved enzymes (*GGP, GalLDH, PGI*, and others) in controlling AsA content and reveal stage dependent expression patterns, which parallel findings in pepper and tomato. Modern meta-analyses have also mapped

168 transcription factors contributing directly to Vitamin C metabolism, illustrating the broader regulatory landscape beyond core biosynthetic genes (Su et al., 2025).

Vitamin E biosynthesis in plants covers both the tocopherol and tocotrienol families, synthesized via tightly regulated pathways integrating the shikimate and methylerythritol phosphate (MEP) routes. While much of our foundational knowledge comes from model species like Arabidopsis and tomato, recent comprehensive genome studies in pepper have advanced for VitE related genes, identifying up to 85 non-redundant candidate genes in *Capsicum baccatum* (Ahmad et al., 2021). Genes like *VTE1, VTE2, VTE3*, and *VTE4* are central to tocopherol assembly, with their expression highly co-regulated during fruit maturation (Dubey et al., 2025).

Transcriptome investigations in tomato fruit development context reveal 41 genes connected to Vitamin E pathways (Bhandari et al., 2013) (Bhandari et al., 2013); this larger set of genes in *Capsicum* suggest gene family expansion and possible lineage-specific adaptations. Combined metabolomic and transcriptomic approaches have uncovered regulatory modules linking tocopherol biosynthetic gene expression to metabolic outcomes and environmental cues, such as stress, ripening, and light exposure (Bhandari et al., 2013).

The integration of *in-silico* transcriptomics, high-throughput RNA sequencing, and experimental validation (qRT-PCR) across fruit development stages provides compelling evidence for direct coupling between biosynthetic gene expression profiles and mature vitamin levels in plant tissues (Aguiar et al., 2023; Ahmad et al., 2021).

Research Gaps:

Genes related to fruit development, ripening, and vitamin biosynthesis in *Capsicum* species have not yet been fully identified, although many such genes have already been reported in other Solanaceae crops like tomato. Therefore, in this study, we aim to conduct a comprehensive analysis of genes and their expression patterns associated with fruit development and vitamin biosynthesis in *Capsicum* species. Additionally, we seek to quantify Vitamin C and Vitamin E content at three fruit developmental stages in three different *Capsicum* species (*C. annuum, C. chinense*, and *C. frutescens*). Using comparative genomics, we will further identify genes involved in Vitamin C and Vitamin E biosynthesis and accumulation across these species. The *Capsicum* genotypes from Northeast India, selected for this study have not been previously investigated for Vitamin C and E biosynthesis/accumulation, nor for the expression of genes associated with fruit development and ripening. This work will provide comprehensive insights into the genetic architecture and gene expression underlying contrasting phenotypes for fruit development and vitamin biosynthesis in *Capsicum species*. Furthermore, molecular markers such as simple sequence repeat (SSR) markers based on genes related to fruit development, ripening, and vitamin biosynthesis/accumulation will be developed, which could be utilized in *Capsicum* breeding programs to manipulate these traits.

Broad Objectives of the Study

To address the research gaps highlighted above in fruit development, ripening, and vitamin biosynthesis, the work has been divided into the following objectives to systematically and stepwise carry out the research work:

1. Identification and expression analysis of genes involved in *Capsicum* fruit development and ripening.

2. Determination of Vitamin C and E content at different fruit developmental stages in contrasting *Capsicum* genotypes.

3. Identification and expression analysis of genes responsible for the regulation of Vitamin C and E in *Capsicum* spp.

CHAPTER 2

MATERIALS & METHODS

turnitin Page 68 of 222 - Integrity Submission Submission ID trn:oid:::1:3347564740

Chapter 2: Materials and Methods

2.1 OBJECTIVE 1: Identification and expression analysis of genes involved in *Capsicum* fruit development and ripening:

2.1.1 Plant materials and growing conditions

In this study, a total of 47 *Capsicum* genotypes, belonging to *C. chinense, C. frutescens,* and *C. annuum,* were selected for experiments. These genotypes were collected from a diverse array of locations, including the Indian regions of Assam, Manipur, Jammu, Mizoram, Uttarakhand, and Nagaland, as well as from the National Bureau of Plant Genetic Resources (NBPGR), Delhi (**Table 2.1 and Fig 3.5**). The selection was designed to represent the broad genetic and phenotypic diversity found within and between these species. To promote healthy and consistent seedling growth, seeds from each genotype were sterilized using a 4% sodium hypochlorite solution, minimizing the risk of microbial contamination. Following sterilization, the seeds were germinated and grown in a glasshouse under controlled conditions, setting the temperatures between 24°C and 26°C, and a 16-hour light period to simulate optimal growth environments. After one month, when the seedlings were four to six-leaf stages, they were transplanted to the experimental fields at Jawaharlal Nehru University (JNU) in New Delhi. Here, the plants were grown under open field conditions until they reached fruit set and maturity. We observed and assessed agronomic traits and fruit development and ripening in natural conditions.

Table 2.1. List of *Capsicum* genotypes used in the study

S. No.	Genotype code number	Genotype name	*Capsicum* Species	Origin	Latitute	Longitude
1	Ca-1	S-19	*C. annuum*	Assam, India	26° 14' N	92° 32' E
2	Ca-2	Sweet long pepper	*C. annuum*	Delhi	28° 35′ N	77° 13' E
3	Cc-3	Lota bhut	*C. chinense*	Assam, India	26° 14' N	92° 32' E
4	Cc-4	Chin-2	*C. chinense*	Manipur, India	24° 44' N	93° 58' E
5	Ca-5	ISC-2	*C. annuum*	Jammu & Kashmir, India	32° 44' N	74° 54' E
6	Ca-6	Mizo-1	*C. annuum*	Mizoram, India	23° 30' N	20° 52' E
7	Cc-7	Chin-9	*C. chinense*	Assam, India	26° 14' N	92° 32' E
8	Ca-8	Can-5	*C. annuum* (Bell Pepper)	Delhi	28° 35' N	77° 13' E
9	Cf-9	CPA	*C. frutescens*	Assam, India	26° 14' N	92° 32' E
10	Cc-10	Chin-3	*C. chinense*	Assam, India	26° 14' N	92° 32' E
11	Ca-11	RANI B	*C. annuum*	Assam, India	26° 14' N	92° 32' E
12	Ca-12	JH-30	*C. annuum*	Assam, India	26° 14' N	92° 32' E
13	Cc-13	NB-5	*C. chinense*	NBPGR, India	28° 35′ N	77° 13' E
14	Ca-14	Kosom moso	*C. annuum*	Manipur, India	24° 44' N	93° 58' E
15	Ca-15	Kalsi	*C. annuum*	Uttarakhand, India	30° 15' N	79° 15' E
16	Ca-16	S-12	*C. annuum*	Assam, India	26° 14' N	92° 32' E
17	Cc-17	S-17	*C. chinense*	Nagaland, India	26° 00' N	94° 20' E
18	Ca-18	19 Assam	*C. annuum*	Assam, India	26° 14' N	92° 32' E
19	Ca-19	NB-23	*C. annuum*	NBPGR, India	28° 35′ N	77° 13' E
20	Cc-20	Chin-37	*C. chinense*	Assam, India	26° 14' N	92° 32' E
21	Ca-21	NB-33	*C. annuum*	NBPGR, India	28° 35′ N	77° 13' E
22	Cc-22	NB-28	*C. chinense*	NBPGR, India	28° 35′ N	77° 13' E
23	Cf-23	JH-6	*C. frutescens*	Assam, India	26° 14' N	92° 32' E
24	Cc-24	JH-21	*C. chinense*	Assam, India	26° 14' N	92° 32' E
25	Ca-25	Can red 90	*C. annuum* (Bell Pepper)	Delhi	28° 35' N	77° 13' E
26	Ca-26	S-20	*C. annuum*	Nagaland, India	26° 00' N	94° 20' E
27	Ca-27	JH-4	*C. annuum*	Assam, India	26° 14' N	92° 32' E
28	Ca-28	Ms Pepper yellow	*C. annuum*	Manipur, India	24° 44' N	93° 58' E
29	Ca-29	Billawar	*C. annuum*	Jammu & Kashmir, India	32° 44' N	74° 54' E
30	Cc-30	G-Assam	*C. chinense*	Assam, India	26° 14' N	92° 32' E
31	Co-31	12 Assam	*Other Capsicum species*	Assam, India	26° 14' N	92° 32' E
32	Cc-32	Umorock	*C. chinense*	Manipur, India	24° 44' N	93° 58' E
33	Ca-33	S-14	*C. annuum*	Assam, India	26° 14' N	92° 32' E

34	Cc-34	10 Assam	*C. chinense*	Assam, India	26° 14' N	92° 32' E
35	Ca-35	NB-6	*C. annuum*	NBPGR, India	28° 35′ N	77° 13' E
36	Cc-36	7 Chinense	*C. chinense*	Assam, India	26° 14' N	92° 32' E
37	Cf-37	Frutescens 4	*C. frutescens*	Manipur, India	24° 44' N	93° 58' E
38	Co-38	S-2	Other *Capsicum species*	Assam, India	26° 14' N	92° 32' E
39	Ca-39	JH-23	*C. annuum*	Assam, India	26° 14' N	92° 32' E
40	Cf-40	Korom moso	*C. frutescens*	Manipur, India	24° 44' N	93° 58' E
41	Ca-41	JH-27	*C. annuum*	Assam, India	26° 14' N	92° 32' E
42	Ca-42	JH-29	*C. annuum*	Assam, India	26° 14' N	92° 32' E
43	Ca-43	JH-5	*C. annuum*	Assam, India	26° 14' N	92° 32' E
44	Ca-44	S-11	*C. annuum*	Mizoram, India	23° 30' N	20° 52' E
45	Ca-45	11 Assam	*C. annuum*	Assam, India	26° 14' N	92° 32' E
46	Cc-46	Chin-33	*C. chinense*	Assam, India	26° 14' N	92° 32' E
47	Ca-47	Dudu	*C. annuum*	Jammu & Kashmir, India	32° 44' N	74° 54' E

2.1.2 Identification of genes involved in fruit development and ripening

(a) Selection of tomato genes for fruit development/ripening and identification of chilli orthologs

We selected a set of genes previously reported to be involved in fruit development and ripening in tomato (Listed in **Table 3.2**). For each of these selected tomato genes, we retrieved the coding sequences (CDS) from the NCBI database, ensuring we worked with well-annotated and reliable genetic data. We then employed this tomato gene sequences as input queries for the BLASTN search (Altschul et al., 1990) against the *C. frutescens* reference genome (GCF_000710875.1_Pepper_Zunla_1_Ref_v1.0_genomic. fa). This comparative genomics approach enabled us to search for orthologous genes in the *Capsicum* genome, expecting similar functions due to their evolutionary relationship and sequence similarity.

To ensure that our identification of potential *Capsicum* orthologs was both accurate and

meaningful, we applied stringent selection criteria: only those matches with an E-value lower than 1e−10 and a sequence identity of at least 80% were considered. By setting these high thresholds, we minimized the risk of including unrelated or distantly related genes, focusing instead on those most likely to share functional roles with their tomato counterparts

(a) Expression analysis of *Capsicum* fruit developmental and ripening genes using transcriptome data

To investigate the expression patterns of genes involved in fruit development and ripening in *Capsicum*, we drew upon transcriptome datasets previously generated in our laboratory for three species: *C. annuum* (PRJNA505972), *C. chinense* (PRJNA327797), and *C. frutescens* (PRJNA327800). These datasets were created using RNA extracted from fruits collected at three key developmental stages namely, early fruit [20 days post anthesis, DPA), breaker fruit (30–45 DPA), and mature fruit (45–60 DPA] (Chhapekar et al., 2020).

The available transcriptome data was analyzed using a standard analysis pipeline (Chhapekar et al., 2020; Conesa et al., 2016; Qin et al., 2014). The raw sequencing reads were first assessed for quality using FastQC (v0.11.5). Briefly, the low-quality reads (with a Phred score below 20) and adapter sequences were trimmed away using TrimGalore (v0.4.4) to ensure that only high-quality data were used for further analysis. The resulting clean reads were then mapped to the *C. annuum* reference genome (Pepper Zunla 1 Ref_v1.0 assembly) using TopHat v2.1.1 (Kim et al., 2013; Trapnell et al., 2012) with default parameters. This rigorous workflow provided a solid foundation for analyzing gene expression profiles related to fruit development and ripening

across different *Capsicum* species and developmental stages. Once the sequencing reads were mapped, we assembled them and calculated transcript abundance using Cufflinks v2.2.1 (Trapnell et al., 2012). To focus on reliably expressed genes, we filtered transcripts with an FPKM of at least 0.2. We then examined how gene expression changed across different fruit developmental stages using Cuffdiff v2.2.1, applying rigorous thresholds of p-value ≤ 0.01 and a false discovery rate (FDR) ≤ 0.05 to ensure that only statistically significant differences were considered.

2.1.3 RNA isolation and quantitative Real Time -PCR analysis

We isolated total RNA from early fruit (20 days post anthesis, DPA), breaker fruit (30–45 DPA), and mature fruit (45–60 DPA)—as well as from leaves and fully opened flowers. These tissues were obtained from three distinct genotypes representing *C. chinense* (NB5), *C. frutescens* (MCM), and *C. annuum* (JH-23) (JH23) with differences in fruit sizes. RNA isolation was performed using the RBC Bioscience RNA isolation kit following the manufacturer's instructions to ensure consistency and high-quality RNA recovery across all samples. To confirm both the quality and quantity of the extracted RNA, we used a two-step assessment process. First, RNA integrity was evaluated by running samples on a 1% agarose gel, which allowed us to visually inspect for intact, high-quality RNA. Next, RNA concentrations were measured using a NanoDrop 1000 spectrophotometer (Thermo Scientific), ensuring that each sample met the requirements for downstream applications. For cDNA synthesis, we used one microgram of total RNA from each sample and performed reverse transcription with the SuperScript III First-Strand System (Invitrogen, USA), carefully adhering to the manufacturer's guidelines. This step

converted our RNA into complementary DNA (cDNA), which serves as the template for quantitative real-time PCR (qRT-PCR). To accurately measure gene expression, we designed specific primers using Primer Express 3.0.1 software, ensuring precise amplification of our target genes. The qRT-PCR assays were conducted using SYBR Premix Ex Taq (Clontech, USA) on the ABI7500 Fast system (Applied Biosystems, USA). The thermal cycling protocol began with an initial denaturation at 95°C for 2 minutes, followed by 40 cycles of 15 seconds at 95°C and 1 minute at 60°C. This protocol was optimized for sensitivity and specificity. To guarantee the reliability of our results, each qRT-PCR experiment included three independent biological replicates and three technical replicates per sample. We used the actin gene as an internal control to normalize the expression data, which is crucial for compensating for any variability in RNA input or reverse transcription efficiency. Finally, we used the $2^{-[\Delta\Delta Ct]}$ method to compare gene expression levels across different tissues in *Capsicum* genotypes to understand how important genes are regulated during fruit development and ripening.

2.1.4 Simple sequence repeats mining and primer designing

To develop a comprehensive set of simple sequence repeat (SSR) markers based on fruit development and ripening related genes in *Capsicum*, we took an inclusive approach by analyzing not only the gene sequences themselves but also extending our search to include 5 kilobases (Kb) of DNA both upstream and downstream of each gene. This strategy was designed to capture SSRs that might be in regulatory or flanking regions, which can be important for correlation of these genes with fruit phenotypes and genetic diversity studies. The identification of SSRs was carried out using the WebSat online tool (Martins et al., 2009), which is well-suited

for mining repetitive DNA motifs. In our search, we focused on regions containing di-, tri, tetra, penta-, and hexa-nucleotide repeats, setting a threshold that each repeat motif must occur at least six times in succession. This criterion ensures that the SSRs we identify are long enough to be informative and reliable for downstream genetic analysis. To further refine our SSR selection, we established a minimum spacing of 100 base pairs (bp) between adjacent SSRs. If two SSRs were found within this distance, they were considered overlapping, and only one was retained for further analysis. This step was important to avoid redundancy and to ensure that each SSR marker would be distinct and easily amplifiable in later experiments. Once we had identified a robust set of SSRs, we proceeded to design PCR primers for their amplification using Primer3 software (Martins et al., 2009), which is integrated into the WebSat platform. We applied several stringent criteria to optimize primer performance: primers were designed to be between 18 and 27 bp in length, with 22 bp as the ideal; the melting temperature (Tm) was set between 50°C and 65°C, aiming for an optimal 60°C; and the GC content was maintained between 40% and 80% to promote stable primer binding. We also specified that the expected PCR product (amplicon) size should range from 100 to 400 bp, which is suitable for routine PCR and gel electrophoresis. To further enhance primer specificity and efficiency, we limited the maximum temperature difference between forward and reverse primers to 1°C and capped the maximum 3′ end stability at 250 (250 ≈ ΔG of –2.5 kcal/mol).

2.1.5 DNA isolation and SSR genotyping and scoring

Genomic DNA was isolated from leaves of 47 distinct *Capsicum* genotypes using CTAB method (Rogers & Bendich, 1985), which is known for its efficiency in yielding pure DNA suitable for downstream applications. To ensure the DNA was both intact and of sufficient quantity, we checked the samples on a 1% agarose gel for visual inspection of integrity and used a Nanodrop spectrophotometer to precisely measure DNA concentration.

For the genotyping of simple sequence repeat (SSR) markers, the following PCR reaction was used: 25 ng of template DNA combined with a carefully balanced mix of reagents: 1× PCR buffer (comprising 50 mM KCl and 20 mM Tris–HCl at pH 8.4) to maintain optimal reaction conditions, 1.5 mM $MgCl_2$ to facilitate the activity of the DNA polymerase, 0.125 mM of each dNTP as the building blocks for new DNA strands, 0.5 µM of each primer for specific targeting of SSR regions, and 0.5 units of iTaq DNA polymerase to catalyze the reaction. The PCR amplifications were performed using an Eppendorf thermal cycler (Germany), following a rigorously optimized temperature profile. The process began with an initial denaturation at 95°C for 4 minutes to ensure complete separation of DNA strands. This was followed by 35 cycles, each consisting of denaturation at 94°C for 2 minutes, annealing at a primer-specific temperature (ranging from 50°C to 65°C) for 45 seconds to allow primers to bind, and extension at 72°C for 30 seconds to synthesize new DNA. A final extension at 72°C for 2 minutes ensured that all DNA fragments were fully extended. Following amplification, we resolved the PCR products using a 10% polyacrylamide gel electrophoresis (PAGE) system from C.B.S Scientific Co., running the gels vertically at 250 V for 6 hours. This high-resolution technique enabled us to separate DNA fragments with even minimal size differences, which is crucial for accurately

distinguishing SSR alleles. The DNA bands were then visualized using silver staining (Tegelström, 1992), a sensitive method that produces clear, sharp bands for analysis. For each SSR marker, we carefully documented the presence or absence of bands and any differences in fragment length among the different *Capsicum* genotypes. These results were then converted into precise allele sizes (measured in base pairs) using a fragment size calculator, creating a detailed genetic profile for each sample.

2.1.6 Genetic diversity analysis of germplasm using SSR markers

To evaluate the genetic diversity and discriminatory power of SSR markers in *Capsicum* germplasm, we calculated the Polymorphism Information Content (PIC) for each marker using the formula: $PIC = 1 - \Sigma xi^2$ where xi is the relative frequency of the ith allele (Anderson et al., 1993). The PIC value serves as a key indicator of marker informativeness, with higher values (>0.5) suggesting greater utility in distinguishing genetic variants (Botstein et al., 1980). These computations were performed using PIC calculation software, following the method described by (Anderson et al., 1993). Markers with high PIC values are particularly valuable for genetic diversity studies, as they provide deeper insights into allelic variations among germplasm accessions (V. Kumar & Yadav, 2019). For the characterization of genetic variation among our 47 *Capsicum* genotypes, we considered the following criteria: --

- Number of alleles per locus (na): This reflects the total allelic richness at each SSR locus, with higher numbers indicating greater genetic variability (İnce & Karaca, 2015).

- Effective number of alleles (ne): This accounts for allele frequency distribution, providing a more accurate measure of allelic diversity (Kimura & Crow, 1964).
- Shannon's Information Index (I): This index measures both allelic richness and evenness, offering a comprehensive view of genetic complexity (Govindaraj et al., 2015; Konopiński, 2020; Morris et al., 2014).
- Expected Homozygosity and Heterozygosity: These metrics reveal the probability of individuals being genetically identical (Ho) or different (He) at a given locus, highlighting population structure and breeding behavior (Nei, 1973).
- Nei's Gene Diversity Index (H): This provides a robust estimate of overall genetic variation within the population, aiding in comparisons across different germplasm collections (Nei, 1973).
- All statistical analyses were conducted using POPGENE32 v1.32 (Yeh et al., 1997), a widely used software for population genetic studies. Genetic diversity assessments of 47 germplasm were calculated as reported earlier (Albrecht et al., 2012; Nankar et al., 2016). The use of these parameters ensures thorough evaluation of SSR marker utility while capturing the full spectrum of genetic diversity in *Capsicum* germplasm (Paran & Van Der Knaap, 2007; Pickersgill, 2007).

2.2 OBJECTIVE 2: Determination of Vitamin C and E content at different fruit developmental stages in contrasting *Capsicum* genotypes.

2.2.1 Plant materials and growing conditions

For this study, we choose a diverse set of 30 *Capsicum* genotypes (different from the genotypes taken in Objective 1), including 10 genotypes each from *C. chinense*, *C. frutescens*, and *C. annuum* (as shown in **Table 2.2 and Fig 3.7**). The plant growing condition described for Objective 1 under 2.1.1 was followed.

To estimate vitamin content, fruits were collected at three different stages of development indicated below: -:

- Early fruit (EF): picked at 20 days post anthesis (DPA), representing the early phase of fruit growth,
- Breaker fruit (BF): gathered between 30-45 DPA, when the fruit begins to ripen and change color, and
- Mature fruit (MF): collected at 45-60 DPA, when the fruit is fully ripe.

For each stage, fruits were sampled from three different plants per genotype, and this process was repeated in three biological replicates (Dubey et al., 2019) to ensure the data were robust and reliable.

Table 2.2. Vitamin C and Vitamin E contents in early fruit (EF), breaker fruit (BF), and mature fruit (MF) stages in 30 *Capsicum* genotypes belonging to *C. chinense* (Cc1 to Cc10), *C. frutescens* (Cf1 to Cf10), and *C. annuum* (Ca to Ca10).

S. No.	Genotype code number	Genotype name	*Capsicum* Species	Origin	Latitute	Longitude
1	Cc-1	BJCF	*C. chinense*	Assam, India	26° 14' N	92° 32' E
2	Cc-2	Chin-2	*C. chinense*	Assam, India	26° 14' N	92° 32' E
3	Cc-3	Chinense	*C. chinense*	Assam, India	26° 14' N	92° 32' E
4	Cc-4	BJ Small	*C. chinense*	Assam, India	26° 14' N	92° 32' E
5	Cc-5	Chin-1(old)	*C. chinense*	Assam, India	26° 14' N	92° 32' E
6	Cc-6	NE-17	*C. chinense*	Nagaland, India	26° 00' N	94° 20' E
7	Cc-7	BJ Medium	*C. chinense*	Assam, India	26° 14' N	92° 32' E
8	Cc-8	NE-3	*C. chinense*	Nagaland, India	26° 00' N	94° 20' E
9	Cc-9	CC GHY	*C. chinense*	Assam, India	26° 14' N	92° 32' E
10	Cc-10	BJ	*C. chinense*	Assam, India	26° 14' N	92° 32' E
11	Cf-1	JH-25	*C. frutescens*	Assam, India	26° 14' N	92° 32' E
12	Cf-2	NE-7	*C. frutescens*	Nagaland, India	26° 00' N	94° 20' E
13	Cf-3	C. Spp 5	*C. frutescens*	Assam, India	26° 14' N	92° 32' E
14	Cf-4	NE-21	*C. frutescens*	Nagaland, India	26° 00' N	94° 20' E
15	Cf-5	C. minimum	*C. frutescens*	Assam, India	26° 14' N	92° 32' E
16	Cf-6	Kherino Khezao	*C. frutescens*	Assam, India	26° 14' N	92° 32' E
17	Cf-7	Frut White	*C. frutescens*	Meghalaya, India	25° 47' N	92° 32' E
18	Cf-8	JH-17	*C. frutescens*	Assam, India	26° 14' N	92° 32' E
19	Cf-9	KOK-1	*C. frutescens*	Assam, India	26° 14' N	92° 32' E
20	Cf-10	Frutescens	*C. frutescens*	Meghalaya, India	25° 47' N	92° 32' E
21	Ca-1	JH-26	*C. annuum*	Assam, India	26° 14' N	92° 32' E
22	Ca-2	S-15	*C. annuum*	NBPGR, India	28° 35′ N	70° 18′ E
23	Ca-3	S-17	*C. annuum*	NBPGR, India	28° 35′ N	70° 18′ E
24	Ca-4	ISC Jammu	*C. annuum*	Jammu & Kashmir, India	32° 44' N	74° 54' E
25	Ca-5	NE-11	*C. annuum*	Nagaland, India	26° 00' N	94° 20' E
26	Ca-6	Kosom Moso	*C. annuum*	Nagaland, India	26° 00' N	94° 20' E
27	Ca-7	NB-32	*C. annuum*	NBPGR, India	28° 35′ N	70° 18′ E
28	Ca-8	S-2	*C. annuum*	NBPGR, India	28° 35′ N	70° 18′ E
29	Ca-9	NB-2	*C. annuum*	NBPGR, India	28° 35′ N	70° 18′ E
30	Ca-10	GHY-YR	*C. annuum*	Assam, India	26° 14' N	92° 32' E

2.2.2 Samples preparation and estimation of Vitamin C and E content through HPLC

(a) Vitamin C samples preparation and analysis

To determine the Vitamin C content in the 30 different *Capsicum* genotypes (listed in, **Table 2.2 and Fig 3.7**), we used a well-established method with some minor modifications (Daood et al.,2006) to suit our study. The process began by collecting about 0.5 grams of fresh fruit tissue from each pepper genotype. To ensure the tissue was thoroughly broken down and to preserve Vitamin C, the samples were ground into a fine powder using a mortar and pestle, with liquid nitrogen added to keep everything cold and prevent any loss of Vitamin C during the grinding process. After the tissue was fully crushed, the powder was transferred into a 15 ml falcon tube containing 10 ml of Metaphosphoric acid (MPA) solution. This acid solution plays a crucial role in stabilizing Vitamin C, protecting it from oxidation and degradation during the extraction. The mixture was then shaken well to mix the tissue with the solution, ensuring that the Vitamin C was fully extracted from the plant material. Next, the samples were placed in a centrifuge and spin at 5000 rpm for 5 minutes. This high-speed spinning separated the solid plant debris from the liquid, leaving a clear supernatant that contained the dissolved Vitamin C. The supernatant was carefully poured into a fresh falcon tube. To remove any remaining tiny particles, the liquid was filtered through a 0.2 µm syringe filter. The filtered solution was then collected in small vials, ready for analysis, for the analysis 1ml samples were transferred to HPLC vials and we used technical triplicates vitamin.

The actual measurement of Vitamin C was performed using High-Performance Liquid Chromatography (HPLC), a precise and reliable technique for detecting and quantifying specific

compounds in a mixture. To ensure accuracy, the researchers prepared a standard curve using freshly made solutions of ascorbic acid (Vitamin C) at known concentrations: 1, 0.5, 0.25, and 0.125 mg/ml. By running these standards through the HPLC, they created a reference for interpreting the results from their pepper samples. Finally, the Vitamin C content in each unknown sample was calculated by comparing its HPLC reading to the standard curve. This thorough and carefully controlled process allowed us to accurately measure and compare Vitamin C levels across all 30 *Capsicum* genotypes, providing valuable insight into the nutritional diversity present in these peppers.

Analysis of Vitamin C content was conducted utilizing the Dionex Ultimate 3000 model HPLC, manufactured by Thermo Fisher. A C18 column with dimensions of 150 × 4.6 mm and a particle size of 3 µm served as the stationary phase in this study. For the determination of Vitamin C, the mobile phase employed was 100% acetonitrile. Separation was achieved at a flow rate of 1 mL/min for Vitamin C. Vitamin C was detected at a wavelength of 254 nm.

(b) Vitamin E samples preparation and analysis

To investigate the Vitamin E content in chili peppers, the research team analyzed the same 30 *Capsicum* genotypes that were previously studied for Vitamin C (see **Table 2.2 & Fig 3.7**). We use a standard protocol (Saini & Keum, 2016) making sure to preserve the integrity of the Vitamin E throughout the process. The procedure began with the careful selection of 0.5 grams of fresh fruit tissue from each genotype. To break down the tissue while protecting the delicate Vitamin E molecules, the samples were ground into a fine powder using a mortar and pestle,

with liquid nitrogen. This step is crucial because Vitamin E is sensitive to both heat and oxidation and light too. After grinding, the powdered tissue was transferred into 15 ml falcon tubes that contained 10 ml of chilled acetone mixed with 1% butylated hydroxytoluene (BHT). The addition of BHT, a powerful antioxidant, helps prevent Vitamin E from breaking down during the extraction process. The tubes were then shaken thoroughly to ensure that the Vitamin E dissolved into the acetone solution. The next step involved centrifuging the samples at 5000 rpm for 5 minutes at 4°C. This rapid spinning separated the solid plant debris from the liquid. The clear supernatant, now rich in extracted Vitamin E, was carefully collected for further processing.

To concentrate the Vitamin E used a rotary vacuum evaporator to gently remove the acetone from the supernatant. This method is gentle enough to avoid damaging the Vitamin E, which can be sensitive to both heat and air. The dried residue was then re-dissolved in 5 ml of chilled acetone (again with 1% BHT) to maximize recovery of the vitamin. The solution was filtered through a 0.2 µm syringe filter to remove any remaining particles, ensuring a clean sample for analysis 1ml samples were transferred in HPLC vials and we used technical triplicates.

For the actual measurement, we used High-Performance Liquid Chromatography (HPLC), the Dionex Ultimate 3000 model, manufactured by Thermo Fisher. A C18 column with dimensions of 150 × 4.6 mm and a particle size of 3 µm served as the stationary phase in this study. To calibrate the instrument and ensure reliable results, we prepared a standard curve using known concentrations of Vitamin E (specifically γ-tocopherol) at 1, 0.5, 0.25, and 0.125 mg/ml. By

comparing the HPLC results from *Capsicum* samples to this standard curve, we determined the Vitamin E content in each *Capsicum* genotype fruits.

For the determination of Vitamin E, the mobile phase consisted of 100% methanol. Separation was achieved at a flow rate of 1.8 mL/min for Vitamin E. Vitamin E was detected at a wavelength of 290 nm.

2.3 OBJECTIVE 3: Identification and expression analysis of genes responsible for regulation of Vitamins C and E in *Capsicum* spp.

2.3.1 Identification of genes involved in Vitamin C and E biosynthesis

The method described for Objective 1 under 2.1.2 Identification of genes involved in fruit development and ripening under section (a) Selection of tomato genes for fruit development/ripening and identification of chili orthologs was followed.

2.3.2 Expression analysis of Vitamin C and E genes in transcriptome data

The method described for Objective 1 under 2.1.2 Identification of genes involved in fruit development and ripening under section (b) Expression analysis of *Capsicum* fruit developmental/ripening genes using transcriptome data was followed.

2.3.3 RNA Isolation and Quantitative Real-Time PCR (qRT-PCR) Analysis

The method described for Objective 1 under section 2.1.3 RNA isolation and qRT-PCR analysis was used.

2.3.4 Development of gene based simple sequence repeat (SSR) markers

The method described for Objective 1 under 2.1.4 Simple sequence repeats (SSR) mining and primer designing was followed.

2.3.5 Identification of Insertion-deletions (InDels) in Vitamin C and E genes

To identify insertion and deletion mutations (InDels), we used Clustal Omega (Madeira et al., 2022), which compares multiple DNA sequences at the same time. We focused on the full gene sequences, as well as the 2 kb upstream promoter regions of the Vitamin C and E genes. This analysis was done using DNA sequences from three different *Capsicum* genomes: *C. baccatum, C. chinense,* and *C. annuum.* By arranging these sequences alongside one another for analysis, through Clustal Omega we found out places where certain DNA segments were either added (insertions) or missing (deletions) in one *Capsicum* species compared to the other species. Detecting these InDels is important because such DNA variations can influence gene activity and potentially determine the levels of Vitamin C and E produced by each *Capsicum* species.

2.3.6 Identification of homologous genes for Vitamin C and Vitamin E biosynthesis and synteny analysis among Solanaceae genomes

To study how Vitamin C and Vitamin E genes are organized and conserved across different plant species, we did analysis of genome wide conserved syntenic segments (CSSs) and collinearity of three *Capsicum* species *C. annuum, C. baccatum*, and *C. chinense* as well as in *S. lycopersicum*. Our main goal was to find out if these vitamin genes are in similar regions and arranged in the same order on the chromosomes of these plants. This type of comparison helps us see how these genes have been inherited and whether they have changed or remain unchanged as the plants have evolved over time. For this purpose and analysis, we used MCScanX (computer program) (Wang et al., 2012) to identify clusters of genes, called conserved syntenic segments (CSSs) that are found in the same order across these species. To do this, we first

collected the complete DNA sequences for all three *Capsicum* species- *C. chinense* (GCA_002271895.2), *C. baccatum* (GCA_002271885.2), and *C. annuum* (GCA_000710875.1) from the NCBI database, and for tomato from the Phytozome website v12(https://phytozome.jgi.doe.gov/pz/portal.html). In the Next step, we wanted to determine which genes in one species are like the genes found in the other species, so we used BLASTp, which compares the total protein products of genes. We set some criteria for what counts as a good match: the genes had to be at least 75% identical, cover at least 80% of the gene's length, and have a very low chance of matching by accident (e-value <1e-03). After finding these comparable vitamin genes, we determined their positions on the chromosomes using TBtools v1.068; (C. Chen et al., 2020). Collinear and homologous vitamin genes, with their physical positions on 12 different chromosomes (chrs), to make our findings easier to understand, we created special diagrams using the circlize R package (Gu et al., 2014), which show how the vitamin genes are connected and positioned across the 12 chromosomes in each spp.

Chapter 3
Results & Discussion

Chapter 3 Results & Discussion

3.1. OBJECTIVE 1: Identification and expression analysis of genes involved in *Capsicum* fruit development and ripening:

3.1.1. Identification of genes involved in fruit development and ripening in *C. annuum* genome

After reviewing many research studies, a total of 32 genes from tomato were chosen that are reported to play a role in tomato fruit development and ripening **(Table 3.1 and Fig. 3.1)** and created a representative network diagram of these genes to help understand their roles **(Fig. 3.1)**. These 32 genes are taken from all 12 tomato chromosomes, i.e. 6 genes from chromosome 10, followed by five in Chromosome 1, 4 in chromosome 3, 3 in chromosome 6, and 2 genes each from chromosomes 2, 4, 5, 7, 8, and 9, respectively. Chromosomes 11 and 12 each have just one gene **(Fig. 3.2)**. The use of these 32 tomato genes against the *Capsicum* genome identified 41 matching genes with at least 80% similarity **(Table 3.2)**. For seven tomato genes—(i) *Tomato Constitutive Triple Response 1* (*TCTR1*), (ii) *Lipoxygenase* (*TOMLOX-C*), (iii) *Riprning Inhibitor* (*MADS-RIN*), (iv) *Chalcone Synthase I* (*CHS-1*), (v) *Glutamate Dehydrogenase*(*GDH-1*), (vi) *DNA Demethylase* (*SlDML-2*), and (vii) *Cysthionine Gama Synthase* (*CGS*)—two or more similar/orthologous genes were identified in the *Capsicum* genome **(Fig. 3.2).** Of the 32 tomato genes taken for study, 18 showed conserved synteny on corresponding *Capsicum* chromosomes, the remaining 14 were scattered across the different chromosomes or unanchored scaffolds of *Capsicum* genome.

Table 3.1 List of selected genes reported to be associated with fruit development and ripening in tomato (Dubey et al. 2019)

S. No.	Tomato fruit development/ripening genes	Function of Gene	Reference/s
1	*Abscisic stress-ripening protein 1(ASR1)*	It is transcriptional regulator of β-hex and thus involved in fruit softening during ripening	(Irfan et al., 2014)
2	*Phytoene synthase 1, chloroplastic (PSY1)*	Activates lycopene biosynthesis	(Giorio et al., 2008)
3	*Ethylene-inducible CTR1-like protein kinase (TCTR1)*	It is ethylene receptor and involved in ethylene signaling during fruit ripening	(Leclercq et al., 2002)
4	*Lipoxygenase (TomloxC)*	It is involved in the generation of volatile C6 aldehyde and alcohol flavor compounds during fruit ripening	(Chen GuoPing et al., 2004)
5	*MADS-box transcription factor (MADS-RIN)*	Associated with ripening time. Shorter ripening time of fruit was observed in *SlMADS1*-silenced tomatoes	(Zhang et al., 2018)
6	*Polygalacturonase-2a (PG2)*	Involved in cell wall metabolism including depolymerization and solubilization and polyuronide degradation during fruit maturation	(Sitrit & Bennett, 1998; Tucker, 1990)
7	*TAGL1 transcription factor (TAGL1)*	Associated with fruit pigmentation. Silencing of the *TOMATO AGAMOUS-LIKE 1 (TAGL1) MADS* box gene results in altered fruit pigmentation	(Itkin et al., 2011)
8	*NAC domain protein (NAC1)*	Functions as a positive regulator of fruit ripening by affecting ethylene synthesis and carotenoid accumulation	(Zhu MingKu et al., 2014)
9	*Ehtylene receptor 2-like (ETR4)*	It is ethylene receptor and involved in ethylene signaling during fruit ripening	(Kevany et al., 2007)
10	*Pramotor binding protein 1-like CNR(Lespl-CNR)*	It is transcription factor that targets *LeMADS-RIN, LeHB1, SlAP2a* and *SlTAGL1*; and thus affects fruit pigmentation and ripening	(W. Chen et al., 2015; Manning et al., 2006)
11	*Chalcone synthase 1 (CHS1)*	Involved in flavonoid biosynthesis	(Ariizumi et al., 2013)
12	*Agamous-like MADS-box protein AGL8 homolog (FUL2)*	Involved in lycopene synthesis, silence line showed altered pigmentation	(Bemer et al., 2012; S. Wang et al., 2014)
13	*MADS-box transcription factor, TDR4 transcription factor (FUL1/TDR)*	Involved in production of cuticle components and volatiles, and glutamic acid (Glu) accumulation during fruit development/ripening	(Fujisawa et al., 2014)
14	*Glutamate dehydrogense (GDH1)*	It is involved in anoxia-reoxygenation during fruit ripening	(Tsai et al., 2016)

15	*Golden 2-like protein (GLK2)*	It is transcription factor that regulates plastid and chlorophyll levels thus responsible for coloration during fruit ripening	(Nguyen et al., 2014)
16	*Protein ros 1-like (SlDML2)*	It is responsible for DNA demethylation and targets *CNR* and *RIN* genes and thus affects fruit development.	(Gallusci et al., 2016; Lang et al., 2017)
17	*Apatela 2-like protein (AP2)*	Transcription factor that regulates fruit ripening via regulation of ethylene biosynthesis and signaling	(Karlova et al., 2011)
18	*Lutescent 2*	Associated with ripening duration. Delayed ripening has been observed in mutant gene	(Barry et al., 2012)
19	*Hydroquinone glucosyltransferase (LOC101244237)*	It is involved in fruit aroma and perhaps targets flavonoid, flavanols, hydroquinone, xenobiotics and chlorinated pollutants	(Louveau et al., 2011)
20	*Short-chain dehydrogenase-reductase (SlscADH1)*	Indirect affects the catabolism of phospholipids and/or integrity of membranes	(Baldassarre et al., 2015)
21	*Senescence-inducible chloroplast stay-green protein 1 (SGR1)*	It is target of *RIN* and also has physical interaction with *phytoene synthase 1 (PSY1)* and promotes the biosynthesis of carotenoids in tomato. Besides, also involved in ethylene signaling	(Wang et al., 2016)
22	*Alternative oxidase 1a (AOX1a)*	It is involved in fruit development, ripening, carotenoids, respiration and ethylene production	(Xu et al., 2012)
23	*Ethylene response factor (ERF6)*	It is involved in carotenoid biosynthesis and ethylene signaling during ripening	(Lee et al., 2012)
24	*Cystathionine gamma synthase (CGS)*	It is involved in methionine synthesis (required for ethylene production) in tomato fruit ripening	(Katz et al., 2006)
25	*Auxin response factor (ARF4)*	Associated with ripening-related fruit quality traits including enhanced fruit density at mature stage, increased firmness, prolonged shelf-life and reduced water (weight) loss at red ripe stage	(Sagar et al., 2013)
26	*Glycosyltransferase (NSGT1)*	It converts the cleavable diglycosides of the smoky-related phenylpropanoid volatiles into non cleavable triglycosides, thereby preventing their deglycosylation and release from tomato fruit upon tissue disruption.	(Tikunov et al., 2013)
27	*Spermidine synthase (spdsyn)*	Associated with reduced shriveling and decay symptom development.	(Nambeesan et al., 2010)
28	*Alpha-mannosidase (LOC100500729)*	It is involved in fruit softening during ripening	(Irfan et al., 2014)
29	*Glycoalkaloid metabolism 1 (GAME1)*	It is involved in steroidal glycoalkaloids (SGAs) metabolism, *GAME1* in the glycosylation of SAs and in reducing the toxicity of SA metabolites to the plant cell during ripening	(Itkin et al., 2011)

30	*Xyloglucan endotransglucosylase-hydrolase (XTH5)*	It is involved in fruit softening and wall-loosening during ripening	(Miedes et al., 2010)
31	*MADS-box protein 1 (LOC543884)*	It interacts with *histone acetyltransferases (HAT)* and *histone deacetylases (HDAC)* and control downstream genes and involved in organ differentiation during fruit development	(Gaffe et al., 2011)
32	*Beta-hexosaminidase 1(LOC100529103)*	β-D-N-acetylhexosaminidase involved in ripening-associated fruit softening	(Meli et al., 2010)

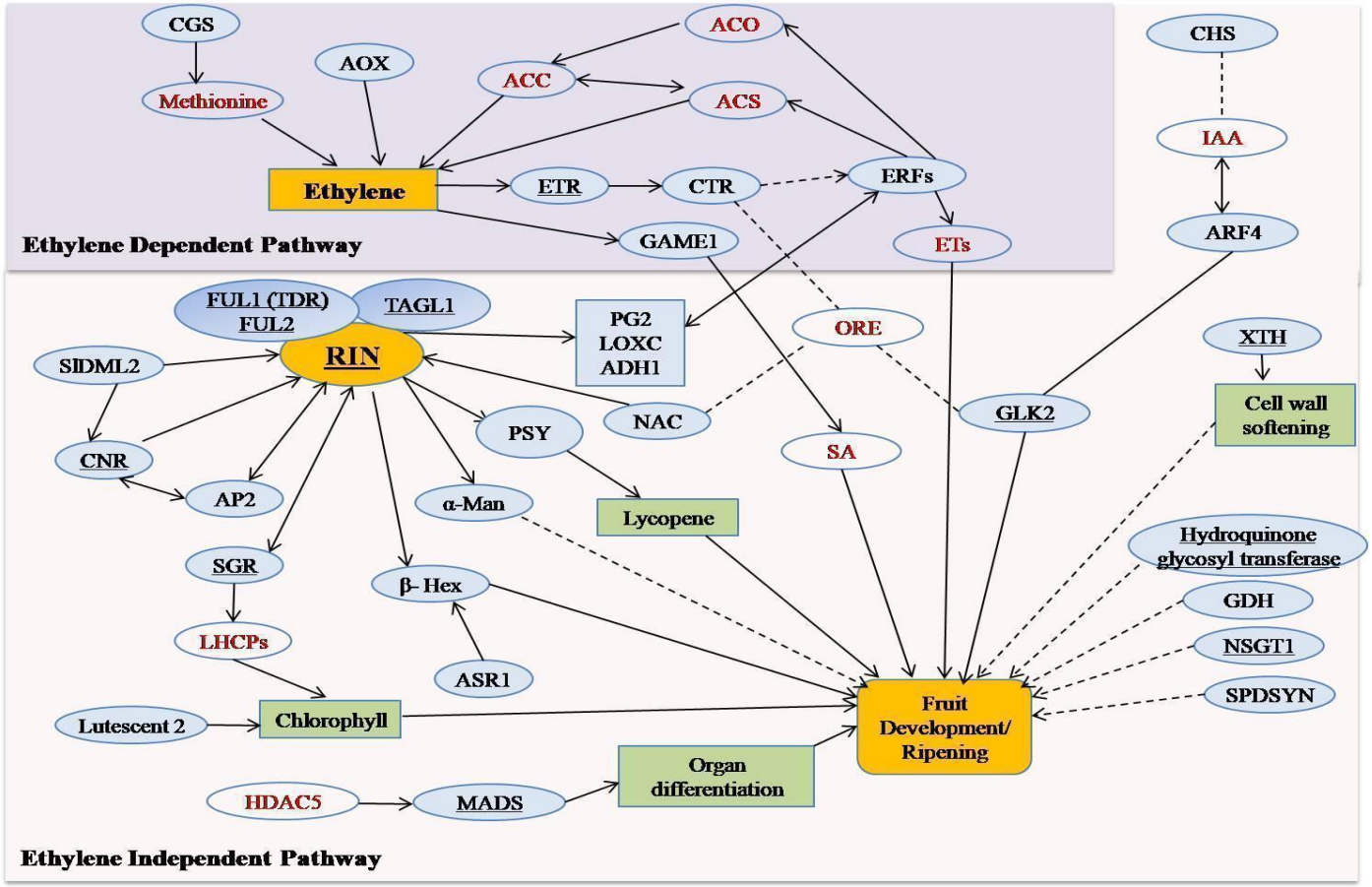


Fig. 3.1 Illustrative representation of the fruit development and ripening pathway in tomato. The figure was prepared based on information compiled from various studies (References listed in **Table 3.1**). A total of 32 genes, shown in black, were selected for ortholog identification in *Capsicum* species (Dubey et al., 2019).

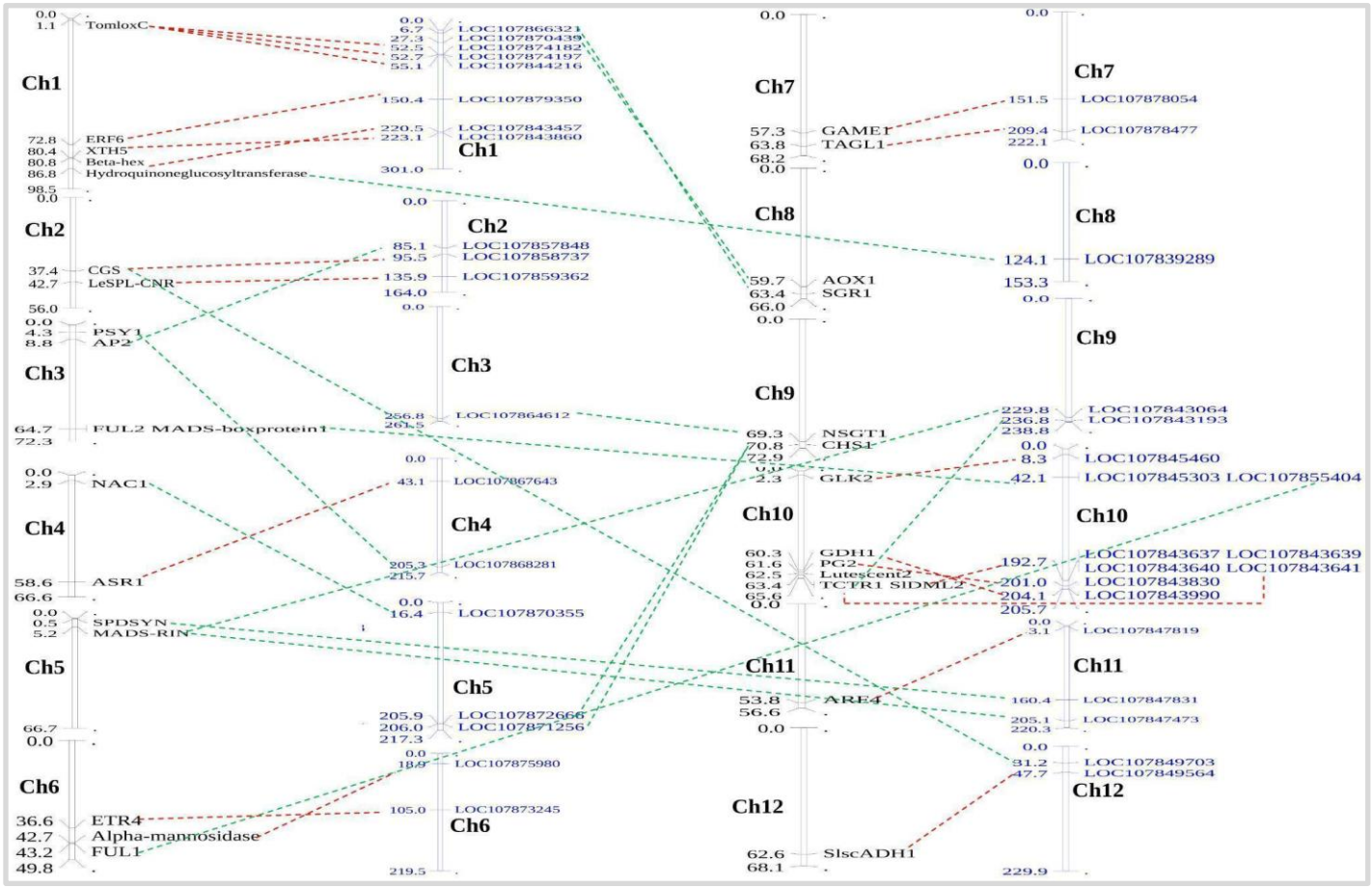


Fig. 3.2 Comparative chromosomal maps of tomato (in black) and *Capsicum* (in blue) showing orthologous genes involved in fruit development/ripening. Name of genes and their chromosomal positions are given on right and left to the bars, respectively. Orthologous genes located on the same chromosomes in both crops are represented by red dashed lines, while those on different chromosomes are shown with green dashed lines (Dubey et al., 2019).

Table 3.2 List of tomato fruit development and ripening genes orthologs identified in *Capsicum* genome (Dubey et al. 2019).

Tomato gene	***Capsicum* ortholog**	**E-value**	**Score (bits)**	**Tomato gene**	***Capsicum* ortholog**	**E-value**	**Score (bits)**
1. *ASR1*	*LOC107867643*	5.34E-174	616	16. *SlDML2*	*LOC107843639*	0	3494
2. *PSY1*	*LOC107868281*	1.59E-121	442		*LOC107843640*	8.51E-62	246
3. *TCTR1*	*LOC107843641*	0	937		*LOC107843637*	0	782
	LOC107843193	2.30E-64	254	17. *AP2*	*LOC107857848*	2.71E-74	285
4. *TomloxC*	*LOC107874197*	0	760	18. *Lutescent 2*	*LOC107854549*	0	839
	LOC107874182	1.50E-158	566	19 *Hydroquinone glucosyltransferase*	*LOC107839289*	0	2593
	LOC107844216	2.16E-82	313	20. *SlscADH1*	*LOC107849564*	0	1496
5. *MADS-RIN*	*LOC107847473*	4.92E-91	340	21. *SGR1*	*LOC107866321*	1.79E-179	634
	LOC107843064	4.95E-86	324	22. *AOX1a*	*LOC107870439*	0	904
6. *PG2*	*LOC107843830*	4.82E-132	477	23. *ERF6*	*LOC107879350*	0	1035
7. *TAGL1*	*LOC107878477*	6.30E-125	453	24. *CGS*	*LOC107858737*	9.62E-104	383
8. *NAC1*	*LOC107870355*	0	854		*LOC107849703*	1.97E-140	505
9. *ETR4*	*LOC107873245*	0	3219	25. *ARF4*	*LOC107847819*	0	1378
10. *LeSPL-CNR*	*LOC107859362*	3.18E-150	536	26. *NSGT1*	*LOC107864612*	0	1197
11. *CHS1*	*LOC107871256*	0	1877	27. spdsyn	*LOC107847831*	2.49E-95	355

	LOC107872666	0	1013	28. *Alpha-mannosidase*	*LOC107875980*	5.53E-109	401
12. *FUL1*	*LOC107855404*	8.69E-92	189	29. *GAME1*	*LOC107878054*	0	1903
13. *FUL2*	*LOC107845304*	8.59E-46	342	30. *XTH5*	*LOC107843860*	0	955
14. *GDH1*	*LOC107843990*	2.87E-130	472	31. *MADS-box protein 1*	*LOC107845303*	2.08E-69	268
	LOC107852706	4.91E-113	414	32. *Beta-hex*	*LOC107843457*	0	1751
15. *GLK2*	*LOC107845460*	2.05E-148	531				

3.1.2. *In silico* expression analysis of fruit development/ripening genes in *Capsicum* spp.

After identifying the genes, we wanted to know how these genes express in different fruit development stages of contrasting *Capsicum* genotypes. For this, we used available transcriptome data generated in our lab from three different developmental stages of each *Capsicum* genotypes fruits (early fruit (EF), breaker fruit (BF), mature fruit (MF), and fully opened flower) belonging to - *C. annuum* (PRJNA505972), *C. chinense* (PRJNA327797), and *C. frutescens* (PRJNA327800). Of the 42 genes analyzed for their expression in transcriptome data, 38 genes were found expressing in different stages of fruit developmental stages and 12 genes showed significant differential expressions. In other words, these genes were either up regulated or down regulated at certain stages compared to other tissue of the plant **(Fig. 3.3)**. These 12 genes showing differential expression were: *LOC107847473* (*MADS-RIN*), *LOC107878477* (*TAGL1*), *LOC107873245* (*ETR4*), *LOC107859362* (*Lespl-CNR*), *LOC107845304* (*FUL2*), *LOC107855404* (*FUL1*), *LOC107845460* (*GLK2*), *LOC107839289* (*Hydroquinone glucosyltransferase*), *LOC107866321* (*SGR1*), *LOC107864612* (*NSGT1*), *LOC107843860* (*XTH5*), and *LOC107845303* (*MADS-protein1*).

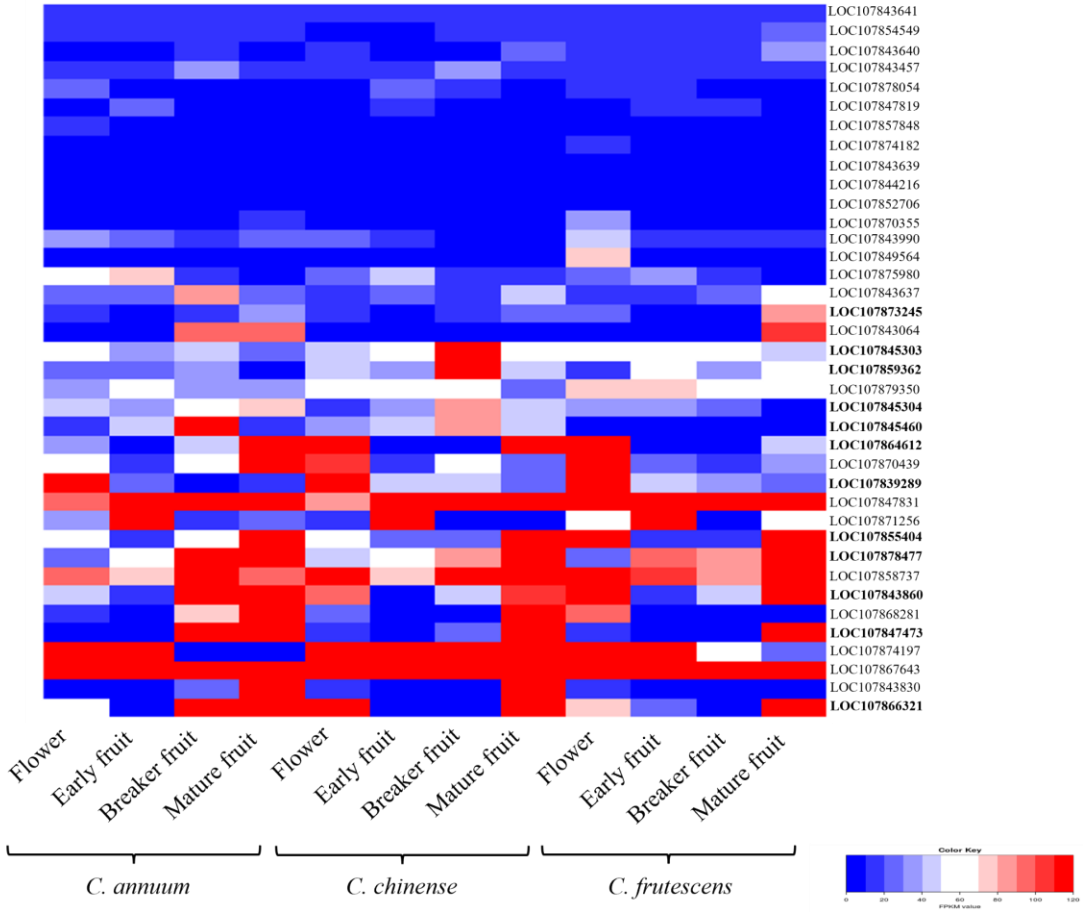


Fig. 3.3 Expression patterns of *Capsicum* orthologs of tomato fruit development/ripening genes as observed in *in-silico* analysis of transcriptome data. Twelve genes given in bold were differentially expressed during fruit development and validated using qRT-PCR analysis (Dubey et al., 2019).

3.1.3 Validation of differentially expressed fruit development/ripening genes using quantitative real time (qRT) PCR in contrasting *Capsicum* species

The 12 selected genes showing differential expression were further validated using quantitative real time PCR (qRT-PCR) in three contrasting genotypes belonging to *C. chinense*, *C. frutescens*, and *C. annuum*. Altogether five tissues (leaf, flower, early fruit, breaker fruit, and

mature fruit) were used for qRT-PCR analysis. Primers used for qRT-PCR analysis are given in **Table 3.3.** To see the relative expression of selected genes among different genotypes belonging to different *Capsicum* spp., leaf tissue of *C. chinense* was used as control. Interestingly, Transcriptome analysis revealed that most genes were differentially expressed across floral and fruit developmental stages. A largely consistent expression profile was shared among the three *Capsicum* species, with only minor exceptions; however, interspecific variation in the intensity of gene expression was evident **(Fig. 3.4)**. On comparison of the same tissues in different *Capsicum* species, it has been observed that some genes showed drastic differences in expression level. For example, *FUL2* and *Golden 2 like gene (GLK2)* had the highest expression in *C. annuum* as compared to *C. chinense* and *C. frutescens*. Similarly, *NSGT1* and *MADS-protein1* had the highest expression in *C. chinense* and *C. frutescens*, respectively.

Table 3.3. List of genes and their primer sequences used for qRT-PCR (Dubey et al. 2019).

***Capsicum* gene**	**Forwar primer seq**	**Reverese primer seq**
LOC107847473	CTCTAGTCGTGGCAAGCTTTAT	CTGATGGTTGGGTTCCTTCA
LOC107878477	AGAGGTAGAGAGAGCACAAGAG	GTTGTTGCGAGCATCTTCATAAT
LOC107873245	AATTGGGATGCAGCGTGTCT	GCAGGGACAAGAACACCAAGA
LOC107859362	GGGAAGGGAAGAGAACCATAAA	TGATGACCCTCCACCAGATA
LOC107845304	GCAGCTTAATGCAACTGATGTC	CCGCATAATGCCTTTGGTTTC
LOC107855404	GATGCTGAGGTTGGTTTGATTG	CCTCTCAGCATACGAGTATCTTTC
LOC107845460	GAAGGGAAGAAGAAGGTGAAGG	CCATCAGTAGCCATAAGCTCTAAA
LOC107839289	TGGCGTGGCTGGATGAA	ACCACTCCCGAACGAGATGTAT
LOC107866321	AACTCACTTTGGCAGTCTCTC	CTTTCTTCCATTCTGCAACCAC
LOC107864612	AGTGGAAGAACTTGGCATAGG	CCTCCACCACCTTCCTTATTC
LOC107843860	GGCATCGTTCATGACACCTA	GGAGATGGGACCGAAATGATAA
LOC107845303	TACCACGAGTATCTGAGGCTAAA	CTCCAGGTCCTTTGTGCTTAAT

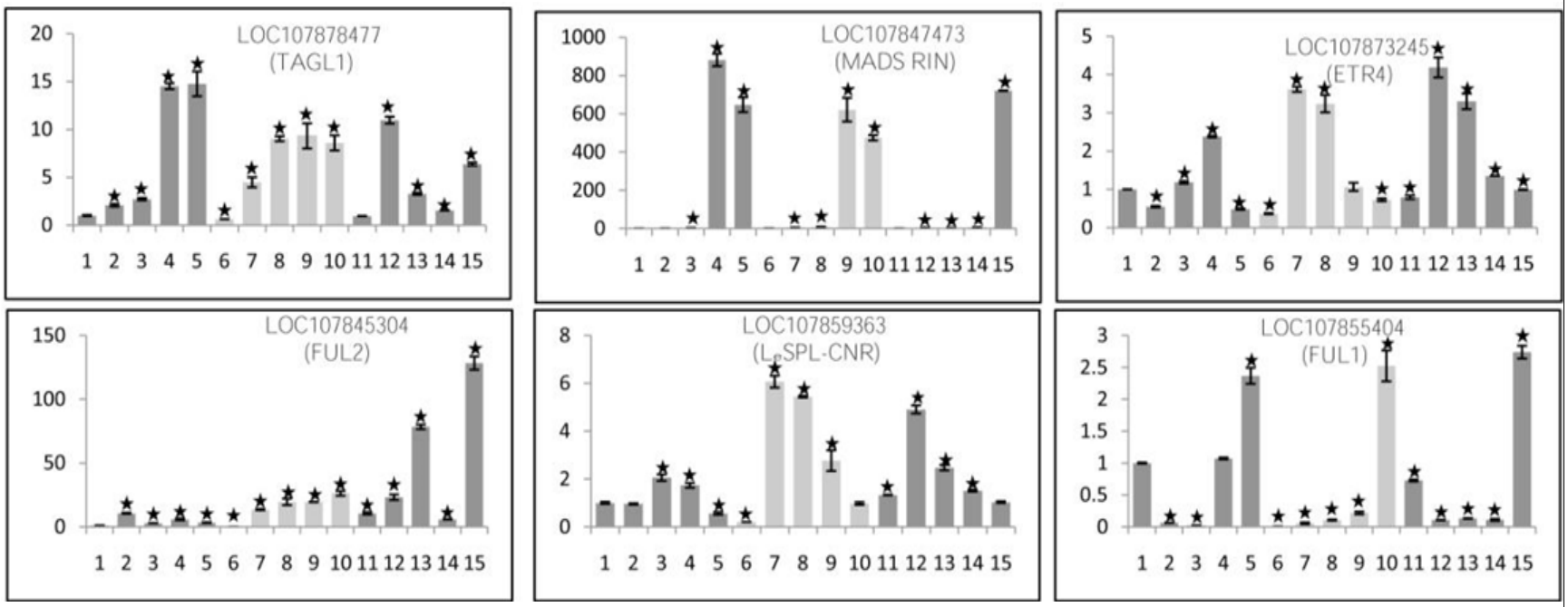


Fig. 3.4 (a) Real time validation of twelve fruit development/ripening genes using qRT-PCR. Y-axis shows relative expression (as compared to NB5 leaf) in terms of fold change of 15 tissues involving three spp. 1 = NB5_leaf, 2 = NB5_flower, 3 = NB5_early fruit, 4 = NB5_breaker fruit, 5 = NB5_mature fruit, 6= MCM_leaf, 7 = MCM_flower, 8 = MCM_early fruit, 9 = MCM_breaker fruit, 10 = MCM_mature fruit, 11= JH23_leaf, 12= JH23_flower, 13 = JH23_early fruit, 14 = JH23_breaker fruit, 15 = JH23_mature fruit; * represents 0.05 level of significance (Dubey et al., 2019)

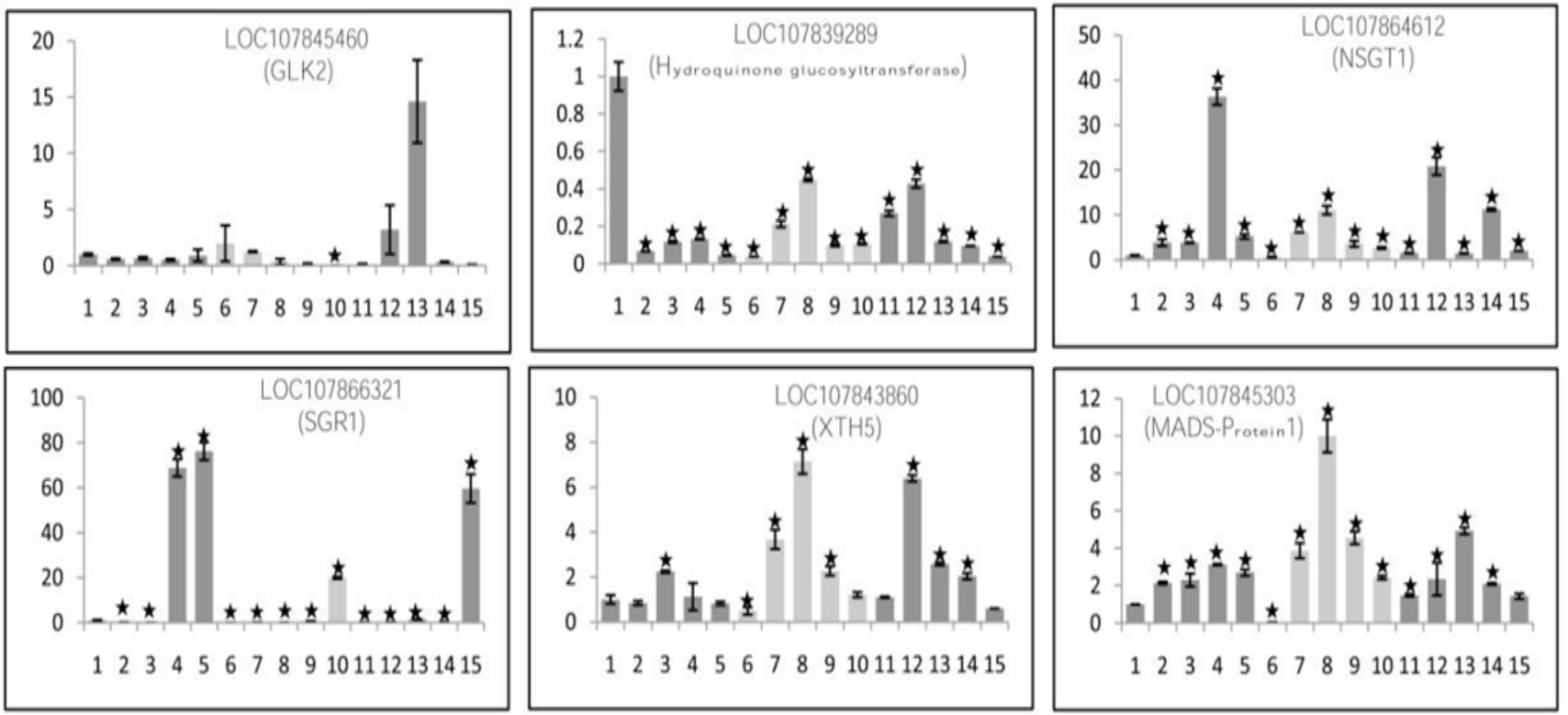


Fig. 3.4 (b) Real time validation of twelve fruit development/ripening genes using qRT-PCR. Y-axis shows relative expression (as compared to NB5 leaf) in terms of fold change of 15 tissues involving three spp. 1 = NB5_leaf, 2 = NB5_flower, 3 = NB5_early fruit, 4 = NB5_breaker fruit, 5 = NB5_mature fruit, 6= MCM_leaf, 7 = MCM_flower, 8 = MCM_early fruit, 9 = MCM_breaker fruit, 10 = MCM_mature fruit, 11= JH23_leaf, 12= JH23_flower, 13 = JH23_early fruit, 14 = JH23_breaker fruit, 15 = JH23_mature fruit; * represents 0.05 level of significance (Dubey et al. 2019).

Table 3.4 Relative expression of selected genes in terms of fold change as compared to leaf of corresponding genotype (Dubey et al.2019).

Capsicum gene ID	Tomato gene	NB5 (*C. chinense*)					MCM (*C. frutescence*)					JH23 (*C. annuum*)				
		A	B	C	D	E	A	B	C	D	E	A	B	C	D	E
LOC107878477	*TAGL1*	1	2.17*	2.71*	14.52*	14.76*	1	7.14*	14.41*	14.41*	13.7*	1	11.52*	3.44*	1.63*	6.71*
LOC107847473	*MADS- RIN*	1	1.44	3.27*	881.45*	647.67*	1	5.73*	9.09*	9.09*	476.23*	1	4.42*	5.2*	6.91*	707.54*
LOC107873245	*ETR4*	1	0.55*	1.18*	2.38*	0.48*	1	9.91*	8.87*	8.87*	1.99*	1	5.27*	4.16*	1.70*	1.25
LOC107845304	*FUL2*	1	10.6*	2.86*	5.82*	3.3*	1	423.74*	610.82*	610.82*	821.49*	1	2.19*	7.41*	0.56*	12.13*
LOC107859362	*LeSPL-CNR*	1	0.96	2.05*	1.73*	0.56*	1	30.1*	26.98*	26.98*	4.84*	1	3.69*	1.86*	1.13	0.78*
LOC107855404	*FUL1*	1	0.06*	0.02*	1.07	2.37*	1	9.29*	17.66*	17.66*	420.11*	1	0.15*	0.18*	0.15	3.77*
LOC107845460	*GLK2*	1	0.59	0.67	0.51	0.91	1	0.63	0.18	0.18	0.002*	1	32.81*	149.68*	3.45	0.48
LOC107839289	*Hydroquinone glucosyltransferase*	1	0.07*	0.12*	0.13*	0.05*	1	5.55*	11.71*	11.71*	2.77*	1	1.59*	0.44*	0.35*	0.14*
LOC107864612	*NSGT1*	1	3.84*	3.84*	36.31*	5.21*	1	9.8*	17.63*	17.63*	4.23*	1	13.52*	0.94	7.27*	1.32
LOC107866321	*SGR1*	1	0.15*	0.02*	68.82*	76.25*	1	171.58*	72.69*	72.69*	20160.1*	1	1.26	6.12*	1.72	255.47*
LOC107843860	*XTH5*	1	0.85	2.24*	1.13	0.83	1	7.48*	14.53*	14.53*	2.47*	1	5.79*	2.34*	1.85*	0.54*
LOC107845303	*MADS-Protein1*	1	2.14*	2.29*	3.13*	2.69*	1	86.34*	223.09*	223.09*	54.26*	1	1.58*	3.32*	1.41	0.97

A = Leaf, B = flower, C = Early fruit, D = Breaker fruit, E = Mature fruit.
* represents 0.05 level of significance.

A significant correlation was observed between gene expression profiles derived from transcriptome data and validation by quantitative real time PCR (qRT-PCR) across three *Capsicum* species *(C. annuum, C. chinense, and C. frutescens).* This correlation validates the reliability of the transcriptome data and highlights key conserved regulatory patterns in fruit development and ripening. For instance, *MADS-RIN* exhibited an exceptionally high, consistent upregulation of over 500-fold in breaker and mature fruit stages across all species in both datasets. Similarly, *SGR1* showed consistent high expression in all fruit tissues, with the highest levels in *C. frutescens*. The expression dynamics of several genes associated with early fruit development and cell wall modification were also recapitulated. Genes including *ETR4, LeSPL, XTH5,* and *MADS-protein1* displayed higher expression in flower, early, breaker with a subsequent decline upon maturity in the qRT-PCR data. This pattern was largely mirrored in the transcriptome data, with minor species-specific variations, such as a notable upregulation of *ETR4* in mature *C. frutescens* fruit.

However, the comparative analysis also revealed specific discrepancies and species-specific divergences for certain genes. While transcriptome data indicated that *TAGL1* was expressed broadly across all developmental stages, qRT-PCR analysis provided precise quantification, revealing that its expression peaked pronouncedly in specific tissues. This tissue-specificity varied by species; for example, *TAGL1* expression peaked in the fruit of *C. frutescens* (14.76-fold higher) but was highest in the flower of *C. annuum* (11.5-fold higher). Similarly, the transcriptome and qRT-PCR data for *FUL1* and *FUL2* agreed on high expression but differed slightly in their precise tissue-specificity among

genotypes. The expression of *Golden 2-like (GLK2)* and *hydroquinone glucosyltransferase* showed opposing regulatory trends during fruit development and ripening between the qRT-PCR and transcriptome data within the same species. For example, qRT-PCR indicated *GLK2* was up-regulated in *C. annuum* but seen less expression in *C. chinense* and *C. frutescens* during development, whereas the transcriptome data suggested downregulation in all three. The expression pattern of *NSGT1* also found differences in transcriptome data and qRT-PCR analysis, indicating potential technical or analytical complexities for these specific transcripts.

In conclusion, the high degree of correlation for many genes confirms the overall robustness of the transcriptome findings. These findings collectively provide a validated framework for understanding the genetic regulation of fruit development ripening in *Capsicum* spp.

3.1.4 Development of gene(s) based Simple sequence repeat (SSR) markers

For utilization of fruit development/ripening genes in *Capsicum* breeding, we developed user-friendly PCR based SSR markers which were either present within the gene or in close vicinity (within 5 Kb regions). Altogether, 49 SSRs were developed, of which, 14 SSRs were from the gene sequence, and 35 SSRs were present in close vicinity (5 Kb upstream and downstream) to the genes **(Table 3.3)**. Out of 14 SSRs (present within gene sequence), one SSR (SSR_CF-14) was present in exon, however, remaining 13 SSRs were intronic SSRs. Furthermore, we analyzed the 1.5 kb promoter region upstream of the transcription start site (TSS) and identified 12 SSRs out of a total of 35 SSRs. Using the

PlantPAN database, we also discovered 11 transcription factors (TF) binding motifs within these promoters and predicted their putative functions **(Table 3.5; Fig. 3.6 and Fig. 3.7).**

Out of 49 SSRs, 38, 10 and 1 were di-, tri- and tetra nucleotides, respectively **(Fig. 3.6)**. For di-nucleotide SSRs five types of motifs i.e GA/TC, CA/TG, AT/TA, AC/GT, and AG/CT were identified. Similarly, for tri- and tetra nucleotide SSRs seven and one kind of motifs were identified, respectively **(Fig. 3.6)**. Among all the above mentioned 13 (5-di nucleotide, 7-tri nucleotide and 1- tetra nucleotide) types of motifs, motif AT/TA had maximum frequency (31 SSRs). Number of motifs repeats are 54 and total motif length 108 were found of 41 SSRs (SSR_CF-47). Above mentioned 14 gene based SSRs were from nine genes i.e. *LOC107843641* (SSR_CF-1), *LOC107874197* (SSR_CF-2), *LOC107843830* (SSR_CF-3), *LOC107845304* (SSR_CF-4, SSR_CF-5), *LOC107855404* (SSR_CF-6, SSR_CF-7, SSR_CF-8, SSR_CF-9), *LOC107843990* (SSR_CF-10), *LOC107845460* (SSR_CF-11), *LOC107854549* (SSR_CF-12, SSR_CF-13), and *LOC107875980* (SSR_CF-14). Remaining 35 SSRs were present within 5 Kb up and downstream regions of 23 genes. The primer sequences, melting temperature, amplicon size (bp) and other details of 49 SSRs are provided in **Table 3.3.** A representative gel image of SSR profiling in *Capsicum* accessions is given in **Table 3. 5; Fig. 3.5, Fig. 3.6, and Fig. 3.7.**

Table 3.5. Summary of SSRs developed from fruit development/ripening genes in *Capsicum* (Dubey et al. 2019)

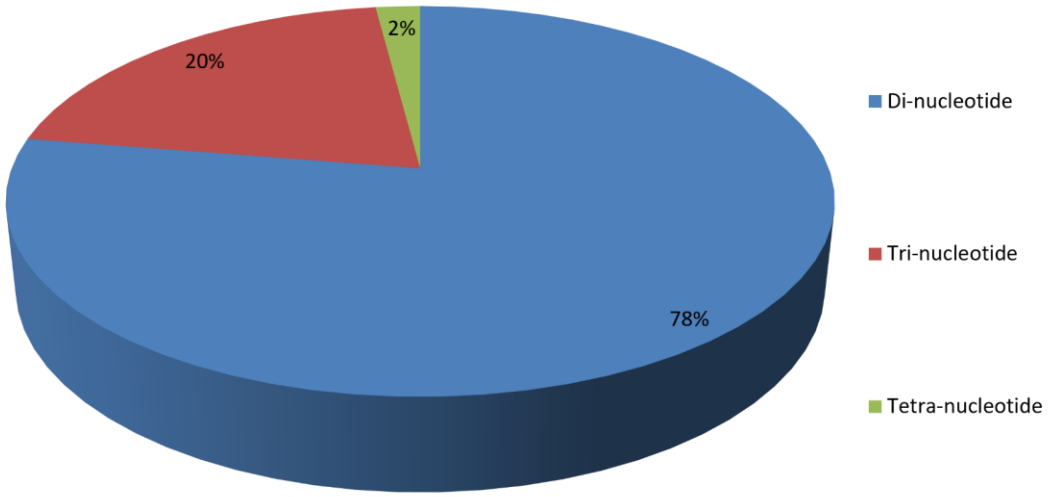
SSR[@]	*Capsicum* gene[$]	Motif	Forward primer	Length	TM	Reverse primer	Length	TM	Amplicon size
SSR_CF-1*	*LOC107843641*	(TA)14	CCAACTTATGTGATCCTCTTCT	22	55	GGTCAATGCACAAGGTAGTT	20	55	385
SSR_CF-2*	*LOC107874197*	(TCT)6	AAGACACTAAGAGAAAGGATGC	22	55	AGTACCCGAGGTTATGGAAT	20	55	384
SSR_CF-3*	*LOC107843830*	(TA)9	GGGGAACTATCAATGGAAAT	20	55	CGTATCGAGTAAGTCAGAGTGAA	23	56	394
SSR_CF-4*	*LOC107845304*	(CAA)6	AGTCTCTTCTTTGTGAGTGTTG	22	54	GACAGCTAACTAGACAGGTTTGT	23	55	383
SSR_CF-5*	*LOC107845304*	(TA)14	CGCACAAACTTCACAACTAA	20	55	CTTTCATACCTCTCAAGAATCC	22	55	354
SSR_CF-6*	*LOC107855404*	(TA)6	AACCGTAGATGGAAGTCCTAC	21	55	TAAGTGTGCTGCTCCTCTTC	20	56	389
SSR_CF-7*	*LOC107855404*	(AAC)6	GTAGTCTCCATCTCCATACCTG	22	55	CGGGTGTAATCAACTCTCTTA	21	55	134
SSR_CF-8*	*LOC107855404*	(AT)6	CGTGAGATTAACATCGTATTCC	22	56	AGTCCTCGTGTTTGTGTGTT	20	56	202
SSR_CF-9*	*LOC107855404*	(AT)11	ATAGTGAACACACAAACACGAG	22	55	GACGTAGCAGGAGTTTTACTTT	22	55	283
SSR_CF-10*	*LOC107843990*	(GT)6	ATAATCTTCAGTATCAGGCTCG	22	55	AACAACCTCTGGATGGTATCT	21	55	253
SSR_CF-11*	*LOC107845460*	(AT)6	TTGAAAGGAGGGAGTATCTT	20	53	CCATCATACATTTGCTTCC	19	54	297
SSR_CF-12*	*LOC107854549*	(AAC)6	GTGATACACCCTTATATGACCC	22	55	GTCGTACTTCTTGTCGTAGGTAT	23	54	270
SSR_CF-13*	*LOC107854549*	(TATC)6	GATTCCCAGAAGTGTGAAAA	20	55	CTACTGTCATTGTTTGGTTGAC	22	55	362
SSR_CF-14*	*LOC107875980*	(GA)6	AGTTCAAGTGAGCGAACATAC	21	55	ACAGTTTAGTGGTAGAACCCTT	22	54	235
SSR_CF-15*	*LOC107843641-up*	(TA)6	TCCTGGTTTGATTATATGGC	20	55	GGTAGTTCGTGGATTACTTTTC	22	55	398
SSR_CF-16*	*LOC107847473-up*	(TA)6	TTCCTACACATGCTCATCACT	21	56	TAAGAATGGAGAAGTGGCAA	20	56	359
SSR_CF-17*	*LOC107878477-up*	(AT)12	CTTTCCTTCAATACCACTTACC	22	55	TTTACTTCGCCATTGCTACT	20	55	316
SSR_CF-18	*LOC107845304-up*	(TA)10	AAAGATAGAGACTTCAGTTGCC	22	55	GGGTTATTTCGGTGTTTTG	19	55	390

SSR_CF-19*	*LOC107845304 -up*	(ATT)8	TCAACTCTTACGACTGACATGA	22	56	AACACCACCATTAAAGACCTC	21	56	326
SSR_CF-20	*LOC107845304 -up*	(AT)6	CCCCTTCGATTAGTTTGTATAG	22	55	GGTTACTGTTGGAGTCGTTAGT	22	55	270
SSR_CF-21*	*LOC107843990 -up*	(AT)10	TAGTGGCCTATTCATGTTG	19	52	GCACGTCTTAATTCTTATCC	20	52	184
SSR_CF-22*	*LOC107845460 -up*	(AGA)6	GATACTTTACTGGATGGTTGCT	22	55	TGTTCTACACTCGTATTTGGG	21	55	269
SSR_CF-23*	*LOC107854549 -up*	(AT)7	TACTATAACAGCAATTACCGCC	22	56	CCGCTTCGATTGTATATGAA	20	56	187
SSR_CF-24*	*LOC107839289 -up*	(TA)12	AACAGTATAAGAACGTGGTGTG	22	54	GACATCGCAGTCAGTAATAAAC	22	54	322
SSR_CF-25*	*LOC107866321 -up*	(TA)7	TAGAAGGTTCATCCAAACTCTC	22	55	GAAGTAGTTGCGTATTGGGTAT	22	55	375
SSR_CF-26	*LOC107866321 -up*	(CA)6	TAACCTTCACTAACACCTCACA	22	55	GGCAAAGAGAATGAGTAGAAAC	22	55	278
SSR_CF-27*	*LOC107879350 -up*	(TA)7	GAGATATTTATGGGGTAAGTCG	22	55	AAGGTGTGTTGTAGGGGTTA	20	55	395
SSR_CF-28*	*LOC107858737 -up*	(TA)10	TACTTGACTGCTTGATTCCTAC	22	54	AAGCCTAACCAAAGCTAAAG	20	53	310
SSR_CF-29	*LOC107858737 -up*	(AG)8	CTAAACTAGCACTTATCCCGAC	22	55	CCTCTTTTGTTACCTCTTTGAC	22	55	250
SSR_CF-30*	*LOC107847819 -up*	(TAC)8	AGGGGTAGTAGTGGAAATTGTT	22	56	ACAGGTGAAGTAGAGGAAGATG	22	55	116
SSR_CF-31*	*LOC107847831 -up*	(AT)9	TGTAACTTGTAACCCAACACG	21	56	GATCTCAAGCTCTTCTTTCTTG	22	55	389
SSR_CF-32*	*LOC107845303 -up*	(TA)9	GTTTGCATGTGAGTTATGTAGG	22	55	AGAATCACTGGGCTATTCAAC	21	56	371
SSR_CF-33	*LOC107843457 -up*	(AAT)6	CTACTTCCGGTTGAAGATTGT	21	56	CAGGATGTTTATCTGTTGCAC	21	56	393
SSR_CF-34*	*LOC107843457 -up*	(AT)21	AAATACCCTCAAATCCTGTG	20	54	ATTCAACAATGGAGTCAACC	20	55	188
SSR_CF-35	*LOC107843457 -up*	(AT)6	CACACGACATAGTCATAGGAAG	22	55	TATATTGAGGGGTCATTTGG	20	55	382
SSR_CF-36*	*LOC107874197 -down*	(AT)6	GAACTTTAGTCCCCAAGCAT	20	56	GTGAACCTCAAACTCTACCATT	22	55	399

SSR_CF-37*	*LOC107847473 -down*	(TA)13	CAACAAGTGTCTGGCAATAAC	21	56	CACGAGTGAAGTACGTGTAGC	21	56	305
SSR_CF-38*	*LOC107870355 -down*	(AT)10	AGAGATCGACGGCAACTTA	19	56	TGACGTGGCGTATGAAAT	18	56	391
SSR_CF-39*	*LOC107873245 -down*	(AT)9	GCACCCGATAATGTAAGAA	19	54	TTTGCATCTCTCATAGCACT	20	54	246
SSR_CF-40*	*LOC107871256 -down*	(TA)9	AAGCCCGTACTAGATTGTTAAG	22	55	ATACTGAAGAAGGATACAACCG	22	55	355
SSR_CF-41*	*LOC107845460 -down*	(AT)14	GATTTTCCCTGTGAGTGGTA	20	55	TGTCCTATTCTCTGTGTGTTTG	22	55	282
SSR_CF-42*	*LOC107843639 -down*	(TA)6	TGGAAGCATCTATTGGAGAA	20	56	GTCACATAAGTTGCGATAAAGC	22	57	383
SSR_CF-43*	*LOC107866321 -down*	(AT)9	ATCTATGGAGTCATTTGGTGAG	22	56	GTTCTTGGGTCATACTTCTTTG	22	56	380
SSR_CF-44	*LOC107866321 -down*	(TC)7	AGTAACCATGTGTGCTGACTAA	22	55	GTGTTTGAGTAGGATTGGAGAT	22	55	282
SSR_CF-45*	*LOC107879350 -down*	(ATC)8	AACAACATATCAGCCTCTGC	20	55	AGGAGTTAGAACAAGAATGCTC	22	55	358
SSR_CF-46*	*LOC107847831 -down*	(CT)6	GGCGATTGCTACTAATAACTCT	22	55	CAAGCTATGTATCGTCAGTCAG	22	55	256
SSR_CF-47*	*LOC107875980 -down*	(AC)54	CGTCAAGTCTACGAGTAAGGA	21	55	CCTCCTGTTTGGTTGTAAGTAG	22	56	315
SSR_CF-48*	*LOC107878054 -down*	(AGA)8	TGGAGAGTTTAGTAGTTTCGTG	22	54	GAGTATGAAGATGAGCGTTAGA	22	54	396
SSR_CF-49*	*LOC107843860 -down*	(TA)14	TGTTAAGAGAGCAGTGTGGTTA	22	55	ACTACCATTACCTTCTCCGAAT	22	56	349

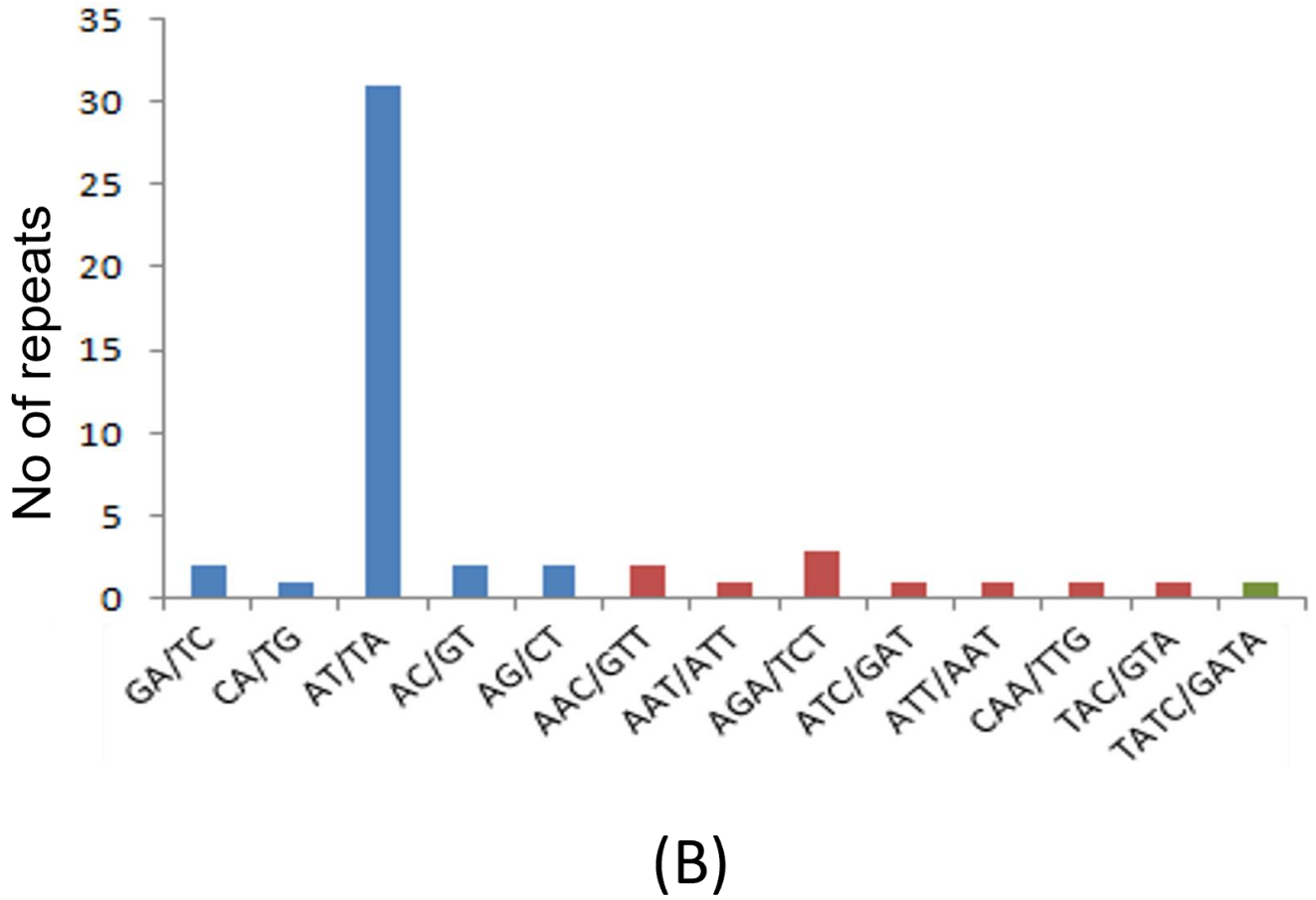
[@] Primer pairs for marked SSRs with star sign were custom synthesized for genotyping; SSRs in bold represented polymorphism

[$] up and down represent sequences 5 Kb upstream from the start codon and 5 Kb downstream from the stop codon of corresponding gene.



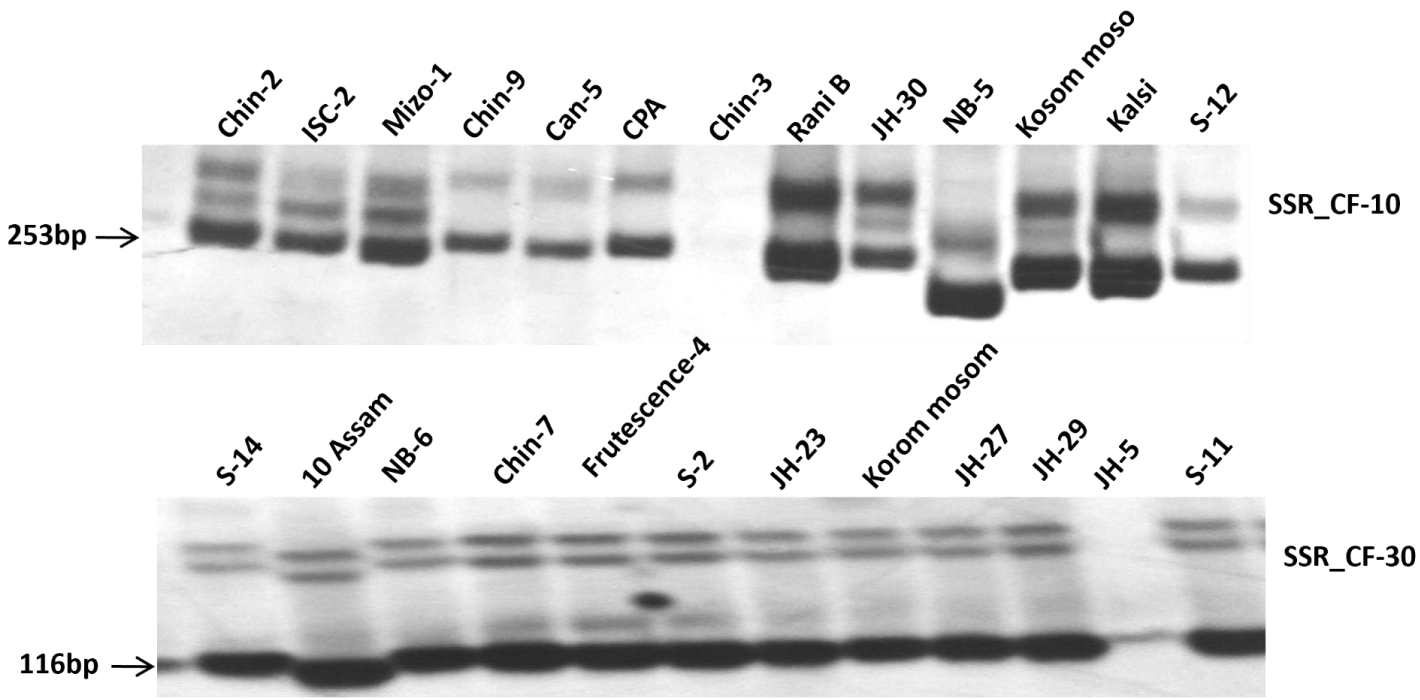
(A)

Fig. 3.5(A) Proportions of different SSRs motifs identified in fruit development and ripening related genes in *Capsicum* (Dubey et al. 2019).



(B)

Fig. 3.5 (B) Distribution of SSR motif frequencies identified in fruit development and ripening related genes in *Capsicum* genome. Motifs written as pairs (e.g., AG/CT) are grouped as complementary sequences (Dubey et al. 2019).



(C)

Fig. 3.5(C) Representative gel image showing polymorphic (SSR CF-30) and non-polymorphic (SSR CF-10) bands while genotyping *Capsicum* germplasm using SSR markers. Amplicon size of each of the markers (in bp) are indicated on the left side (Dubey et al. 2019).

3.1.5 Plant materials used for SSR genotyping and diversity analysis

In this study, a total of 47 *Capsicum* genotypes, belonging to *C. chinense, C. frutescens*, and *C. frutescens,* were selected for SSR genotyping experiments. These contrasting genotypes were collected mainly from the Northeast India (**Table 2.1 and Fig. 3. 6 for accession details)**. The selection was designed to represent the broad genetic and phenotypic diversity found within and between these species.

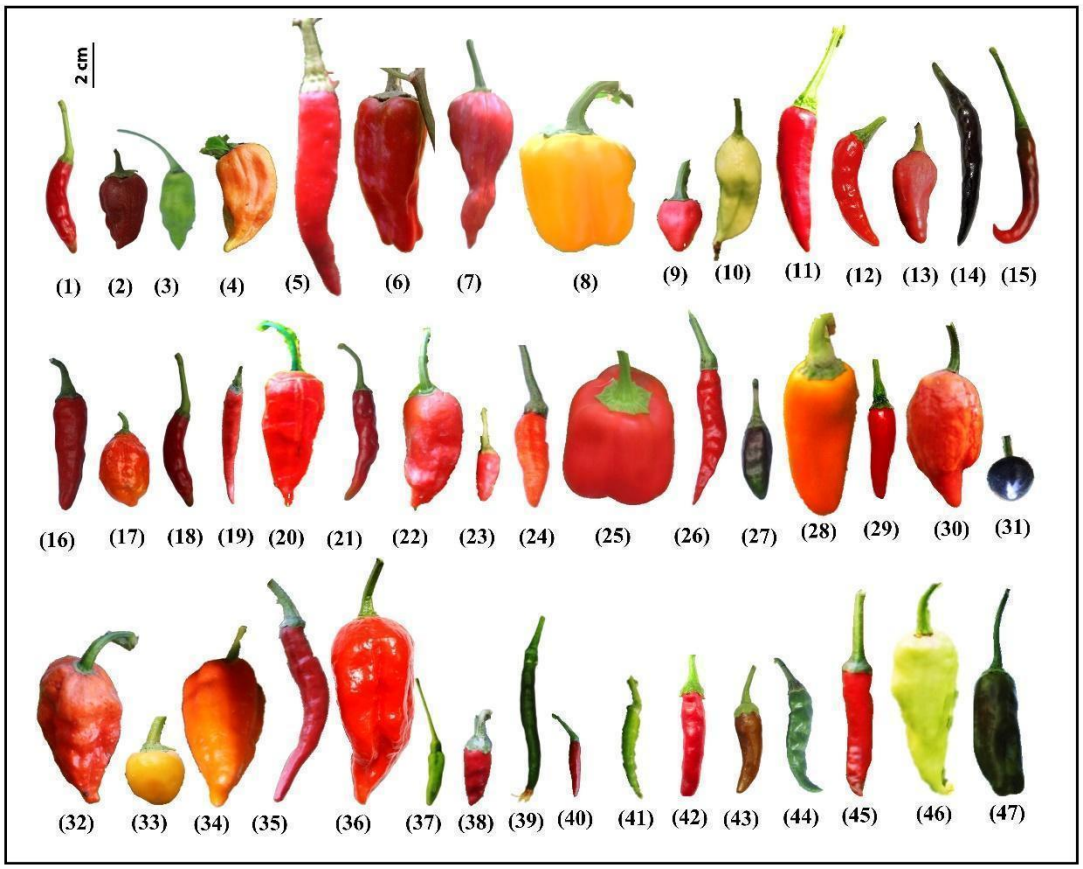


Fig. 3. 6 Pictures of 47 Genotypes of *Capsicum* belonging to three *Capsicum* species- ***C. annuum*** (Ca-1, Ca-2, Ca-5, Ca-6, Ca-8, Ca-11, Ca-12, Ca-14, Ca-15, Ca-16, Ca-18, Ca-19, Ca-21, Ca-25, Ca-26, Ca-27, Ca-28, Ca-29, Ca-33, Ca-35, Ca-39, Ca-41, Ca-42, Ca-43, Ca-44, Ca-45, & Ca-47), ***C. chinense*** (Cc-3, Cc-4, Cc-7, Cc-10, Cc-13, Cc-17, Cc-20, Cc-22, Cc-24, Cc-30, Cc-32, Cc-34, Cc-36, & Cc-46), ***C. frutescens*** (Cf-9, Cf-23, Cf-37, & Cf- 40), and **Other *Capsicum* species** (Co-31 & Co- 38) used in the study.

3.1.6 Genotyping of SSRs and diversity analysis

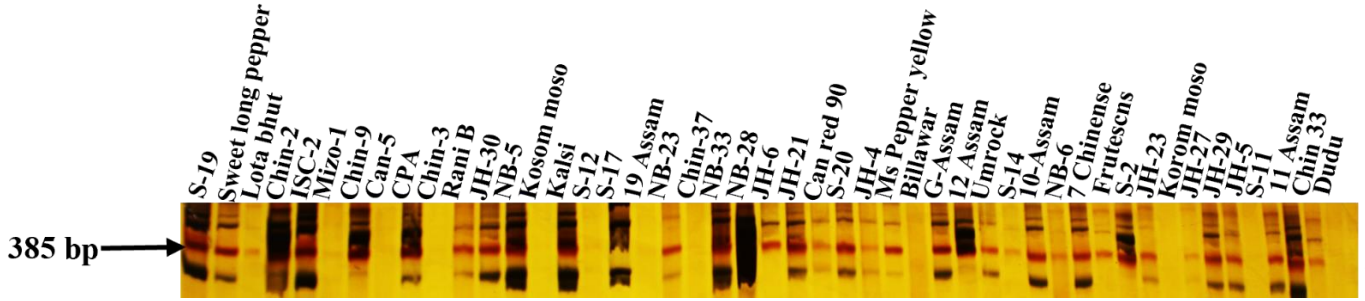
For SSRs genotyping, we have selected 42 SSRs out of 49 SSRs (including all the 14 genic SSRs and 28 non-genic SSRs, **(Table 3.5 and Fig. 3.7)**. Out of 42 SSRs, 23 SSRs were found to be polymorphic including six genic SSRs. Genetic diversity parameters such as effective number of alleles (ne), Shannon index (I), expected homo- and heterozygosity, Nei gene diversity and polymorphic information content (PIC) for each of the 23 polymorphic SSRs are summarized in **Table 3.6.** A total of 50 alleles with 2-3 alleles per locus were scored. Shannon diversity index (I) ranged from 0.51 (SSR_CF-30) – 0.97 (SSR_CF-5) with an average 0.64. Average Nei's gene diversity was 0.42 (ranged 0.26 to 0.58). Maximum PIC was identified for SSR_CF-5 (0.51) followed by SSR_CF-3 (0.44), however, minimum PIC was observed for SSR_CF-30 (0.22).

Table 3.6. Diversity parameters observed in 47 *Capsicum* accessions using 23 polymorphic SSR markers.

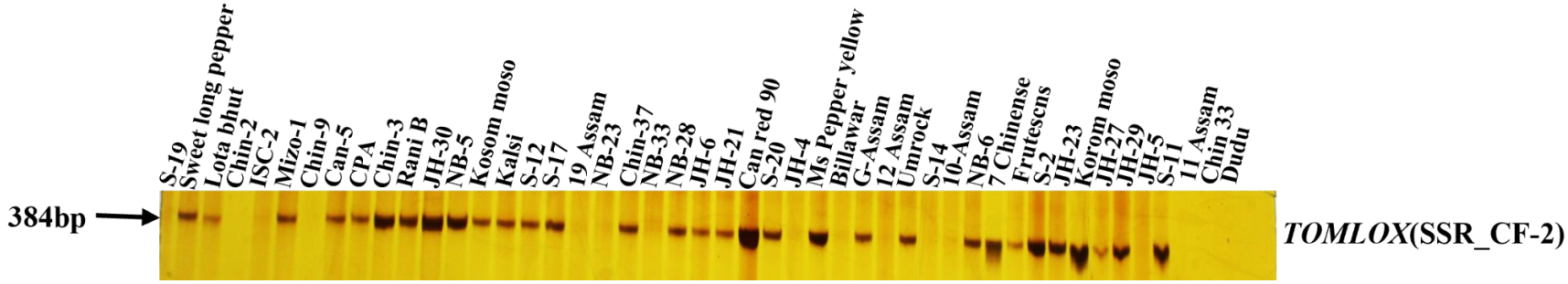
SSR[#]	No of observed allele	Effective number of alleles	Shannon index	Expected homozygosity	Expected heterozygosity	Nei gene diversity	PIC[@]
SSR_CF-1	2	1.873	0.659	0.529	0.471	0.466	0.357
SSR_CF-11	2	1.719	0.609	0.577	0.423	0.418	0.331
SSR_CF-15	2	1.555	0.542	0.639	0.361	0.357	0.293
SSR_CF-17	2	1.614	0.568	0.616	0.384	0.38	0.308
SSR_CF-19	2	1.858	0.654	0.533	0.467	0.462	0.355
SSR_CF-2	2	1.719	0.609	0.577	0.423	0.418	0.331
SSR_CF-21	3	1.68	0.712	0.591	0.409	0.405	0.359
SSR_CF-22	2	1.572	0.55	0.632	0.368	0.364	0.298
SSR_CF-25	2	1.719	0.609	0.577	0.423	0.418	0.331
SSR_CF-27	2	1.941	0.678	0.51	0.49	0.485	0.367
SSR_CF-3	3	2.036	0.852	0.486	0.514	0.509	0.44
SSR_CF-30	3	1.356	0.51	0.735	0.265	0.263	0.22
SSR_CF-31	2	1.873	0.659	0.529	0.471	0.466	0.357
SSR_CF-32	2	1.815	0.641	0.546	0.454	0.449	0.348
SSR_CF-37	2	1.957	0.682	0.506	0.494	0.489	0.369
SSR_CF-38	2	1.516	0.524	0.656	0.344	0.34	0.282
SSR_CF-4	2	1.614	0.568	0.616	0.384	0.38	0.308
SSR_CF-42	2	1.896	0.666	0.522	0.478	0.473	0.361
SSR_CF-43	2	1.504	0.518	0.661	0.339	0.335	0.279
SSR_CF-45	2	1.815	0.641	0.546	0.454	0.449	0.348
SSR_CF-46	2	1.978	0.688	0.5	0.5	0.494	0.372
SSR_CF-47	2	1.815	0.641	0.546	0.454	0.449	0.348
SSR_CF-5	3	2.399	0.972	0.411	0.589	0.583	0.511

[#] SSRs in bold represent genic SSRs

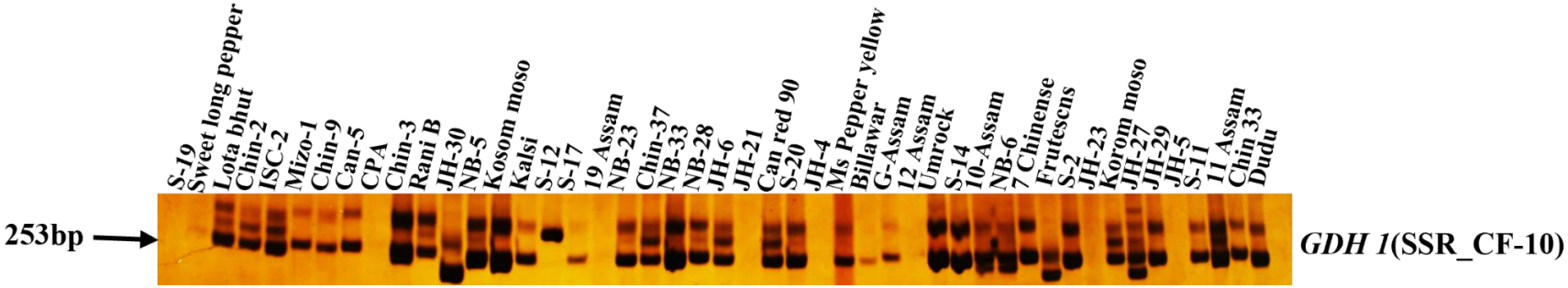
[@]PIC represents polymorphic information content.


S-19
Sweet long pepper
Lota bhut
Chin-2
ISC-2
Mizo-1
Chin-9
Can-5
CPA
Chin-3
Rani B
JH-30
NB-5
Kosom moso
Kalsi
S-12
S-17
19 Assam
NB-23
Chin-37
NB-33
NB-28
JH-6
JH-21
Can red 90
S-20
JH-4
Ms Pepper yellow
Billawar
G-Assam
12 Assam
Umrock
S-14
10-Assam
NB-6
7 Chinense
Frutescns
S-2
JH-23
Korom moso
JH-27
JH-29
JH-5
S-11
11 Assam
Chin 33
Dudu
385 bp
TCTR1(SSR_CF-1)

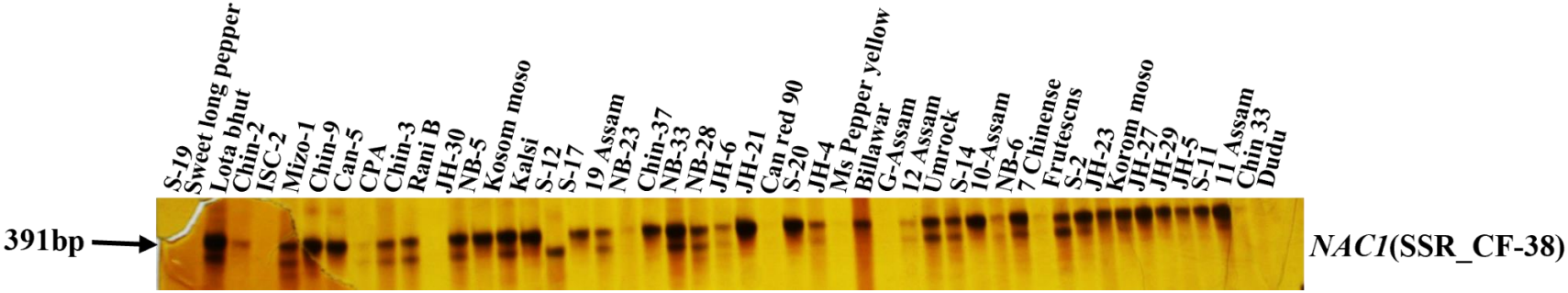
(A)


S-19
Sweet long pepper
Lota bhut
Chin-2
ISC-2
Mizo-1
Chin-9
Can-5
CPA
Chin-3
Rani B
JH-30
NB-5
Kosom moso
Kalsi
S-12
S-17
19 Assam
NB-23
Chin-37
NB-33
NB-28
JH-6
JH-21
Can red 90
S-20
JH-4
Ms Pepper yellow
Billawar
G-Assam
12 Assam
Umrock
S-14
10-Assam
NB-6
7 Chinense
Frutescns
S-2
JH-23
Korom moso
JH-27
JH-29
JH-5
S-11
11 Assam
Chin 33
Dudu
384bp
TOMLOX(SSR_CF-2)

(B)


S-19
Sweet long pepper
Lota bhut
Chin-2
ISC-2
Mizo-1
Chin-9
Can-5
CPA
Chin-3
Rani B
JH-30
NB-5
Kosom moso
Kalsi
S-12
S-17
19 Assam
NB-23
Chin-37
NB-33
NB-28
JH-6
JH-21
Can red 90
S-20
JH-4
Ms Pepper yellow
Billawar
G-Assam
12 Assam
Umrock
S-14
10-Assam
NB-6
7 Chinense
Frutescns
S-2
JH-23
Korom moso
JH-27
JH-29
JH-5
S-11
11 Assam
Chin 33
Dudu
253bp
GDH 1(SSR_CF-10)

(C)


S-19
Sweet long pepper
Lota bhut
Chin-2
ISC-2
Mizo-1
Chin-9
Can-5
CPA
Chin-3
Rani B
JH-30
NB-5
Kosom moso
Kalsi
S-12
S-17
19 Assam
NB-23
Chin-37
NB-33
NB-28
JH-6
JH-21
Can red 90
S-20
JH-4
Ms Pepper yellow
Billawar
G-Assam
12 Assam
Umrock
S-14
10-Assam
NB-6
7 Chinense
Frutescns
S-2
JH-23
Korom moso
JH-27
JH-29
JH-5
S-11
11 Assam
Chin 33
Dudu
391bp
NAC1(SSR_CF-38)

(D)

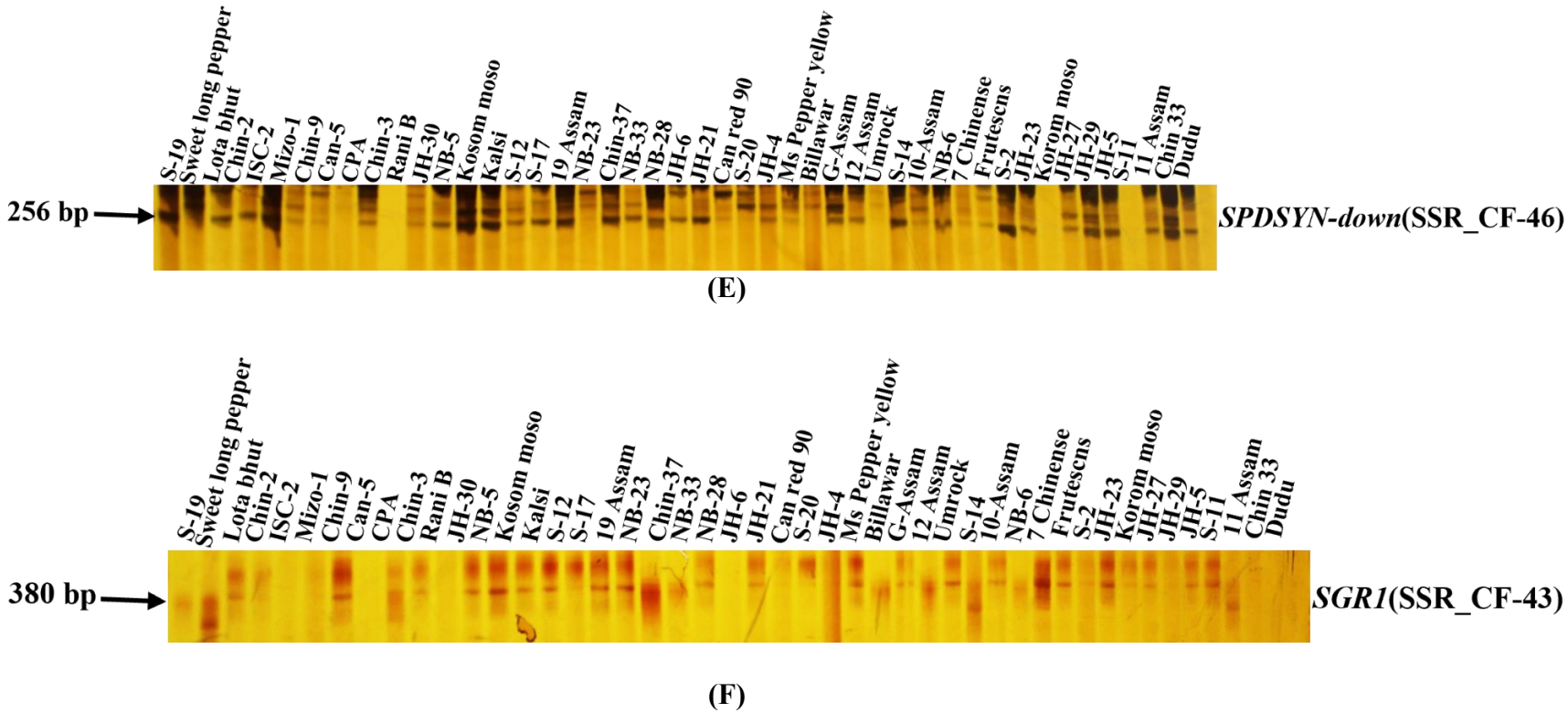


Fig. 3.7 Polymorphic gel images of *Capsicum* germplasm shown while genotyping with **(A)** SSR_ CF-1 (*TCTR1*), **(B)** SSR_ CF-2 (*TOMLOX*), **(C)**SSR_ CF-10 (*GDH1*), **(D)** SSR_ CF-38 (*NAC1*), **(E)** SSR_ CF-46 (*SPDSYN-down*), and **(F)** SSR_CF-43 (

3.2 OBJECTIVE 2: Determination of Vitamin C and Vitamin E content at different fruit developmental stages in contrasting *Capsicum* genotypes.

3.2.1. Vitamin C content in *Capsicum* species

Vitamin C content was found highly variable across the three fruit developmental stages i.e., early fruit (EF), breaker fruit (BF), and mature fruit (MF) in the 30 genotypes belonging to three *Capsicum* species **[Table 3.7; Fig. 3.8, and Fig. 3.9 (a), (b), (c)& (d)]**. At the EF stage, maximum Vitamin C content of 38.41 mg/g was observed in the genotype Cc-9 followed by Ca-3 (35.87 mg/g) and Cc-2 (35.26 mg/g) and the minimum Vitamin C content of 15.21 mg/g was found in Ca-4. In case of BF stage, maximum Vitamin C (34.38 mg/g) was found in Ca-8 and Cc-9 (31.82 mg/g) followed by Ca-5 (27.34 mg/g) and minimum Vitamin C content was estimated in Cf-10 (16.19 mg/g). In case of MF stage, the highest Vitamin C content was found (41.76 mg/g) in Ca-8 and Cc-6 (34.08 mg/g), followed by Cc-3 (32.04 mg/g) and minimum was found in Cf-8 (1 6.94 mg/g). Overall, the highest Vitamin C content was observed in Ca-8 (MF), Cc-9 (EF), Ca-8 (BF), and Ca-3 (EF); and the lowest was found in Ca-4 (EF), respectively and overall average highest VitC content was found in *C. chinense* (25.33 mg/g) followed by *C. annuum* (24.25 mg/g) and *C. frutescense* (22.09 mg/g).

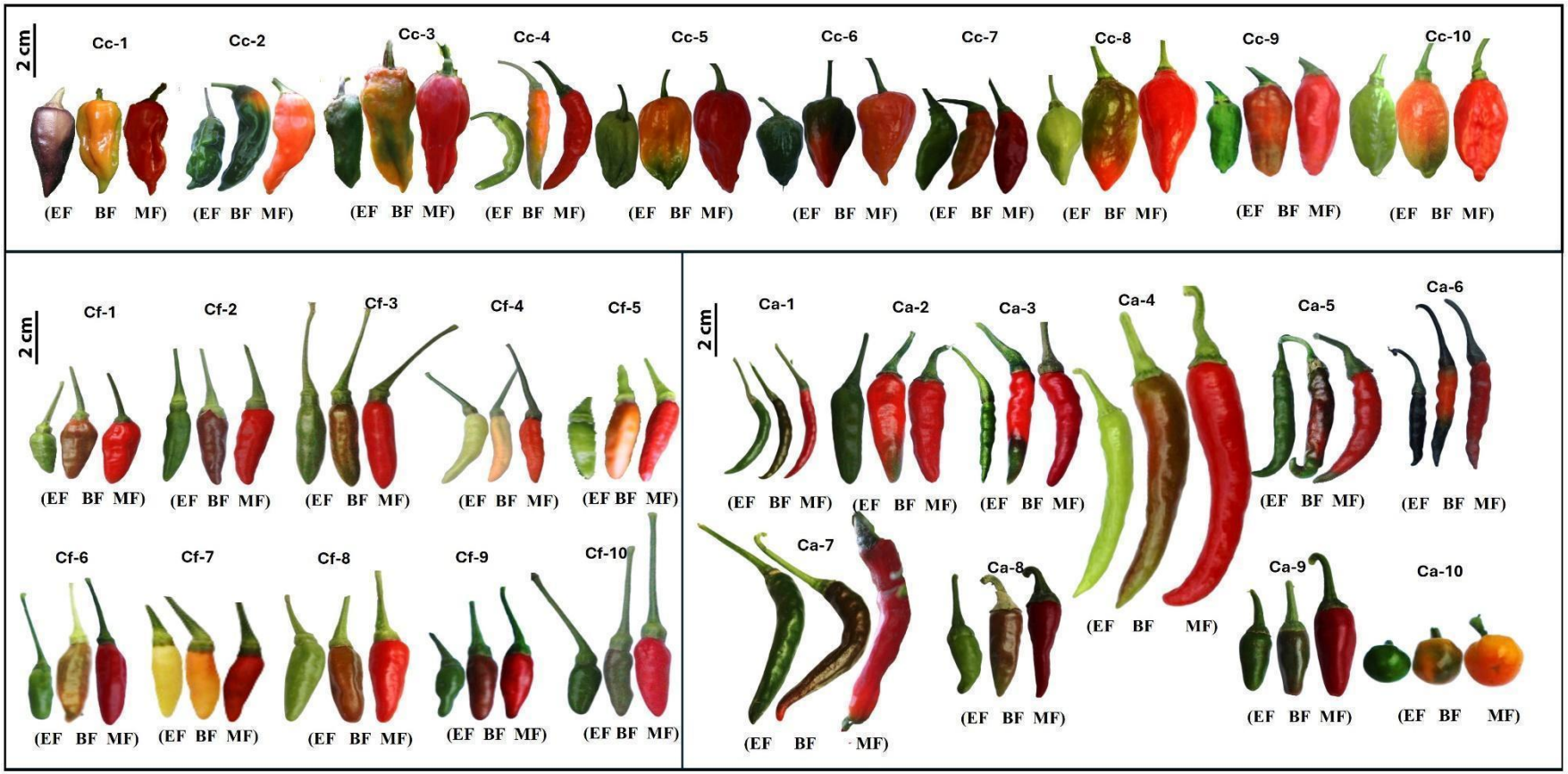


Fig. 3.8 Pictures of 30 *Capsicum* Genotypes belonging to 3 *Capsicum* species- *C. chinense* (Cc-1 to Cc-10), *C. frutescens* (Cf-1 to Cf-10), and *C. annuum* (Ca-1 to Ca-10) used for Vitamin C and Vitamin E estimation in three developmental stages-Early fruit (EF), Breaker fruit (BF), and Mature fruit (MF).

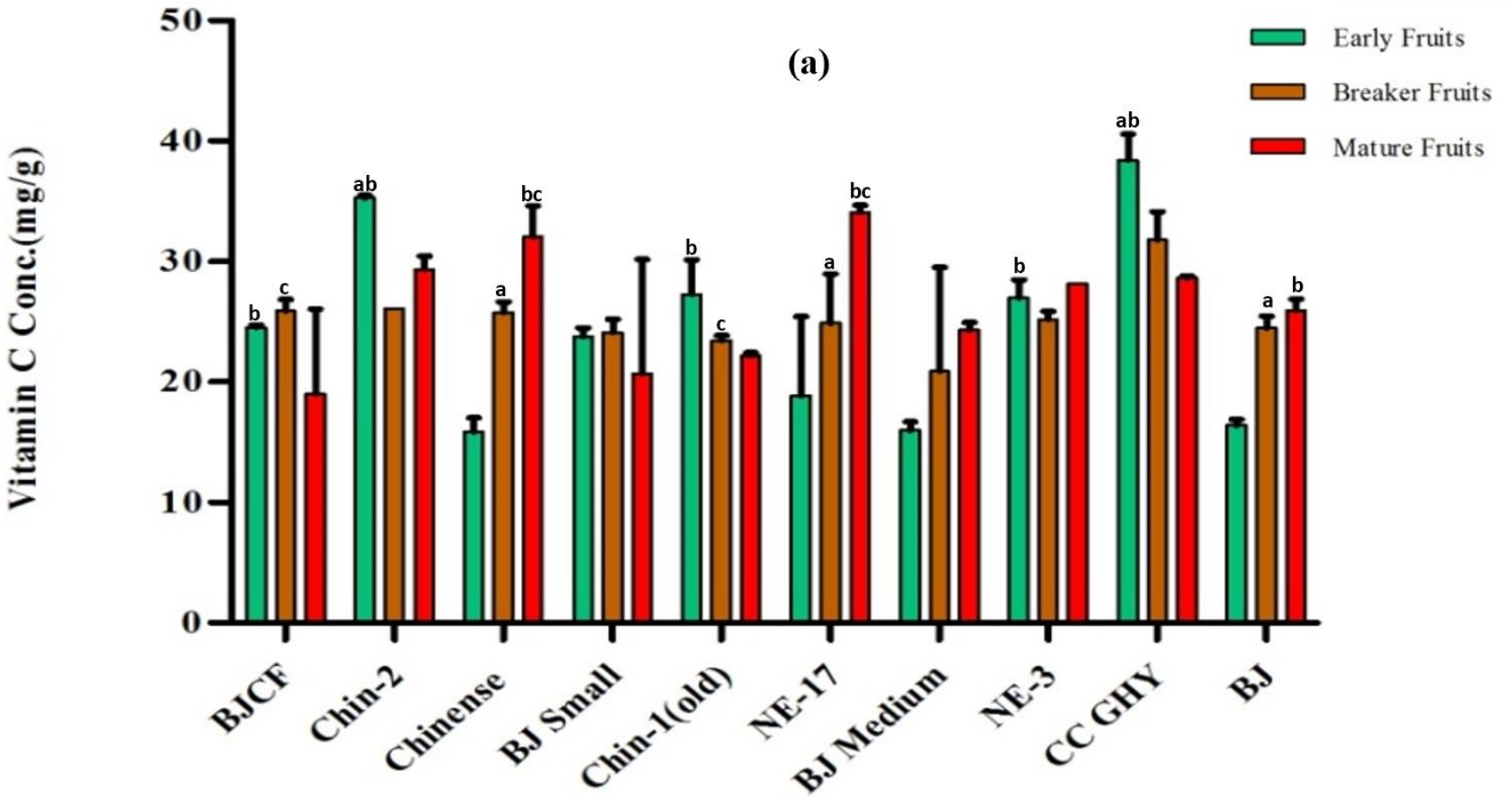


Fig. 3.9 (a) Vitamin C contents in early fruit (EF), breaker fruit (BF), and mature fruit (MF) in 10 *Capsicum* genotypes belonging to *C. chinense*. Letters 'a', 'b' and 'c' represent significance with respect to the fruit stage 'a-between early and breaker', 'b-between early and mature' and 'c-between breaker and mature' using 2-way ANOVA Test (Dubey et al., 2025).

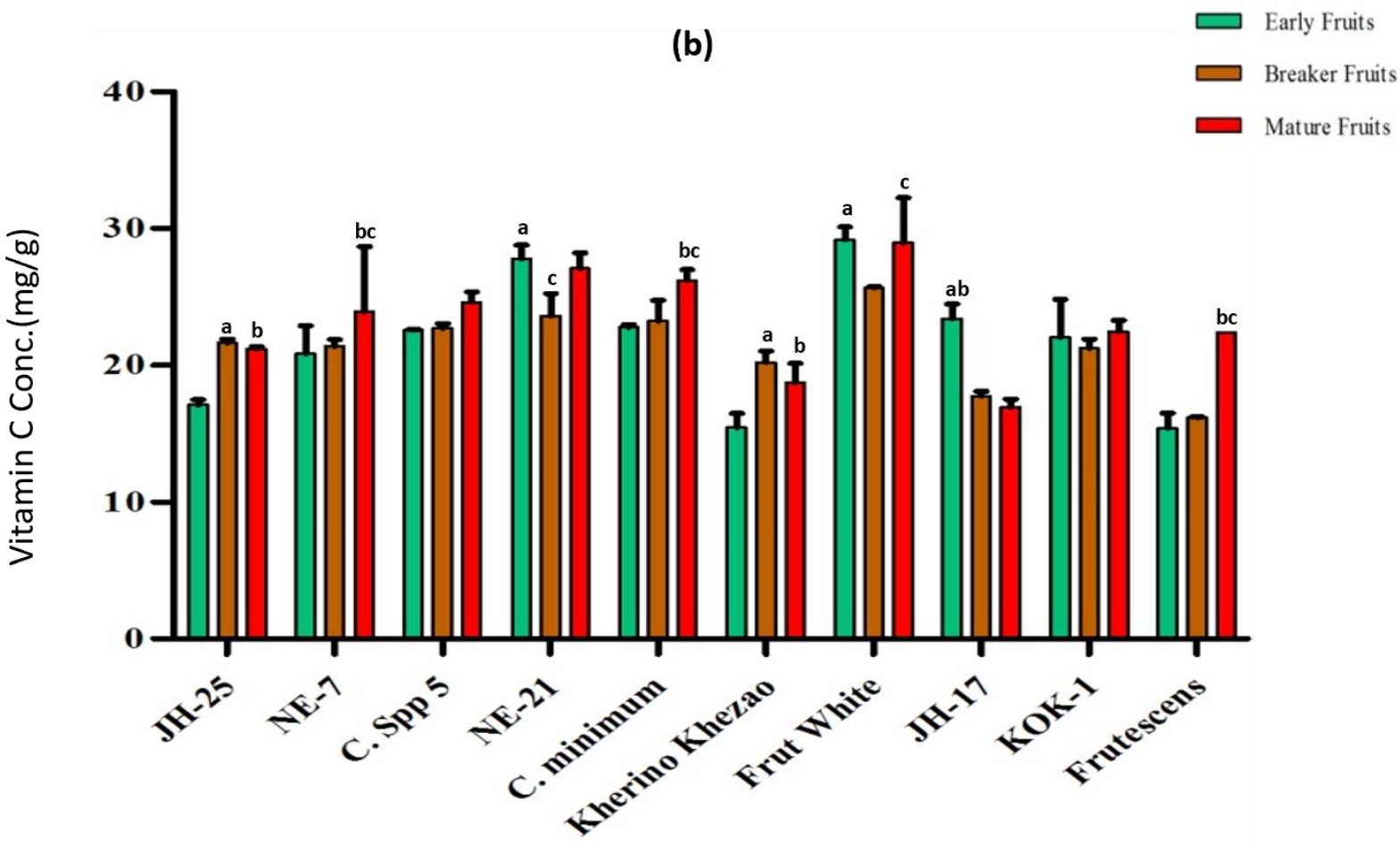


Fig. 3.9 (b) Vitamin C contents in early fruit (EF), breaker fruit (BF), and mature fruit (MF) in 10 *Capsicum* genotypes belonging to *C. frutescens*. Letters 'a', 'b' and 'c' represent significance with respect to the fruit stage 'a-between early and breaker', 'b-between early and mature' and 'c-between breaker and mature' using 2-way ANOVA Test (Dubey et al. 2025)

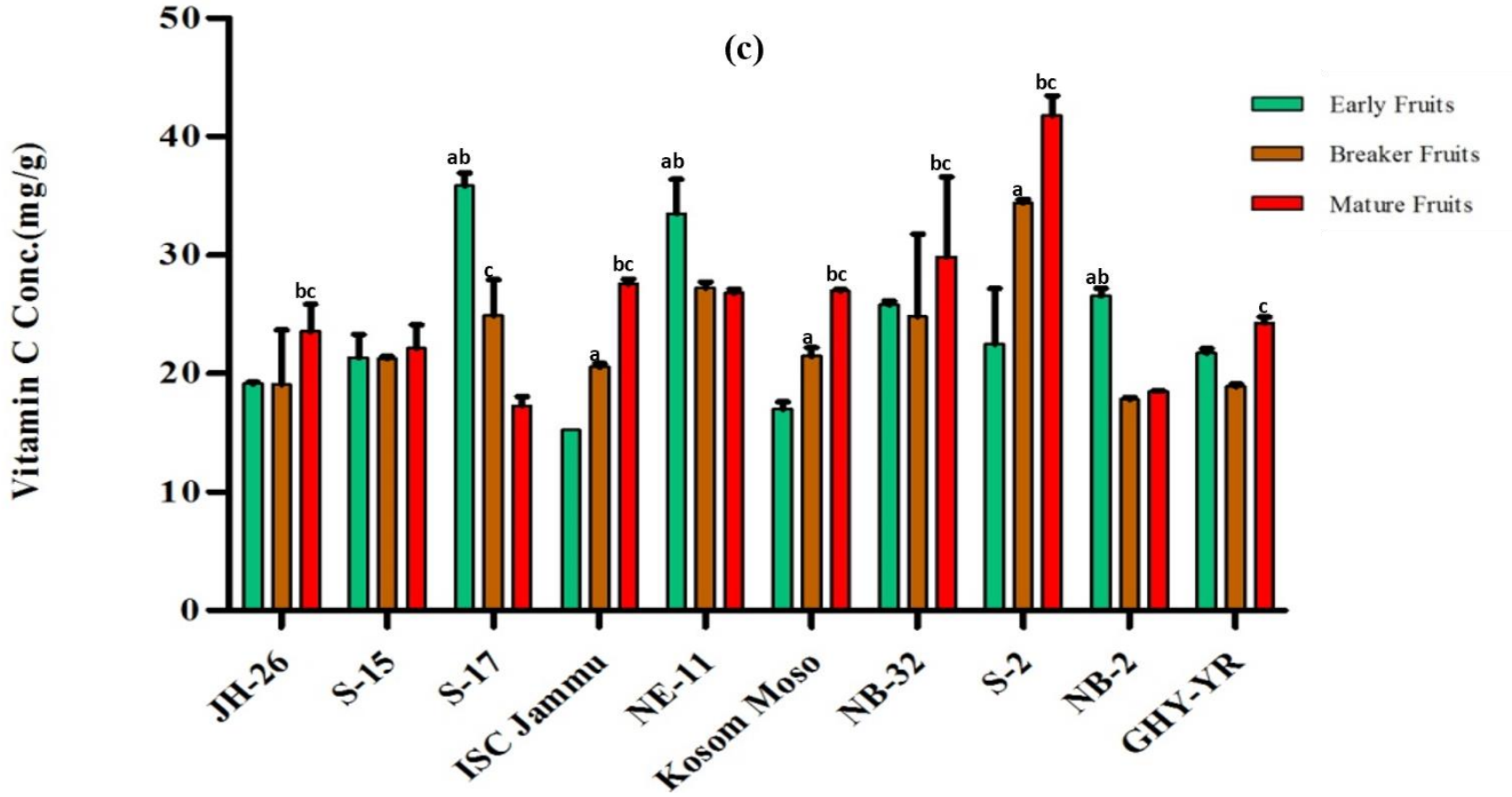


Fig. 3.9(c) Vitamin C contents in early fruit (EF), breaker fruit (BF), and mature fruit (MF) in 10 *Capsicum* genotypes belonging to *C. annuum*. Letters 'a', b' and 'c' represent significance with respect to the fruit stage 'a-between early and breaker', 'b-between early and mature' and 'c-between breaker and mature' using 2-way ANOVA Test (Dubey et al. 2025)

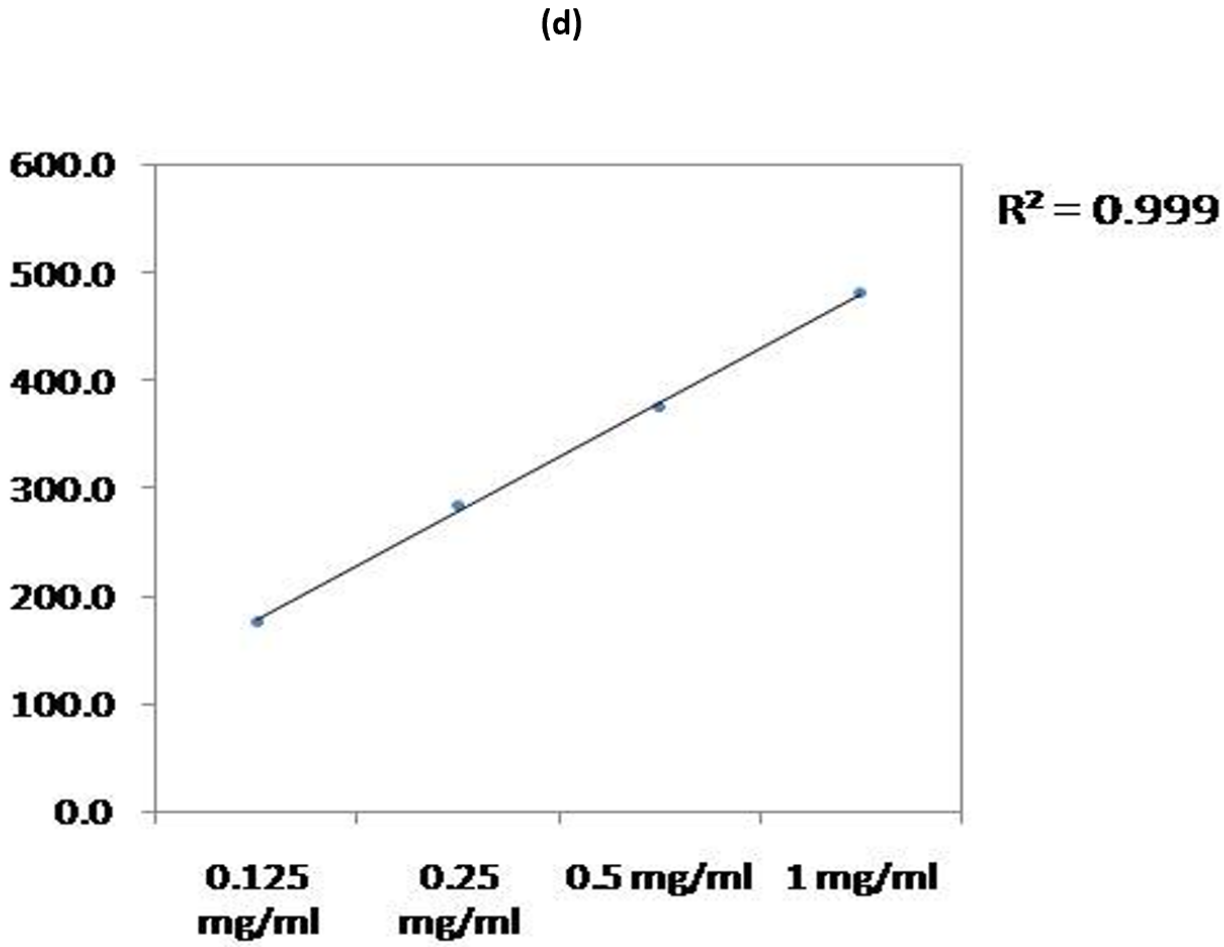


Fig. 3.9(d) Standard curve used for calculating and comparing Vitamin C content in 30 *Capsicum* germplasm through HPLC readings (Dubey et al. 2025)

3.2.2 Vitamin E content in *Capsicum* species

Similar to Vitamin C, Vitamin E content was also found highly variable across the tissues of 30 genotypes belonging to three species of *Capsicum* **[Table 3.7; Fig. 3.8, and Fig. 3.10 (a), (b), (c)& (d)]**. In the case of EF stage, maximum Vitamin E content was observed in Ca-5 (111.4 mg/g) followed by Ca-6 (111.12 mg/g), and Ca-10 (109.3 mg/g); and minimum content was estimated in Cf-10 (30.68 mg/g). In the BF stage, maximum Vitamin E content was found in Cc-9 (108.7 mg/g) and Cc-2 (108.4 mg/g) followed by Cf-7 (108.2 mg/g); and the minimum was found in Cc-4 (6.59 mg/g). In MF stage, the highest Vitamin E content was found in Cf-9 (112.4 mg/g) and Cc-6 (109.28 mg/g) followed by Ca-1 (108.2 mg/g), and the minimum Vitamin E was found in Cf-8 (3.32 mg/g). Overall, the highest Vitamin E content was observed in Cf-9 (MF) and the lowest was observed in Cf-8 (MF), respectively, and overall average highest VitE content was found in *C. annuum* (85.63 mg/g) followed by *C. frutescense* (85.22 mg/g.) and *C. chinense* (83.41 mg/g).

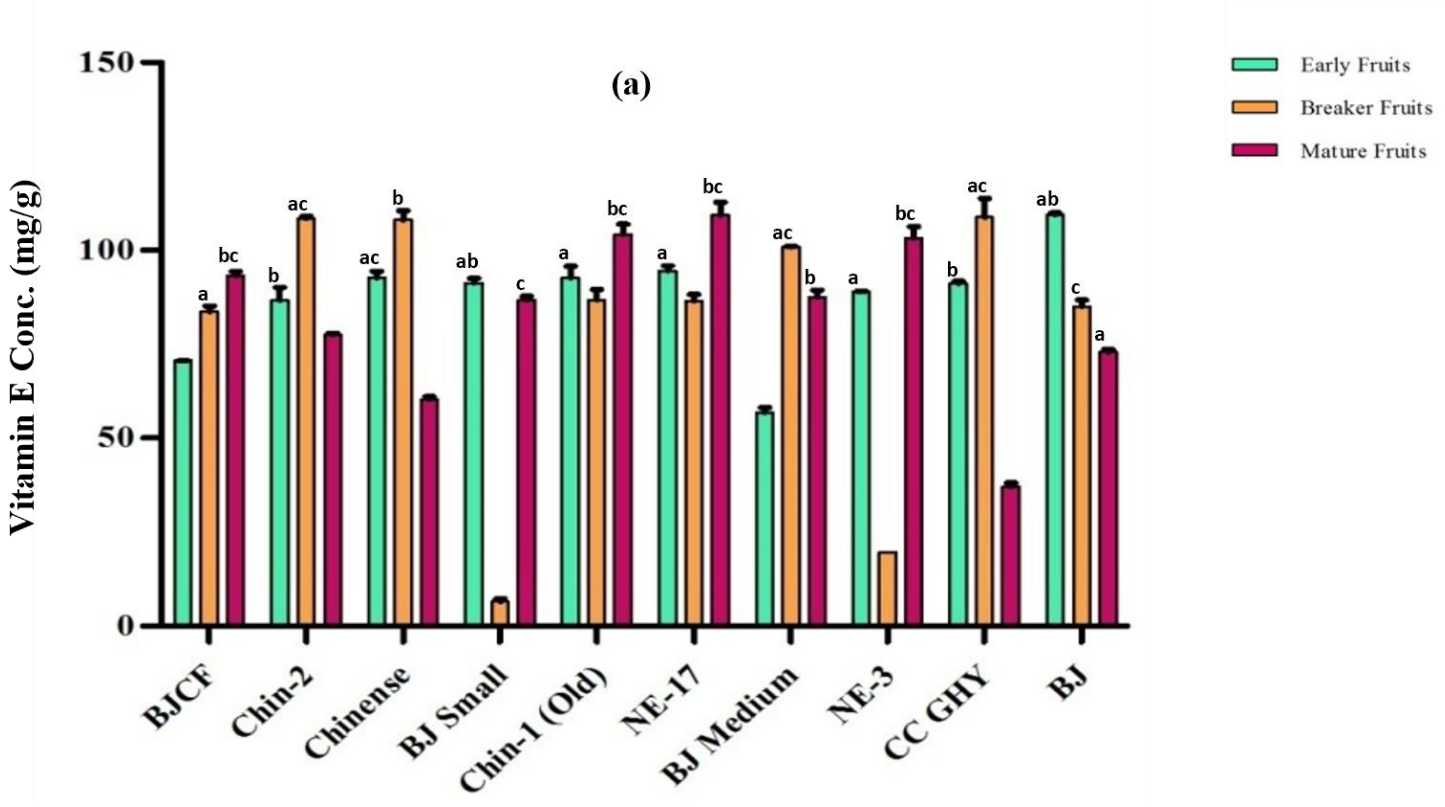


Fig. 3.10(a) Vitamin E contents in early fruit (EF), breaker fruit (BF), and mature fruit (MF) in 10 *Capsicum* genotypes belonging to *C. chinense*. Letters 'a', 'b' and 'c' represent significance with respect to the fruit stage 'a-between early and breaker', 'b-between early and mature' and 'c-between breaker and mature' using 2-way ANOVA Test (Dubey et al., 2025) (Dubey et al. 2025)

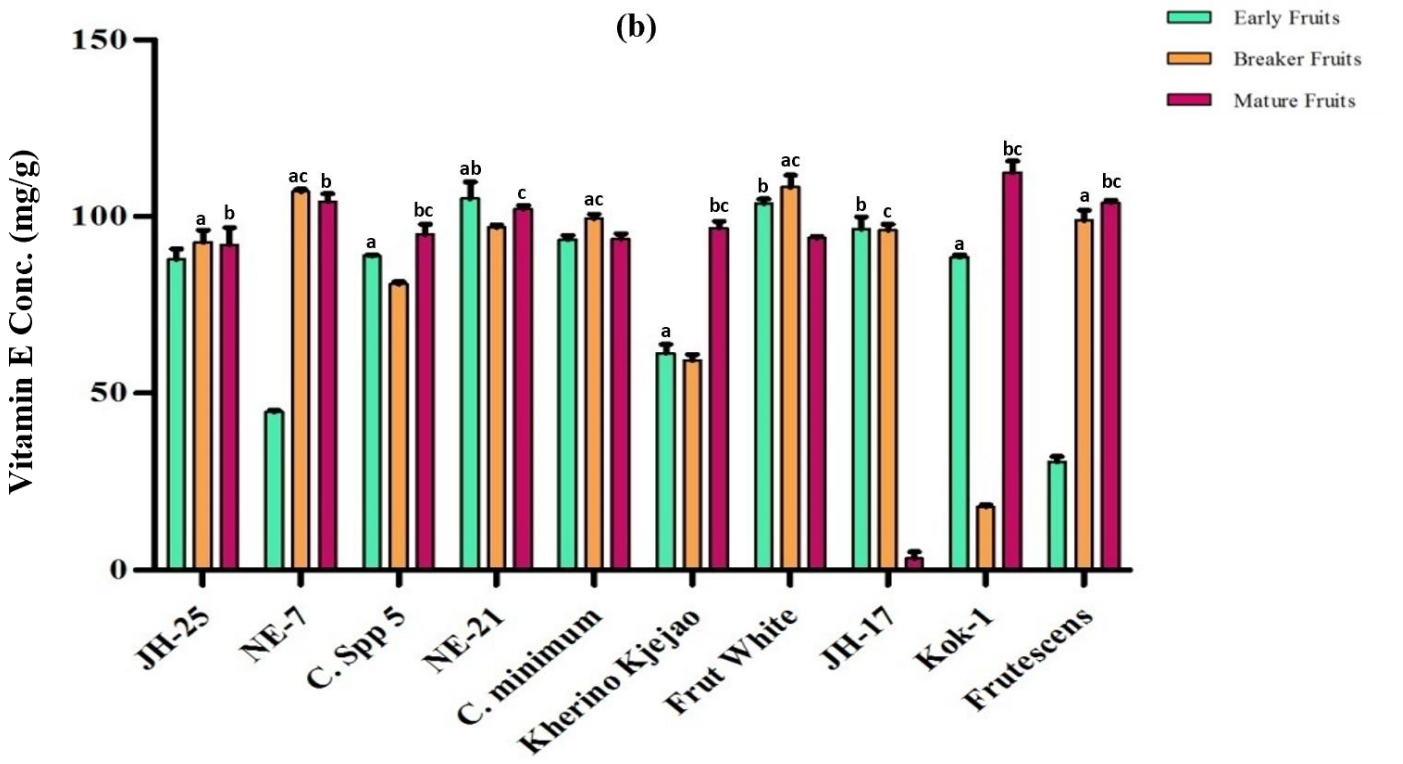


Fig. 3.10(b) Vitamin E contents in early fruit (EF), breaker fruit (BF), and mature fruit (MF) in 10 *Capsicum* genotypes belonging to *C. frutescens*. Letters 'a', b' and 'c' represent significance with respect to the fruit stage 'a-between early and breaker', 'b-between early and mature' and 'c-between breaker and mature' using 2-way ANOVA Test (Dubey et al. 2025)

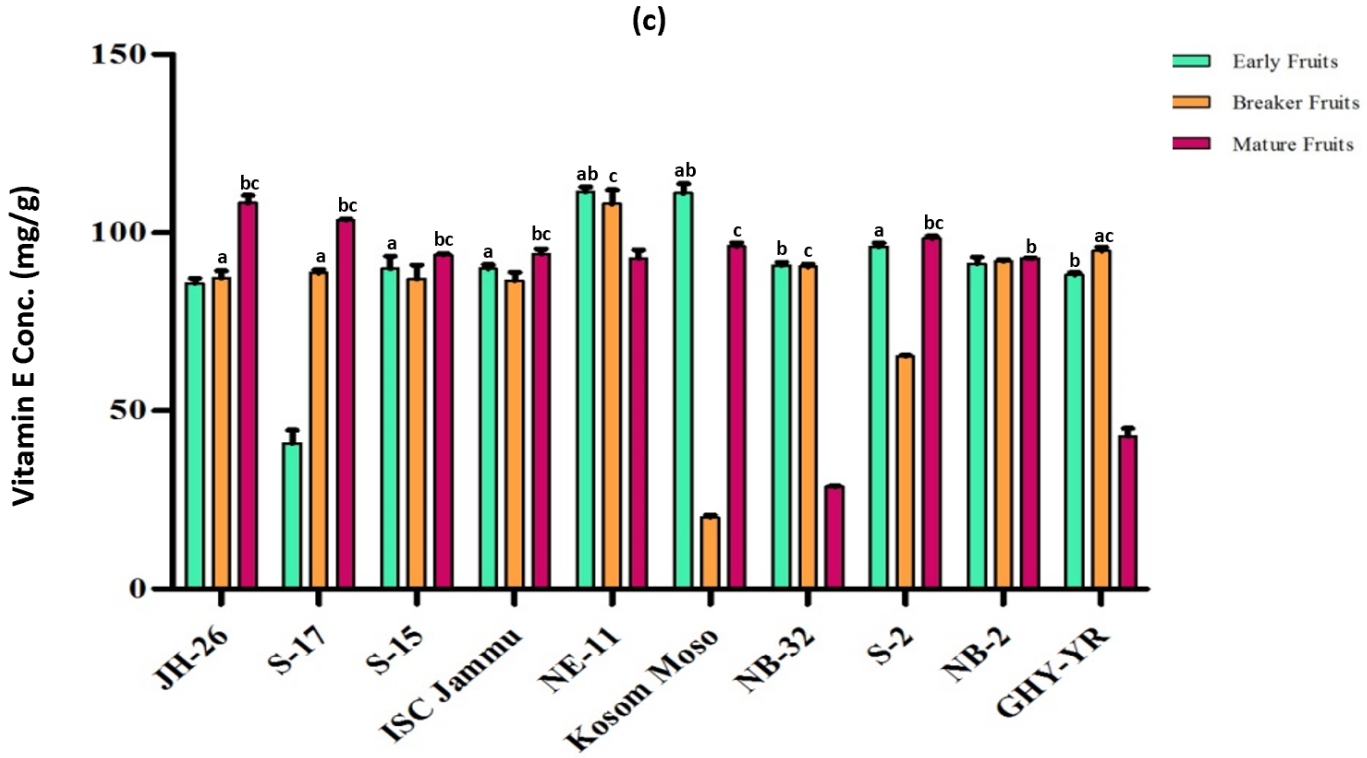


Fig. 3.10(c) Vitamin E contents in early fruit (EF), breaker fruit (BF), and mature fruit (MF) in 10 *Capsicum* genotypes belonging to *C. annuum*. Letters 'a', 'b' and 'c' represent significance with respect to the fruit stage 'a-between early and breaker', 'b-between early and mature' and 'c-between breaker and mature' using 2-way ANOVA Test (Dubey et al. 2025)

(d)

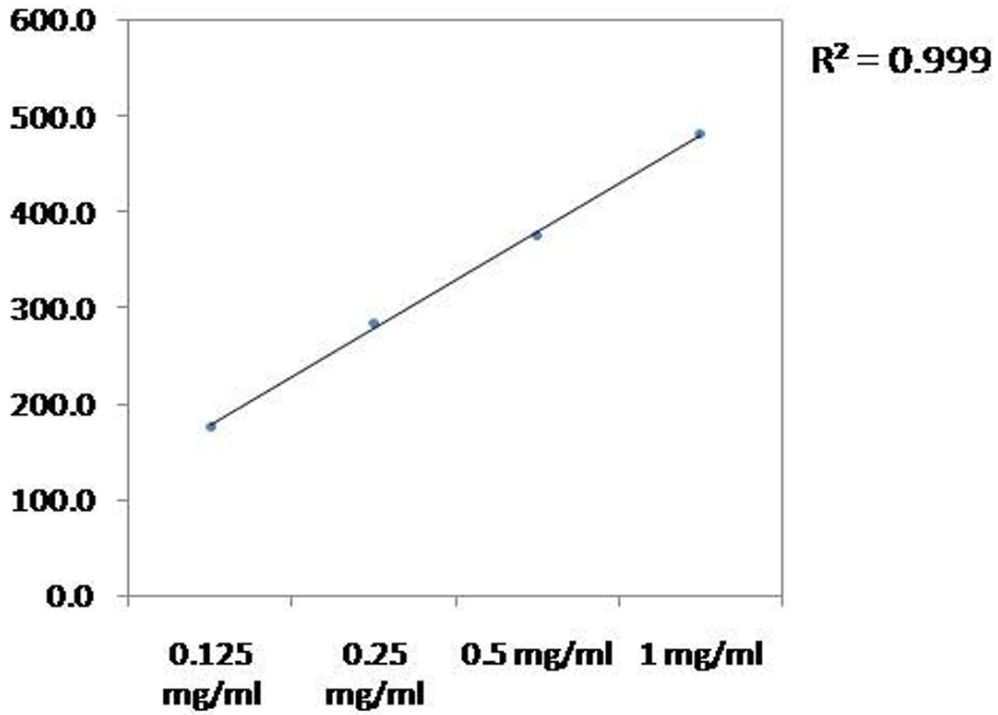


Fig. 3.10(d) Standard curve used for calculating and comparing Vitamin E content in 30 *Capsicum* germplasm through HPLC readings (Dubey et al. 2025).

Table 3.7 Vitamin C and Vitamin E contents in early fruit (EF), breaker fruit (BF), and mature fruit (MF) stages in 30 *Capsicum* genotypes belonging to C. chinense (Ca1 to Ca10), C. frutescens (Cf1 to Cf10), and C. annuum (Ca to Ca10)

Vitamins Content								
S. No	***Genotypes* Code number**	***Capsicum* Species**	**Vitamin C Content (mg/g)**			**Vitamin E Content (mg/g)**		
			Early Fruit Stage	**Breaker Fruit Stage**	**Mature Fruit Stage**	**Early Fruit Stage**	**Breaker Fruit Stage**	**Mature Fruit Stage**
1	Cc-1	*C. chinense*	24.51±0.23	25.88±0.97	19.02±7.04	70.66±0.07	83.70±1.43	93.32±1.17
2	Cc-2	*C. chinense*	35.26±0.21	26.07±0.00	29.37±1.05	86.66±3.54	108.49±0.55	77.53±0.42
3	Cc-3	*C. chinense*	15.86±1.15	25.75±0.89	32.04±2.61	92.72±1.69	108.11±2.50	60.29±0.71
4	Cc-4	*C. chinense*	23.78±0.71	24.08±1.13	20.67±9.52	91.29±1.28	6.59±0.71	86.84±1.00
5	Cc-5	*C. chinense*	27.27±2.87	23.43±0.42	22.19±0.24	92.73±3.10	86.76±2.85	104.28±2.61
6	Cc-6	*C. chinense*	18.86±6.58	24.89±4.07	34.08±0.62	94.51±1.40	86.52±1.80	109.29±3.51
7	Cc-7	*C. chinense*	15.97±0.72	20.87±8.55	24.32±0.60	56.81±1.41	100.88±0.19	87.58±1.86
8	Cc-8	*C. chinense*	26.96±1.50	25.16±0.71	28.16±0.00	89.10±0.11	19.68±0.00	103.25±3.07
9	Cc-9	*C. chinense*	38.41±2.14	31.82±2.31	28.63±0.14	91.21±0.59	108.78±4.97	37.21±1.03
10	Cc-10	*C. chinense*	16.38±0.52	24.46±1.00	25.93±0.98	109.39±0.52	85.02±1.83	72.97±0.71
11	Cf-1	*C. frutescens*	17.10±037	21.66±025	21.21±0.18	87.96±2.86	92.89±3.32	92.11±4.80
12	Cf-2	*C. frutescens*	20.85±2.03	21.38±0.52	23.90±4.77	44.78±0.45	107.01±0.68	104.32±2.11
13	Cf-3	*C. frutescens*	22.58±0.04	22.71±0.35	24.62±0.75	88.97±0.27	80.97±0.4	94.90±2.88
14	Cf-4	*C. frutescens*	27.78±1.00	23.60±1.64	27.09±1.11	105.00±4.57	96.99±0.71	102.21±0.99
15	Cf-5	*C. frutescens*	22.78±0.21	23.23±1.53	26.20±0.76	93.56±1.13	99.45±1.18	93.65±1.47
16	Cf-6	*C. frutescens*	15.44±1.06	20.22±0.81	18.74±1.42	61.32±2.53	59.35±1.3	96.79±1.94
17	Cf-7	*C. frutescens*	29.19±0.91	25.68±0.07	28.94±3.30	103.65±1.31	108.26±3.38	93.88±0.61
18	Cf-8	*C. frutescens*	23.41±1.06	17.77±0.35	16.94±0.58	96.48±3.46	96.17±1.60	3.32±1.84
19	Cf-9	*C. frutescens*	22.04±2.77	21.24±0.8	22.49±0.78	88.62±0.52	17.90±0.51	112.44±3.20
20	Cf-10	*C. frutescens*	15.43±1.08	16.19±0.06	22.41±0.00	30.68±1.45	98.97±2.83	104.00±0.71
21	Ca-1	*C. annuum*	19.12±0.14	19.04±4.67	23.55±2.32	85.76±1.44	87.34±2.01	108.27±2.12
22	Ca-2	*C. annuum*	21.33±1.97	21.27±0.14	22.17±1.92	42.00±3.54	88.66±0.85	103.55±0.20
23	Ca-3	*C. annuum*	35.87±1.07	24.89±3.03	17.28±0.80	89.88±3.59	87.01±3.91	93.86±0.36
24	Ca-4	*C. annuum*	15.21±0.00	20.58±0.26	27.58±0.41	89.94±1.17	86.59±2.39	94.10±1.34
25	Ca-5	*C. annuum*	33.50±2.90	27.24±0.50	26.80±0.33	111.43±1.33	108.07±3.89	92.81±2.40
26	Ca-6	*C. annuum*	16.97±0.73	21.49±0.71	26.99±0.11	111.12±2.54	20.23±0.59	96.24±0.94
27	Ca-7	*C. annuum*	25.80±0.36	24.79±6.96	29.86±6.74	90.87±0.76	90.43±0.71	28.75±0.14
28	Ca-8	*C. annuum*	22.47±4.71	34.38±0.32	41.77±1.65	96.11±0.99	65.33±0.21	98.40±0.4
29	Ca-9	*C. annuum*	26.56±0.61	17.81±0.14	18.53±0.04	91.36±1.78	92.09±0.35	92.81±0.11
30	Ca-10	*C. annuum*	21.730.39	18.89±0.21	24.27±0.52	88.14±0.64	94.88±1.00	4 2.93±2.08

Altogether, in our analysis, the Vitamin C content was found to be the highest in *C. chinense,* species followed by *C. annuum* and, *C. frutescens*, and in the case of Vitamin E, the highest content was found in *C. annuum* and *C. frutescens* followed by *C. chinense* species

3.3 OBJECTIVE 3: Identification and expression analysis of genes responsible for regulation of Vitamins C and E in *Capsicum* spp.

3.3.1 Identification of genes involved in biosynthesis and accumulation of Vitamin C and Vitamin E in *Capsicum* species.

A set of 70 tomato genes involved in Vitamin C and E biosynthesis in tomato were selected and further used to identify orthologous genes in *Capsicum*. Out of 70 tomato genes, 21 were from Vitamin C biosynthesis (Ioannidi et al. 2009) and 49 were from Vitamin E biosynthesis pathway (Quadrana et al. 2013). Based on BLASTp search analysis using the tomato genes, a total of 29, 44 and 36 genes were identified for Vitamin C biosynthetic pathway in *C. annuum*, *C. baccatum*, and *C. chinense* genome, respectively **(Supplementary Table 1)**. Similarly, 81, 85 and 70 genes were identified for Vitamin E biosynthesis in the three *Capsicum* genomes, respectively **(Supplementary Table 1)**. Of ll the genes identified for Vitamin C and Vitamin E, 106 (97.24%) were physically mapped on 12 different chrs. of *Capsicum*, and the rest were mapped on scaffolds of the *C. annuum* genome **[Fig. 3.11 (A), (B), and (C)]**. In *C. baccatum,* 111 (86.04%) genes and in *C. chinense* 100 (96.15%) genes were physically mapped on the 12 *Capsicum* chrs and the rest were mapped on scaffolds of the two respective genomes. In *C. annuum,* most of the Vitamin C biosynthetic genes were present on chr 3 (6 genes) and the maximum no. of Vitamin E biosynthetic genes were present on chrs 2, 3, and 6 (10 genes on each chr) **[Fig. 3.11 (A); (D)]**. In *C. baccatum,* for Vitamin

C, most of the genes were found on chr 9, and the minimum number of genes were on chrs 5 and 10 (1 gene on each chr); for Vitamin E, most of the genes were present on chrs 11, 9 and 10, and the minimum number of genes were present on chrs 7, 9 and 10 (3 genes on each chr) **[(Fig. 3.11 (B); (D)]**. However, in *C. chinense*, most of the genes for Vitamin C were present on chr 6, and the minimum number of genes were on chrs 8 and 10 (1 gene on each chr); for vitamin E, most of the genes were present on chr 6, and the minimum number was on chrs 9 and 10 (2 genes on each chr) **[Fig. 3.11 (C); (D)]**.

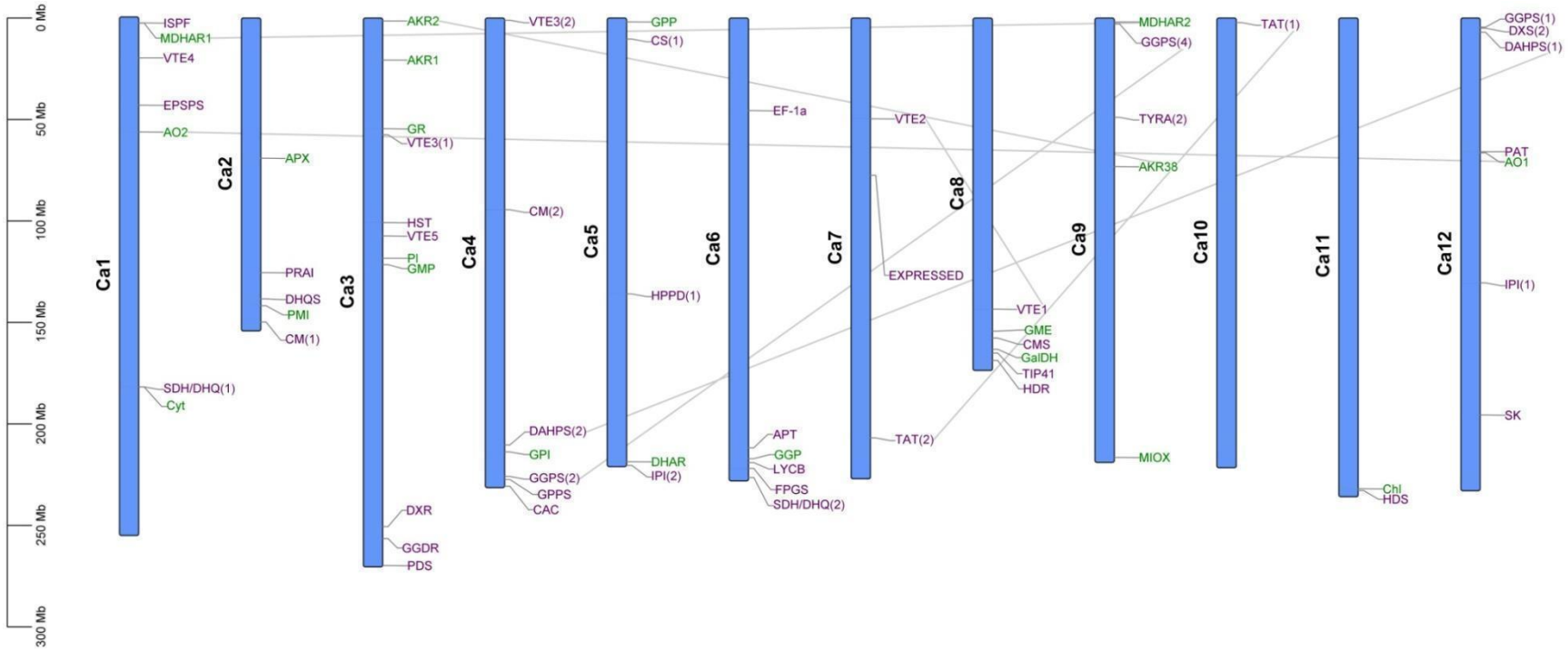


Fig. 3.11(A) Distribution of genes involved in Vitamin C (purple) and Vitamin E (green) biosynthesis on 12 chromosomes (Chr) of *Capsicum annuum.* Scale in mega bases (Mb) is given on the left side (Dubey et al. 2025).

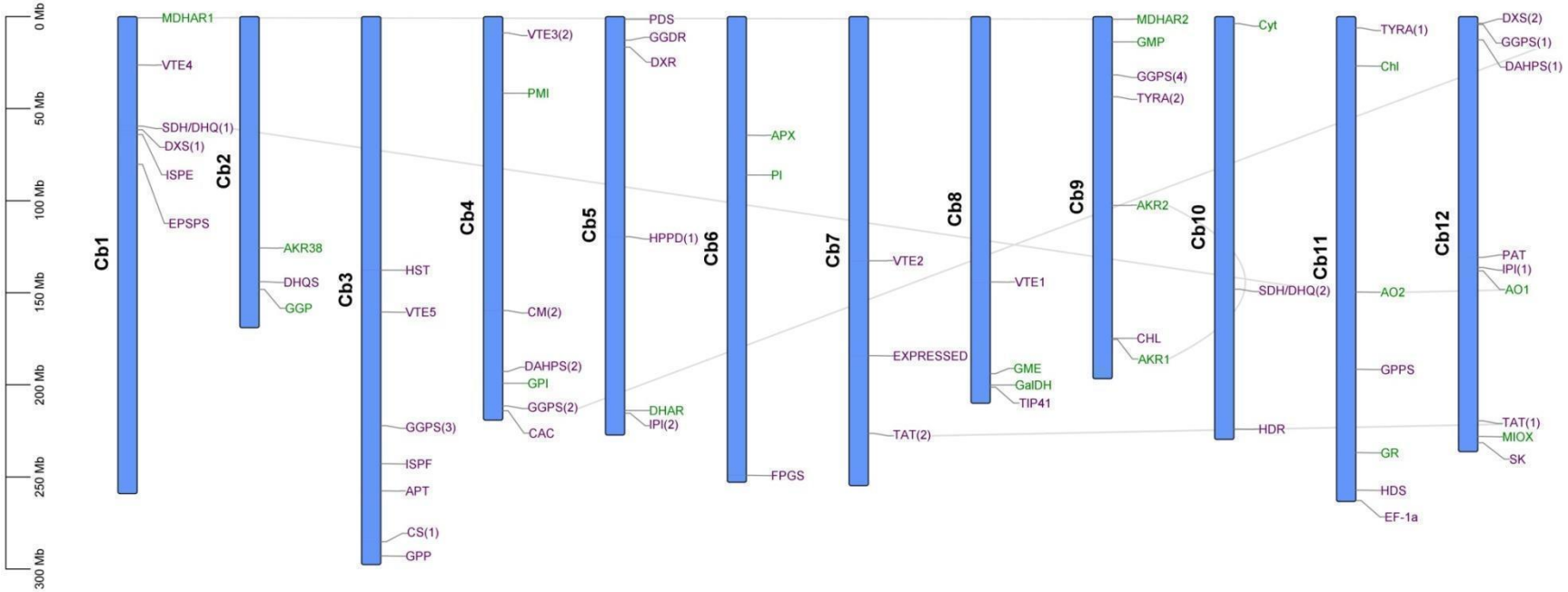


Fig. 3.11(B) Distribution of genes involved in Vitamin C (purple) and Vitamin E (green) biosynthesis on 12 chromosomes (Chr) of *Capsicum baccatum.* Scale in mega bases (Mb) is given on the left side (Dubey et al. 2025).

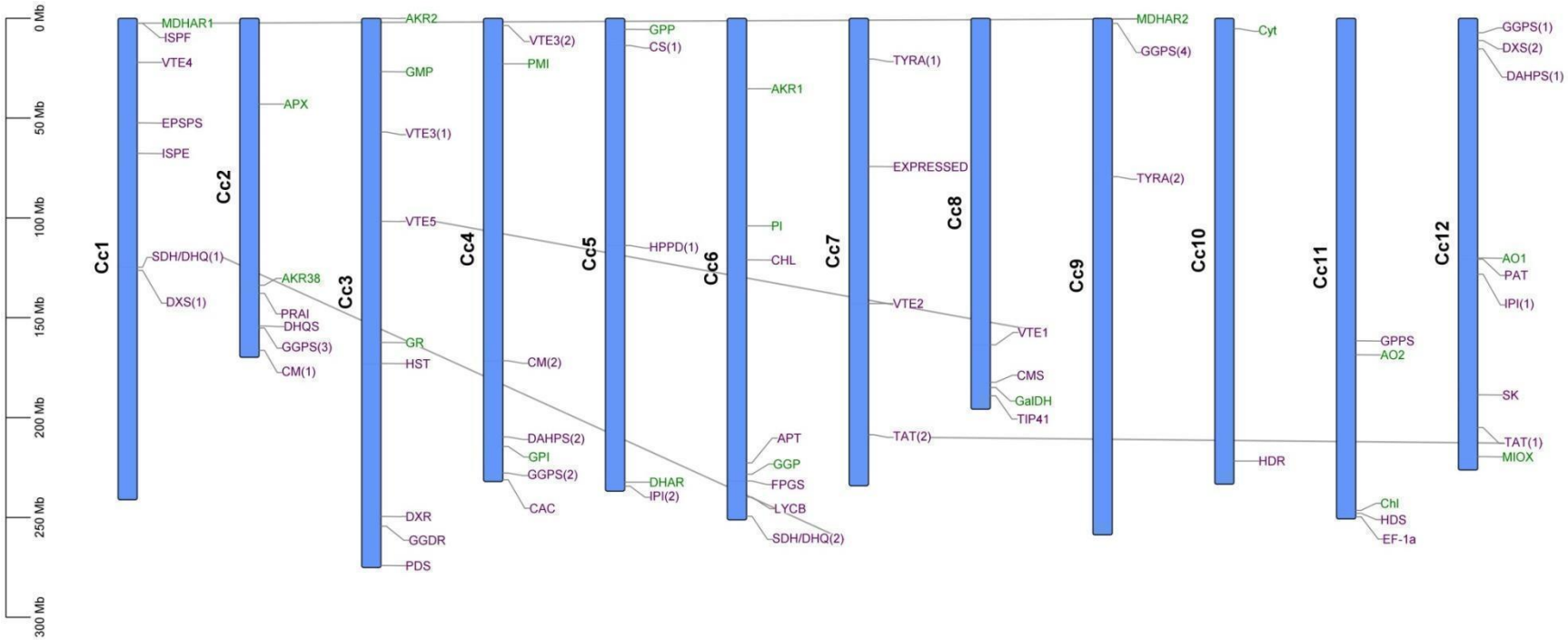


Fig. 3.11(C) Distribution of genes involved in Vitamin C (purple) and Vitamin E (green) biosynthesis on 12 chromosomes (Chr) of *Capsicum chinense.* Scale in mega bases (Mb) is given on the left side (Dubey et al. 2025).

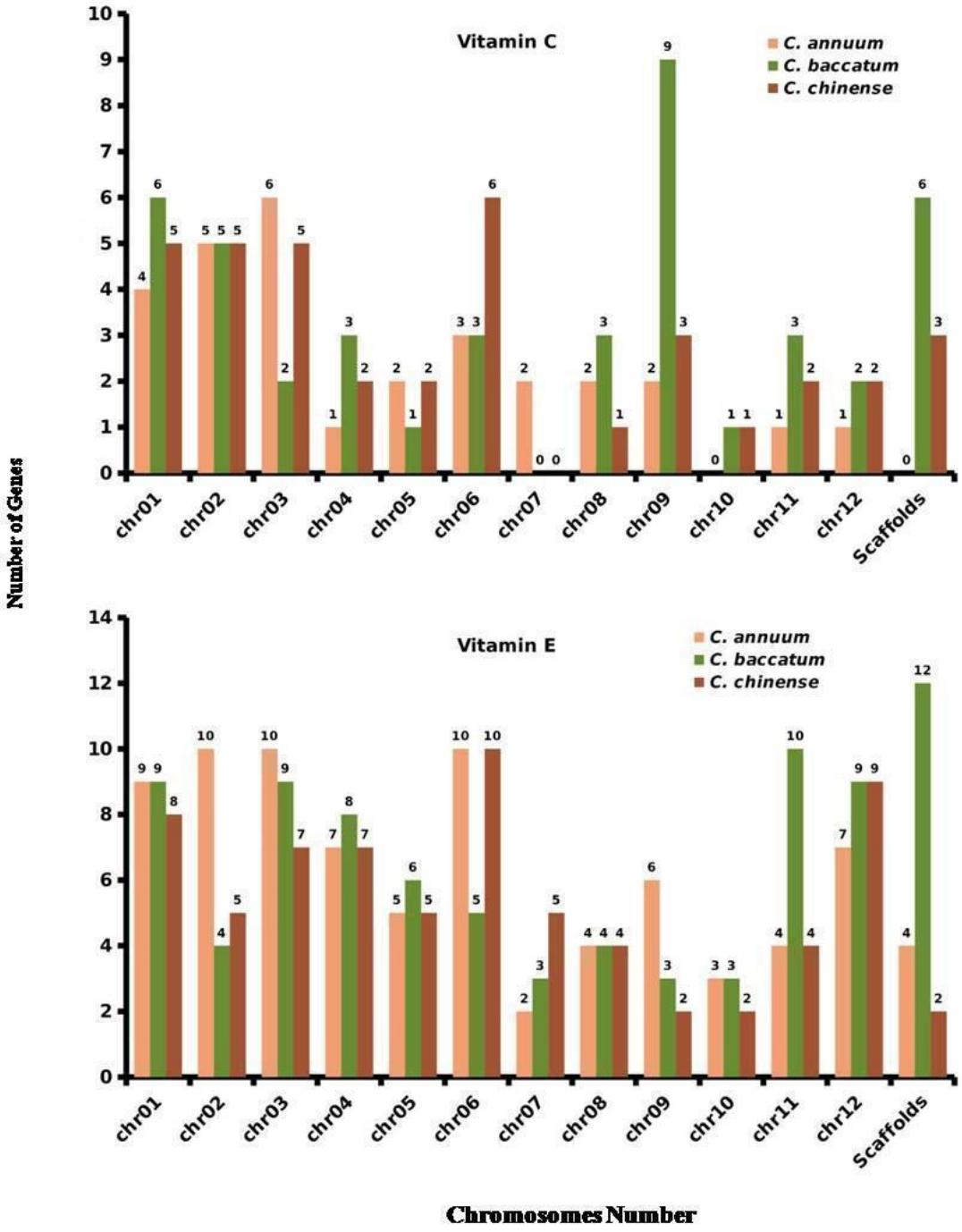


Fig. 3.11(D) Distribution of genes involved in Vitamin C and E biosynthesis on 12 chromosomes (Chr) of *C. annuum, C. baccatum*, and *C. chinense* (Dubey et al. 2025).

3.3.2 Expression analysis of genes involved in Vitamin C and Vitamin E biosynthesis /accumulation in transcriptome data

The expression analysis of genes involved in Vitamin C and E biosynthesis in the transcriptomes of different fruit developmental stages from three *Capsicum* species revealed the differential expression of vitamin biosynthesis genes **(Fig. 3.12)**. Variation of transcript abundance was observed for Vitamin C biosynthesis genes in early, breaker and mature fruit stages of *C. annuum, C. chinense* and *C. frutescens* species. A higher abundance of transcripts was observed for *GDP-mannose-3,5-epimerase (CaGME), Dehydro ascorbate reductase (CaDHAR)*, and *Monodehydro ascorabte reductase (CaMDHAR1)* in all the fruit developmental stages across the three species. Transcript of gene *SOD chloroplast superoxide dismutase (CaChl)* was also found higher in all the tissues except BF of *C. annuum*. On the other hand, consistent lower transcript abundance was identified for genes *Aldo/keto reductase (CaAKR38), thylakoid-bound ascorbate peroxidase (CaAPX), myo-inositol oxygenase (CaMIOX), ascorbate oxidase (CaAO1) and CaAO2*. Some genes like *L-galactose-1-phosphate phosphatise (CaGPP) and glucose-6-phosphate isomerise (CaGPI)* showed higher expression in EF as compared to BF and MF in all the three species. Gene *CaAKR1* showed higher expression in *C. chinense* (in all the fruit developmental stages) as compared to *C. annuum* and *C. frutescens*.

The Vitamin C content in the three *Capsicum* species were 7.50 mg/g (EF), 9.96 mg/g (BF), and 2.64 mg/g (MF) for *C. annuum*; 12.07 mg/g (EF), 7.40 mg/g (BF), 12.02 (MF) for *C. chinense*; and 10.11 mg/g (EF), 17.23 mg/g (BF); and 24.18 mg/g (MF) for *C. frutescens*. The

Correlation coefficient(r) between the transcript abundance and Vitamin C content was estimated and was found in the range -1 to +1 in the individual *Capsicum* species **(Fig. 3.12)**. At the individual gene level, 18 out of 21 genes showed a strong positive correlation (r > 0.9) with Vitamin C content in one or more *Capsicum* species. The remaining three genes showed a weak positive correlation with Vitamin C content.

Similarly, Vitamin E biosynthetic genes showed variable transcript expression in the three *Capsicum* species. Higher abundance of transcript was observed for genes *γ-tocopherol C-methyl transferase (CaVTE4), 4-hydroxy-3-methylbut-2-enyl-diphosphate reductase (CaHDR)* and *elongation factor 1-alpha (CaEF-1α)* in all the developmental stages of fruits in three *Capsicum* species. However, lower transcript abundance was observed in the case of chlorophyllase *(CaCHL), tyrosine amino transferase (CaTAT(2)), arogenate dehydrogenase(CaTYRA(2)), geranyl pyrophosphate synthase (CaGPPS), clathrin adapter complexes medium subunit (CaCAC), isopentenyl diphosphate d-isomerase (CaIPI(2), CaVTE5, homogentisatesolanesyl transferase (CaHST)* and *2-C-methy-D-erythritol-4-phosphatecitydylyl transferase (CaCMS)* in all the tissues studied. Some genes showed consistent stage-specific expression. For instance, *lycopene b-cyclase (CaLYCB), phytoene saturase (CaPDS), and (CaTAT(1))* showed enhanced expression in MF; and isopentenyl diphosphate d-isomerase prephenate aminotransferase *(CaIPI), TIP-41 like family protein (CaTIP41), chorismate mutase [CaCM(1)],* and *3-deoxy-D-arabino-heptulosonate-7-P synthase (CaDAHPS)* showed higher expression in EF. Species-specific expression was found for *CaVTE* gene which showed significantly higher expression in *C. chinense* and *C. frutescens*. Similarly, gene *1-deoxy-D-xylulose-5-P synthase (CaDXS)* was specifically expressed in *C.*

chinense only.

The Vitamin E content for the three *Capsicum* species were as as follows- 7.34 mg/g (EF), 8.15 mg/g (BF), 4.25 mg/g (MF) for *C. annuum,* 11.26 mg/g (EF), 8.70 mg/g (BF), and 9.20 (MF) for *C. chinense*; and 6.28 mg/g (EF), 6.90 mg/g (BF) and 7.94 mg/g (MF)] for *C. frutescens*. The majority (32 out of 81) of Vitamin E biosynthesis genes showed a strong positive correlation (r>0.9) with Vitamin E content in one or more *Capsicum* species, although some of these genes showed weak positive correlation and some were negatively correlated **(Fig. 3.12)**.

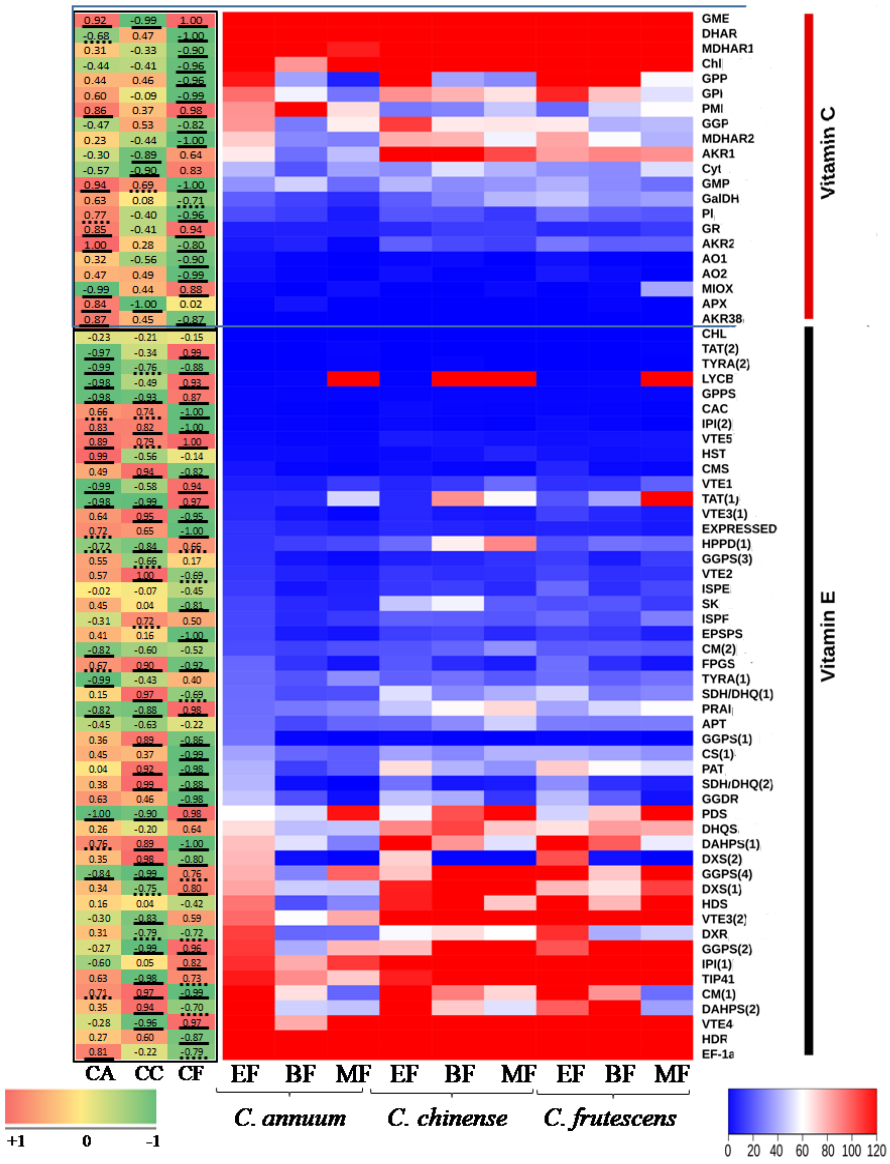


Fig. 3.12. Heat map showing different levels of expression of genes involved in Vitamin C and Vitamin E biosynthesis in transcriptome of different fruit tissues - early fruit (EF), breaker fruit (BF), and mature fruit (MF) of three *Capsicum* spp. viz. *C. annuum, C. chinense, C. frutescens* (blue- very low or negligible; red- high). Heat map in red-yellow-green color showing correlation of gene expression and vitamin content in *C. annuum* (CA), *C. chinense* (CC) and *C. frutescens* (CF) genotypes. Values underlined with solid (-) and dash (...) represent significance at $p<0.01$ and $p<0.05$ level respectively.

3.3.3 Validation of expression of genes involved in Vitamin C biosynthesis/accumulation using quantitative Real-Time PCR (qRT- PCR) in contrasting *Capsicum* genotypes (species)

For qRT-PCR validation, we selected 11 genes from the Vitamin C biosynthetic pathway (*CaPMI, CaGMP, CaAKR1, CaAKR2, CaGPI, CaGPP, CaGGP, CaAKR38, CaGME, CaGalDH* and *CaMIOX*) whose orthologs were reported as the most important in tomato (Ioannidi et al. 2009). Sequences used for the primers of qRT-PCR analysis are given in **Table 3.8.** Out of 11 genes selected, 7 genes are known to be involved in the main Vitamin C biosynthetic pathway, while 4 are known to be associated with the alternate pathway of ascorbate biosynthesis. The gene expression patterns of these 11 genes were highly variable across the tissues as well as the *Capsicum* species. A few genes such as *CaAKR38*, *CaAKR1*, *CaAKR2*, and *CaGMP* showed a significant reduction in expression in fruit tissues as compared to the leaves in all the three species **(Fig. 3.13)**. Unlikely, some genes showed higher expression in the fruit developmental stages compared to the leaf tissue. For instance, *CaMIOX* showed up to 40-fold higher expression in MF as compared to leaf in *C. chinense*. Gene *CaGME* also got upregulated in MF of *C. chinense* and *C. frutescens*, although down regulated in MF of *C. annuum*. Similarly, other genes like *CaGPI*, *CaGPP, CaGalDH*, and *CaGGP* showed similar expression patterns in *C. chinense* and *C. frutescens* and opposite expression patterns in *C. annuum*. The correlation analysis between Vitamin C gene expression and Vitamin C content revealed a positive correlation for 7 out of 11 genes ($r>0.7$ and p value< 0.01) in one or more *Capsicum* species as shown in **Fig. 3.13 (G)**.

Table 3.8 Primer sequences used for qRT-PCR analysis for Vitamin C and E biosynthetic genes in *Capsicum* species.

S. No.	Gene	Forward primer	Reverse primer
Vitamin C			
1	*CaGGP*	TACGGAAAGTTTGTTGCTGATGA	TCCTCCCATTCCCCAAGAAG
2	*CaGMP*	AGAGGAGTCCGCATCAAAAAAC	CGAGCCCATTGTCCAACAGT
3	*CaGME*	TGACGATGAAGGTGTGGATGA	AAGAAGGTTGCTCCTGAAGTCTAAAA
4	*CaMIOX*	AAAGTGCTACTTCACCCAAGTTTTG	AAAAGCACATCCAAGAGGAAATG
5	*CaAKR38*	CATGTGTCCGGATTGGAATTC	TTGGTGTTGCTCGTTTCATTG
6	*CaAKR2*	CACTAGGGCGTGGCTTTTTG	TGGCAGATGCTTTCGGTAGTC
7	*CaGPI*	GGTTTACTTCCGGCTGCACTT	CTGTTTGCTTCGTCCATCGA
8	*CaGPP*	GCAGGTGGCGCGGTAATA	TGGGCTGTGATGTCAAATTCC
9	*CaGalDH*	ATTGCAGCTCCGACCACTTG	AAACCTTGCCGAGTGGAGAA
10	*CaAKR1*	GGGACCTCCATGACTTGCA	CCGCTTGAGGCTTTCTTCAC
11	*CaPMI*	TGTCCCGCTAAACAATCTTGTG	GCATACGGCTGATTCTCATCAA
Vitamin E			
1	*CaGGDR*	TTTTGTAGAGATGTGCGCAGATG	TCCCTGGTGCCACTTTCTTG
2	*CaHPPD*	GTGGCTGTGGCTGAAGTTCA	ATGAGGCTGTCCGCATCTCT
3	*CaIPI*	GGCTCATCGCCTCCCCTATA	GGATGATTTGGTTAGCCGTAGTG
4	*CaTAT*	CCTGGTTTCCCAATTTATGCA	CCCTTGTCTGGAACGAGATCA
5	*CaTYRA*	TTGGCTTTTGAGGCTTTGAAG	TCCGCTTTCCCAAACAGTTG
6	*CaVTE1*	GCCTGCTGCTTTTCCTGTATTT	CCCACTCTATCCAGCCTGTTG
7	*CaVTE2*	TGTCGTCTTTTGCCATTTTGAG	ACTGATGAAAAGAGCCCAAAACAA
8	*CaVTE3*	CATGGGCTTATCATGGGATGT	GCCTTTGGACCGAGCTGTAG
9	*CaVTE4*	TCAACTTCCACCATCCAATCC	TGTGGTCTATGGAAAGTGGAGATAAT
10	CaVTE5	GGCCTTTCGTTGGCTACTGA	GGCCCTCTGAGCAATTCTTCT

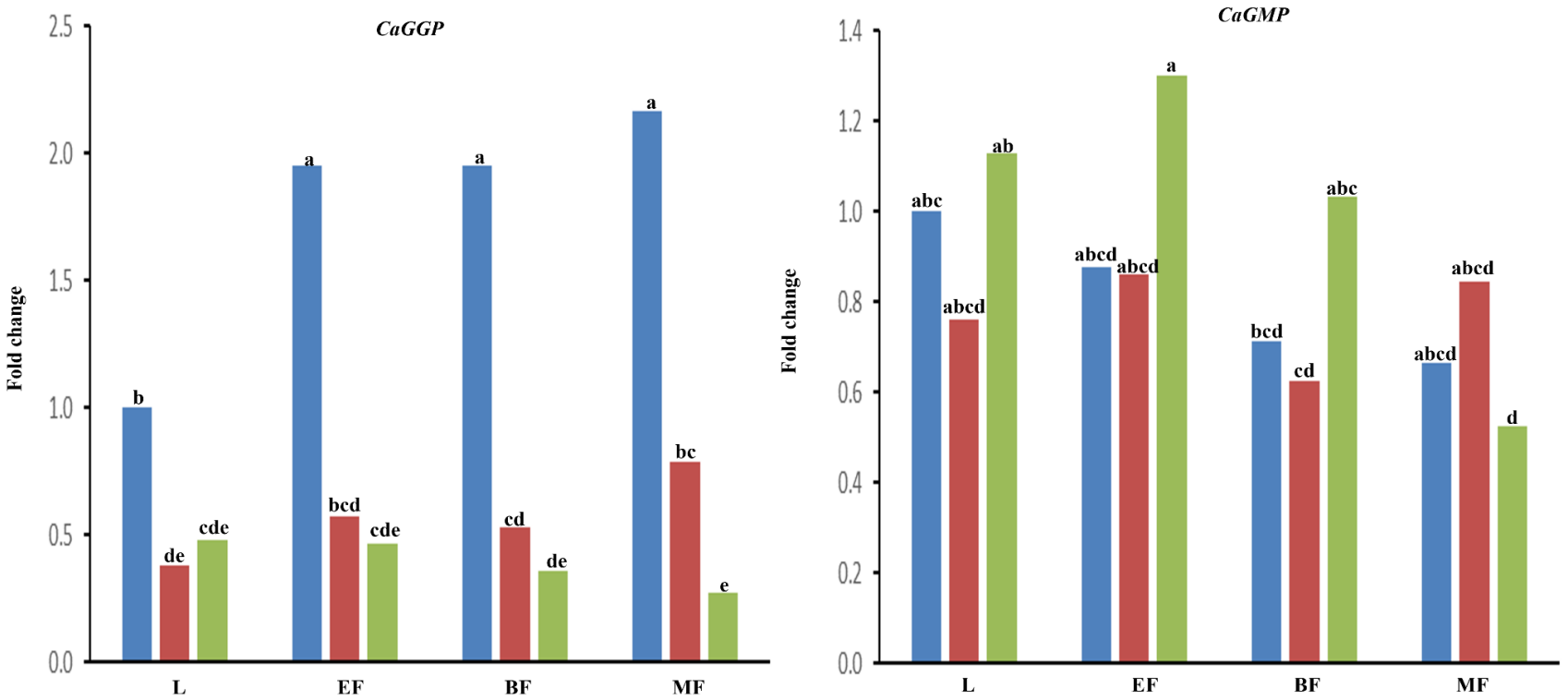


Fig. 3.13 (A) Expression analysis of genes (*CaGGP* and *CaGMP*) involved in Vitamin C biosynthetic pathway in four tissues - leaf (L), early fruit (EF), breaker fruit (BF), and mature fruit (MF) in *C. annuum* (blue color), *C. chinense* (red color), and *C. frutescens* (green color). For each gene, samples with the same letter (a to g) refer to the difference between the samples not statistically significant; and samples with different letters refer to the significant difference ($p < 0.05$) estimated using Tukey's HSD test (Dubey et al. 2025).

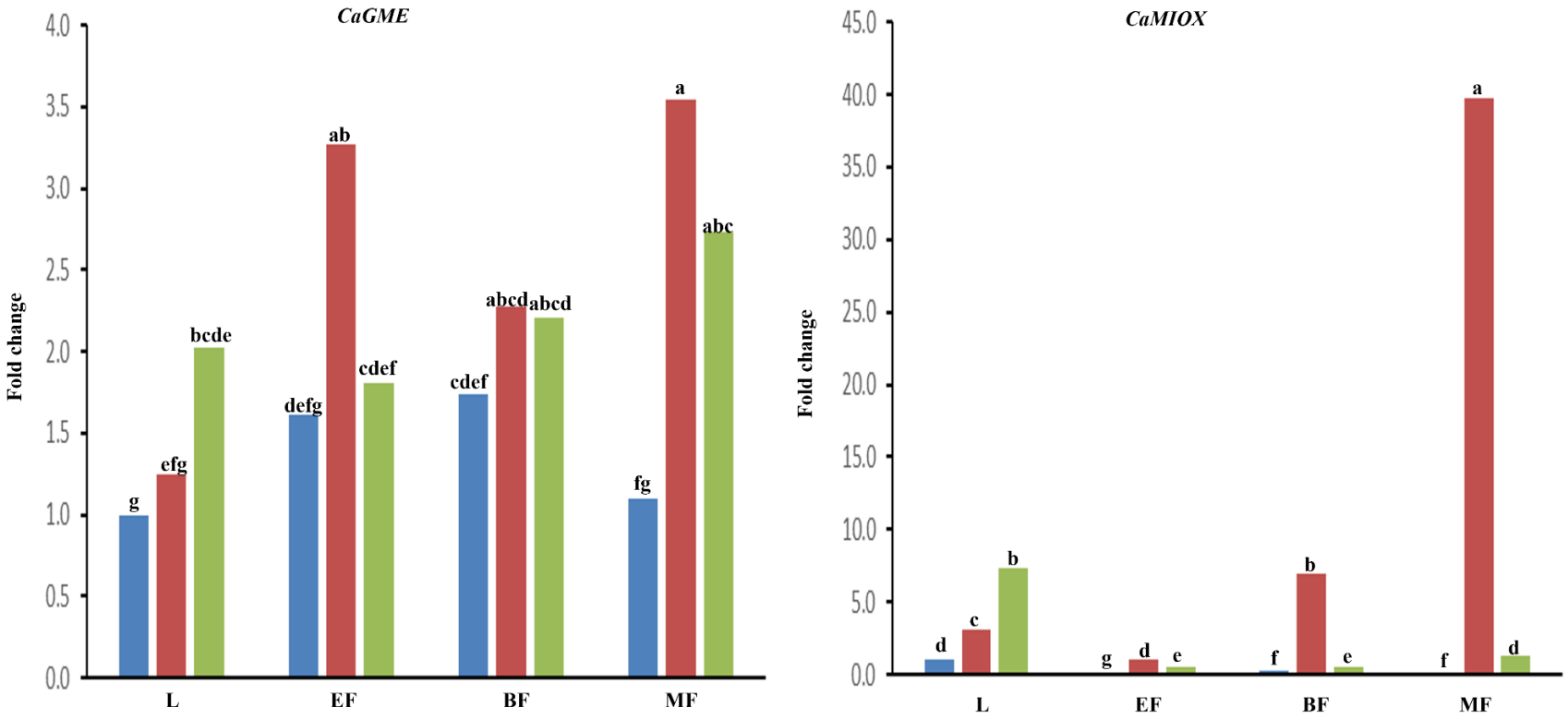


Fig. 3.13 (B) Expression analysis of genes (*CaGME* and *CaMIOX*) involved in Vitamin C biosynthetic pathway in four tissues - leaf (L), early fruit (EF), breaker fruit (BF), and mature fruit (MF) in *C. annuum* (blue color), *C. chinense* (red color), and *C. frutescens* (green color). For each gene, samples with the same letter (a to g) refers to the difference between the samples not statistically significant; and samples with different letters refer to the significant difference ($p < 0.05$) estimated using Tukey's HSD test (Dubey et al., 2025).

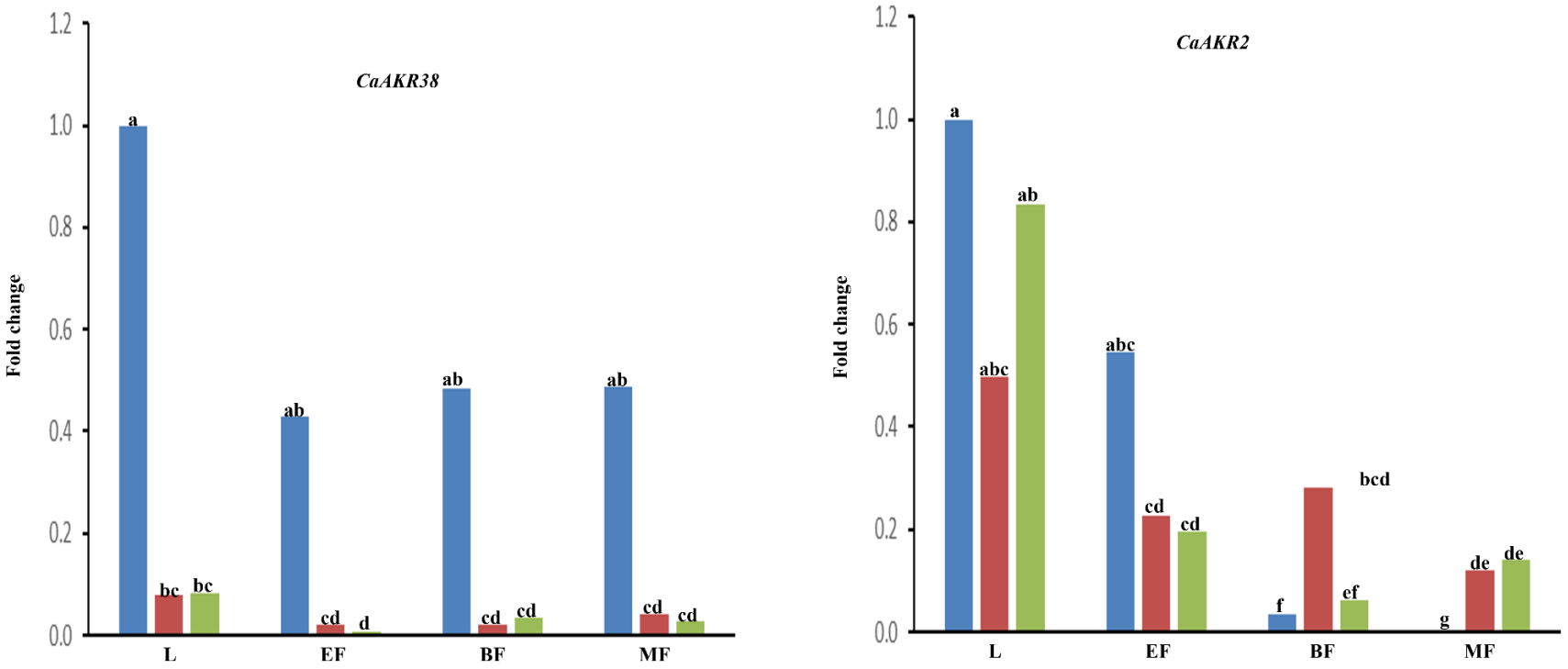


Fig. 3.13 (C) Expression analysis of genes (*CaAKR38* and *CaAKR2*) involved in Vitamin C biosynthetic pathway in four tissues - leaf (L), early fruit (EF), breaker fruit (BF), and mature fruit (MF) in *C. annuum* (blue color), *C. chinense* (red color), and *C. frutescens* (green color). For each gene, samples with the same letter (a to g) refers to the difference between the samples not statistically significant; and samples with different letters refer to the significant difference ($p < 0.05$) estimated using Tukey's HSD test (Dubey et al. 2025).

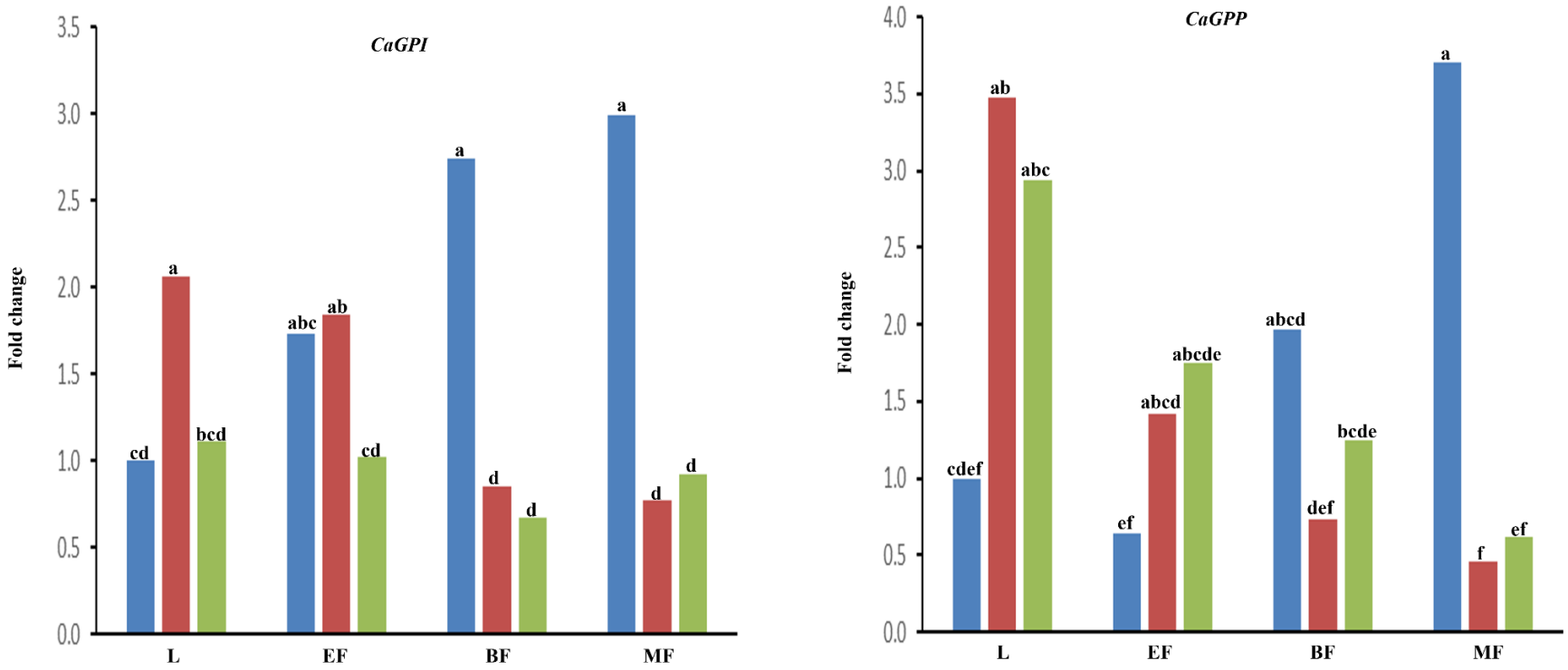


Fig. 3.13(D) Expression analysis of genes (*CaGPI* and *CaGPP*) involved in Vitamin C biosynthetic pathway in four tissues - leaf (L), early fruit (EF), breaker fruit (BF), and mature fruit (MF) in *C. annuum* (blue color), *C. chinense* (red color), and *C. frutescens* (green color). For each gene, samples with the same letter (a to g) refers to the difference between the samples not statistically significant; and samples with different letters refer to the significant difference ($p < 0.05$) estimated using Tukey's HSD test (Dubey et al. 2025).

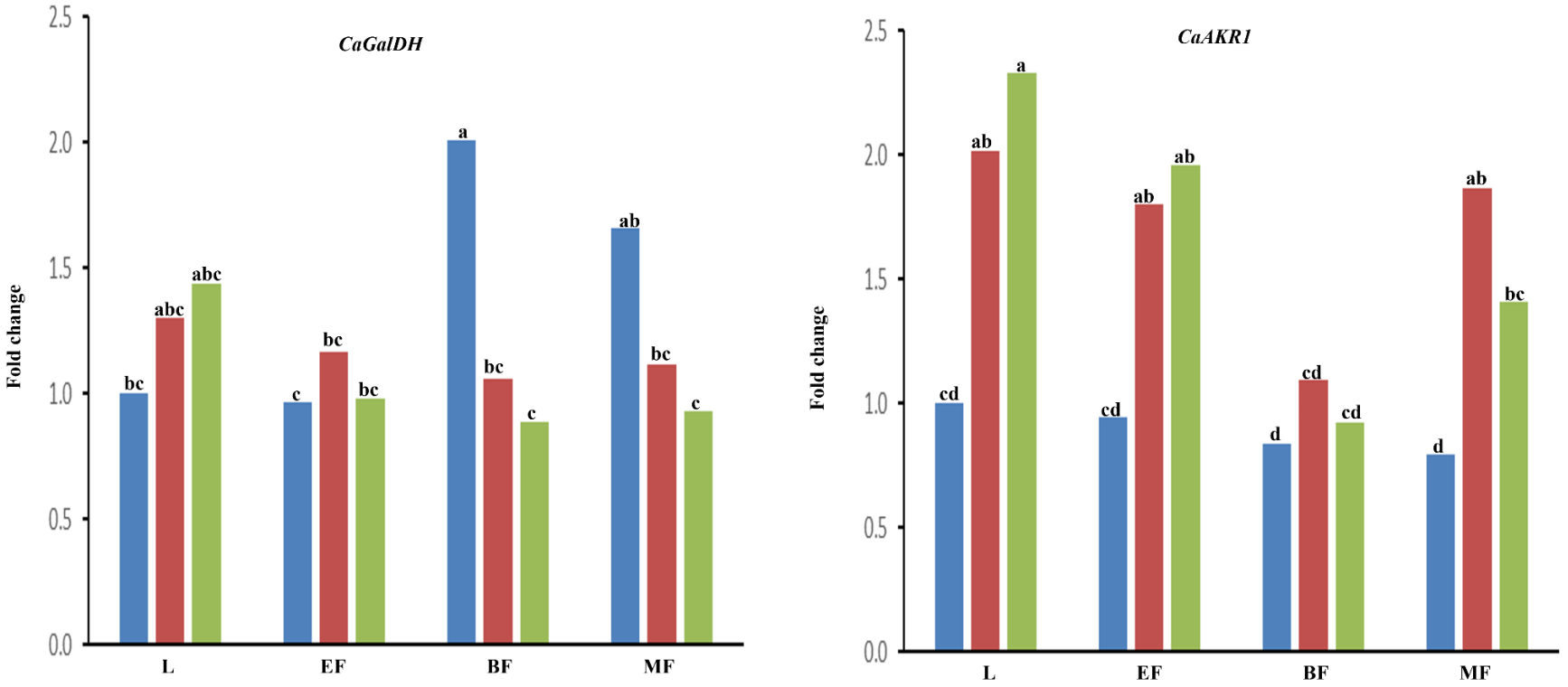


Fig. 3.13 (E) Expression analysis of genes (*CaGalDH* and *CaAKR1*) involved in Vitamin C biosynthetic pathway in four tissues - leaf (L), early fruit (EF), breaker fruit (BF), and mature fruit (MF) in *C. annuum* (blue color), *C. chinense* (red color), and *C. frutescens* (green color). For each gene, samples with the same letter (a to g) refers to the difference between the samples not statistically significant; and samples with different letters refer to the significant difference ($p < 0.05$) estimated using Tukey's HSD test (Dubey et al. 2025).

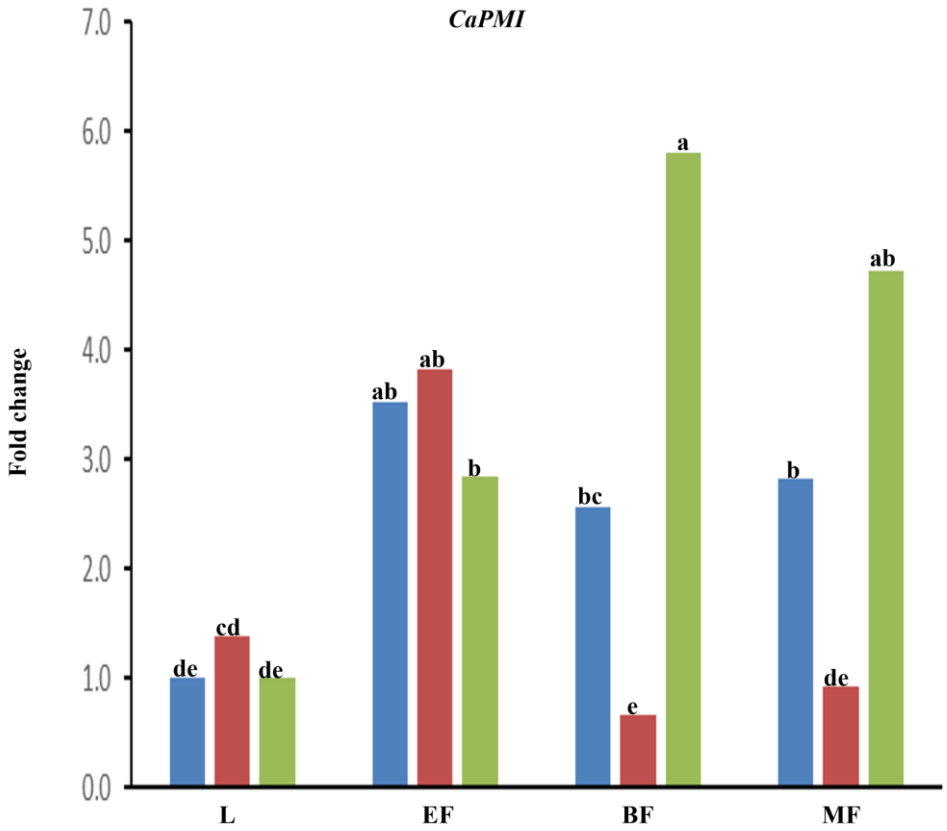
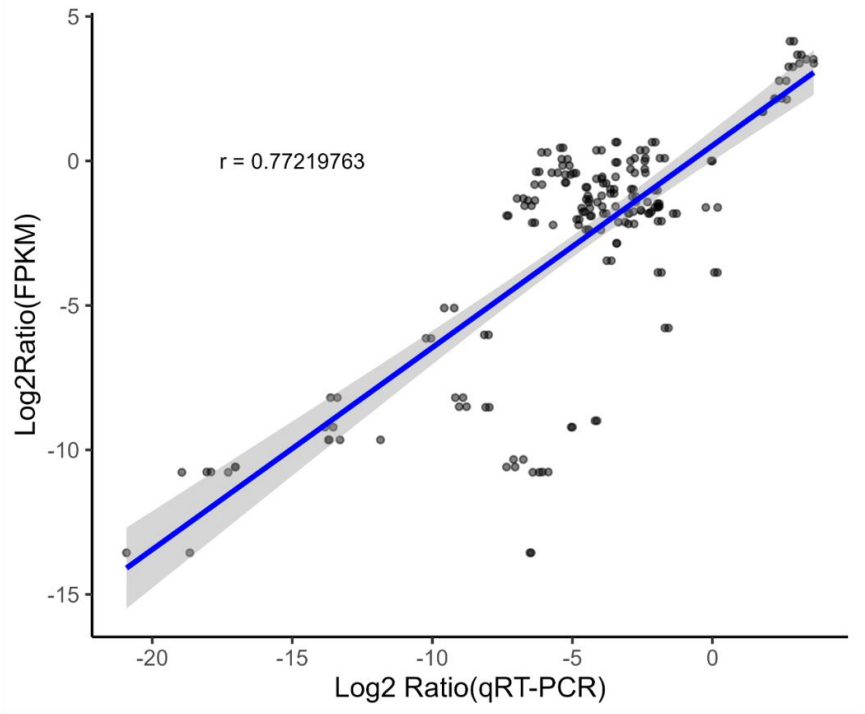


Fig. 3.13(F) Expression analysis of gene (*CaPMI*) involved in Vitamin C biosynthetic pathway in four tissues - leaf (L), early fruit (EF), breaker fruit (BF), and mature fruit (MF) in *C. annuum* (blue color), *C. chinense* (red color), and *C. frutescens* (green color). For each gene, samples with the same letter (a to g) refers to the difference between the samples not statistically significant; and samples with different letters refer to the significant difference ($p < 0.05$) estimated using Tukey's HSD test (Dubey et al. 2025).

CA	CC	CF	
-0.60*	-0.39	-1.00**	*CaGGP*
0.57	0.50	-0.93**	*CaGMP*
0.37	-0.42	0.95**	*CaGME*
0.49	-0.10	-0.48	*CaMIOX*
0.43	0.86**	-0.29	*CaAKR38*
0.49	0.57	-0.61*	*CaAKR2*
-0.55	0.75**	-0.53	*CaGPI*
-0.76**	0.83**	-0.89**	*CaGPP*
-0.13	0.96**	-0.62*	*CaGalDH*
0.60*	0.91**	-0.71**	*CaAKR1*
-0.44	0.19	0.76**	*CaPMI*



(a) **(b)**

Fig. 3.13(G) (a) Heat map in red-yellow-green color showing correlation of gene expression and Vitamin C content in *C. annuum* (CA), *C. chinense* (CC) and *C. frutescens* (CF) genotypes. * and ** showed significance at 0.05 and 0.01 levels, respectively. **(b)** The scatter plot shows the Pearson correlation between qRT-PCR data and transcriptome data for 11 Vitamin C genes (Dubey et al. 2025).

Out of total Vitamin E biosynthesis genes, 10 genes (*CaVTE4, CaVTE2, CaGGDR, CaVTE5, CaVTE3(1), CaHPPD(1), CaVTE1, CaTAT(1), CaIPI(1)* and *CaTYRA(1)*) whose orthologues were involved in advanced steps of tocopherol biosynthesis in tomato (Quadrana et al. 2013) were selected for qRT-PCR validation. Sequences for qRT-PCR primers are given in **Table 3.8**. Like Vitamin C biosynthesis genes, variable expressions were also observed for Vitamin E biosynthesis genes **(Fig. 3.14)**. Gene *CaGGDR* got down regulated (up to 263-fold lower) as fruit matures, however, gene *CaTAT(1)* showed higher expression (up to 71-fold) in advanced fruit developmental stages in all the three *Capsicum* species. Fifteen-fold higher expression of *CaHPPD* was observed in BF as compared to the leaf in *C. chinense*. Correlation analysis between Vitamin E gene expression and Vitamin E content revealed positive correlation for 6 out of 10 genes ($r>0.8$ and p value< 0.01) in one or more *Capsicum* species **(Fig. 3.14 (F)**.

The correlation analysis of gene expressions in transcriptome and qRT-PCR data showed a similar trend of expression for Vitamin C genes in three fruit stages among the three *Capsicum* species with correlation coefficient greater than 0.77 **[Fig. 3.14 (F)]**. Similarly, for Vitamin E biosynthesis genes, we observed significant correlation or similar trend of expression in transcriptome and qRT-PCR analysis with correlation coefficient greater than 0.52 **[Fig. 3.14 (F)]**.

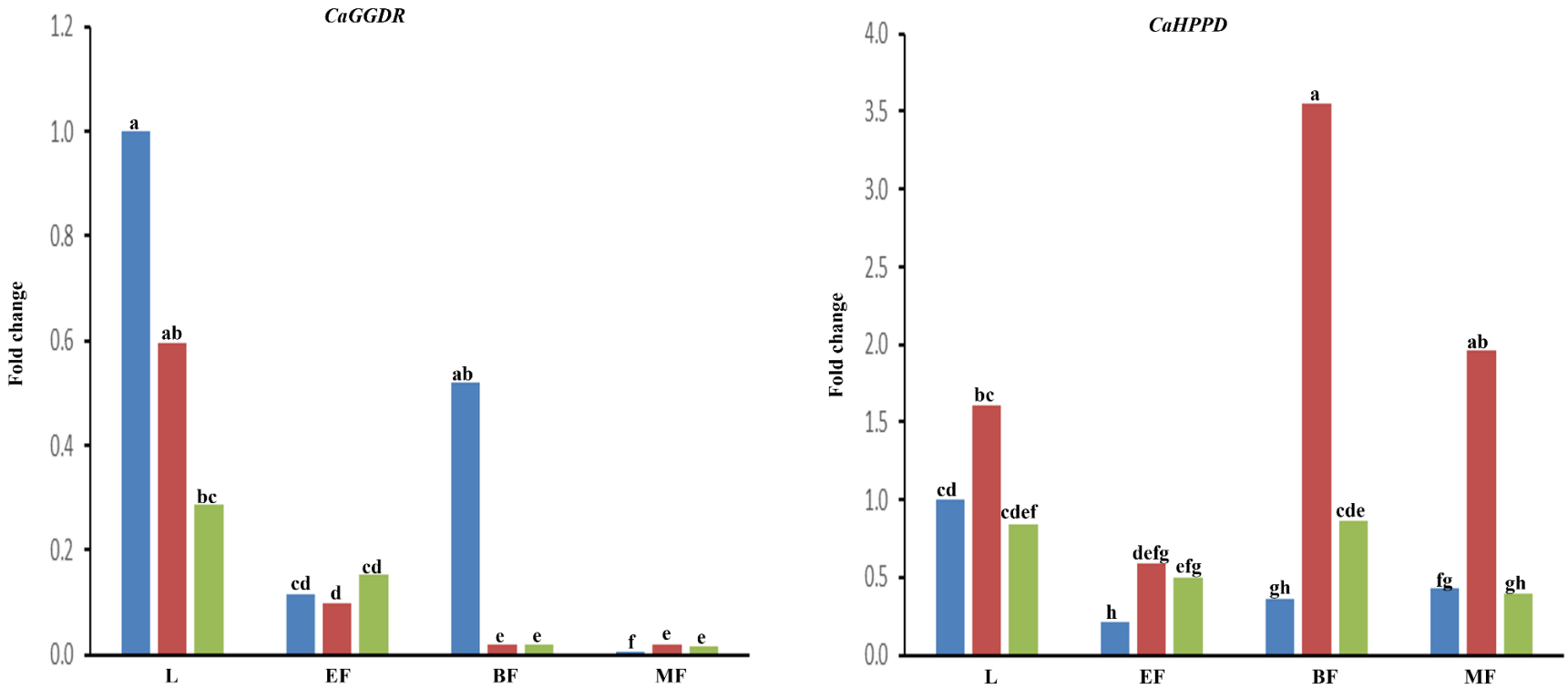


Fig. 3.14(A) Expression of genes involved (*CaGGDR* and *CaHPPD*) in Vitamin E biosynthetic pathway in four tissues - leaf (L), early fruit (EF), breaker fruit (BF), and mature fruit (MF) in *C. annuum* (blue color), *C. chinense* (red color), and *C frutescens* (green color). For each gene, samples with the same letter (a to h) refer to the difference between the samples not statistically significant; and samples with different letters refer to the significant difference ($p < 0.05$) estimated using Tukey's HSD test (Dubey et al. 2025).

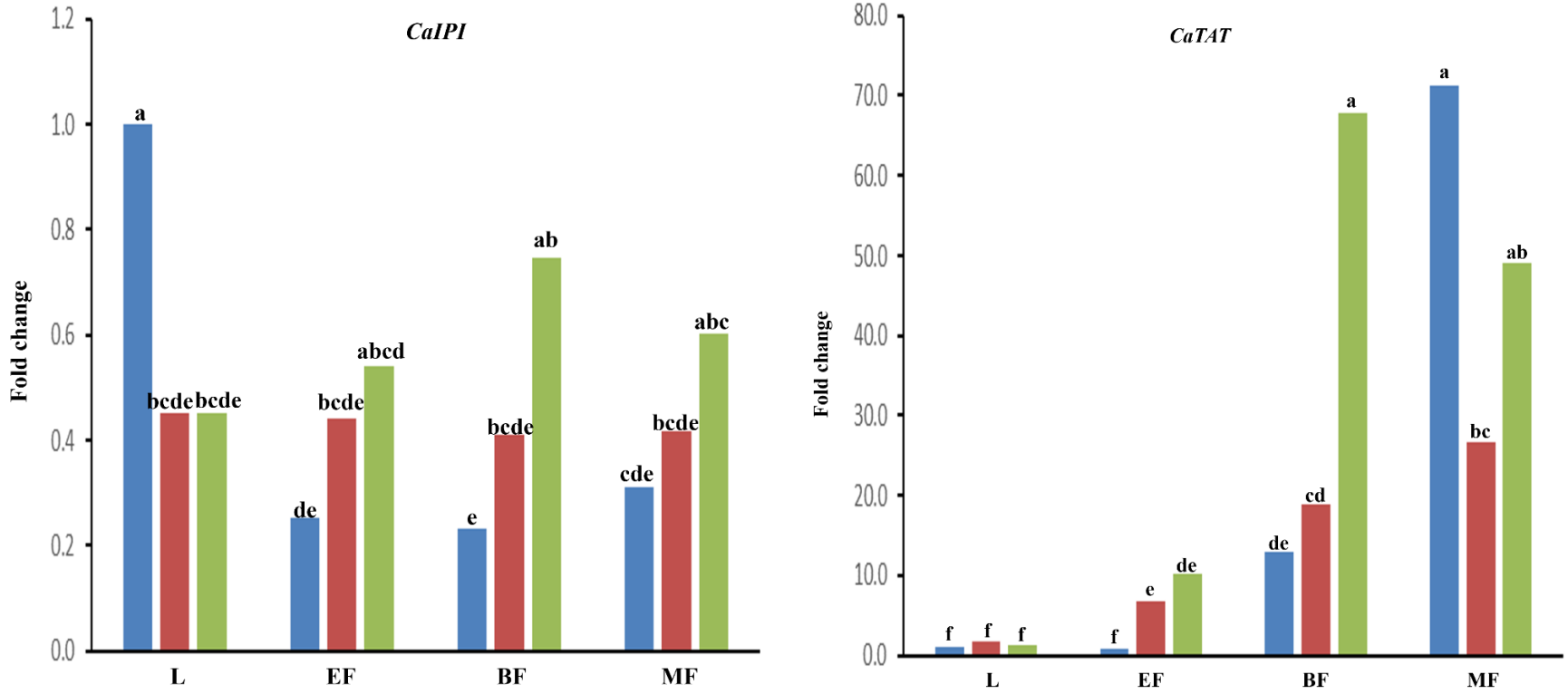


Fig. 3.14(B) Expression of genes involved (*CaIPI* and *CaTAT*) in Vitamin E biosynthetic pathway in four tissues - leaf (L), early fruit (EF), breaker fruit (BF), and mature fruit (MF) in *C. annuum* (blue color), *C. chinense* (red color), and *C frutescens* (green color). For each gene, samples with the same letter (a to h) refer to the difference between the samples not statistically significant; and samples with different letters refer to the significant difference ($p < 0.05$) estimated using Tukey's HSD test (Dubey et al. 2025).

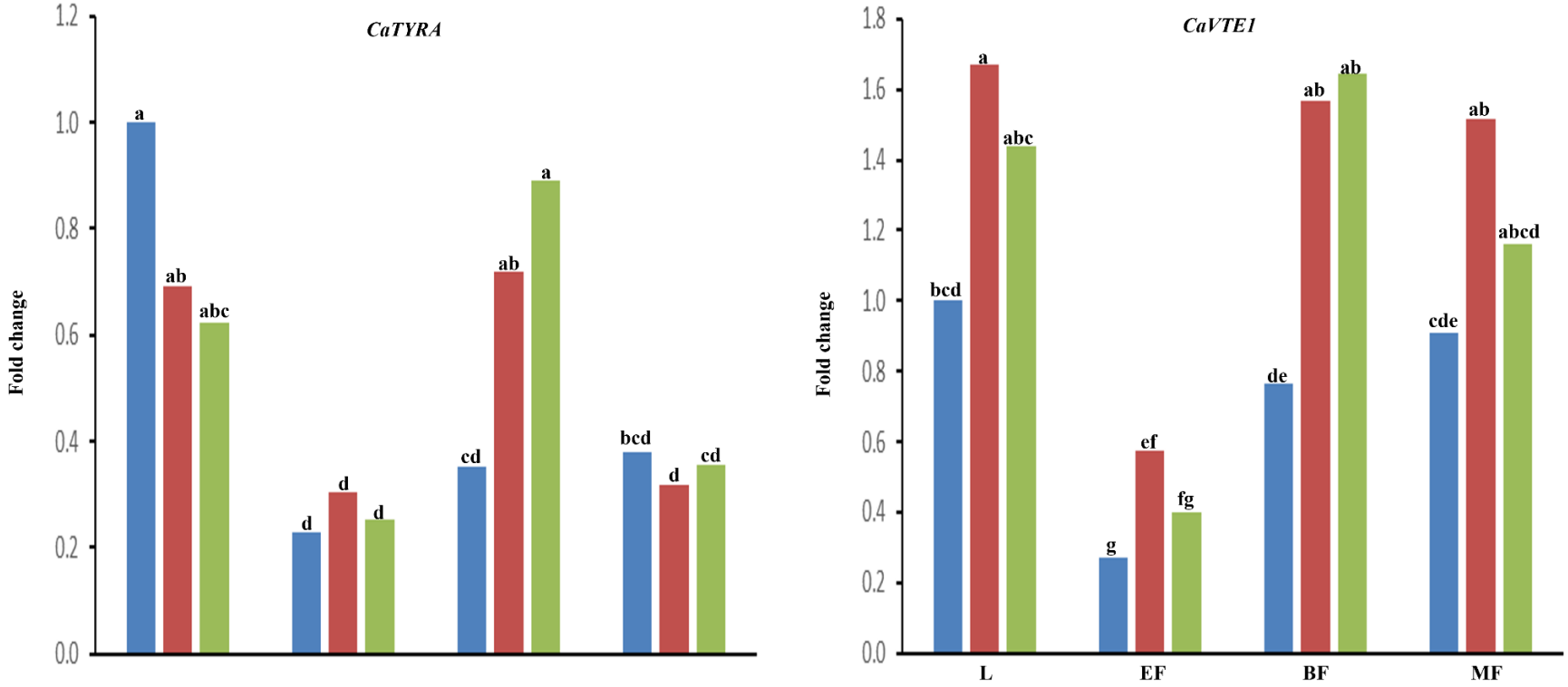


Fig. 3.14(C) Expression of genes involved (*CaTYRA* and *CaVTE1*) in Vitamin E biosynthetic pathway in four tissues - leaf (L), early fruit (EF), breaker fruit (BF), and mature fruit (MF) in *C. annuum* (blue color), *C. chinense* (red color), and *C frutescens* (green color). For each gene, samples with the same letter (a to h) refer to the difference between the samples not statistically significant; and samples with different letters refer to the significant difference ($p < 0.05$) estimated using Tukey's HSD test (Dubey et al. 2025).

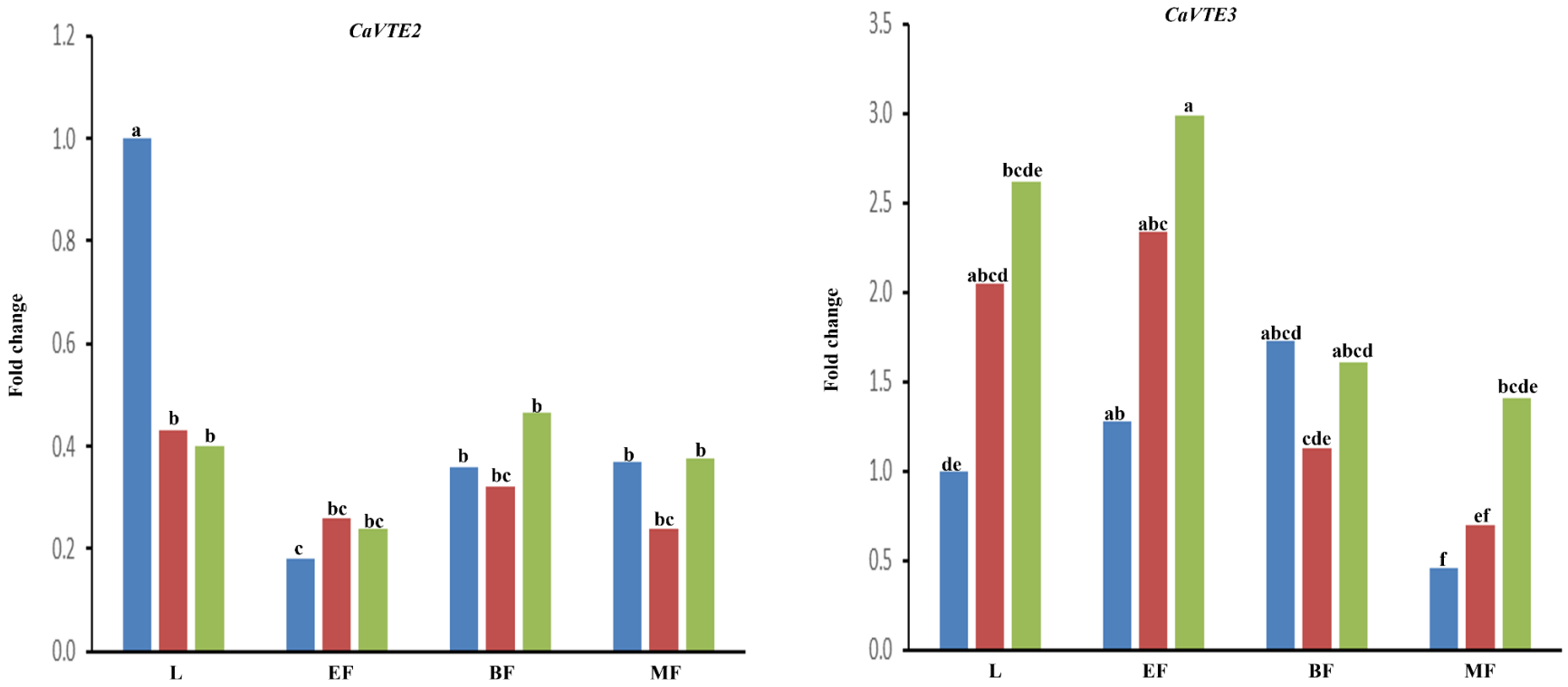


Fig. 3.14(D) Expression of genes involved (*CaVTE2* and *CaVTE3*) in Vitamin E biosynthetic pathway in four tissues - leaf (L), early fruit (EF), breaker fruit (BF), and mature fruit (MF) in *C. annuum* (blue color), *C. chinense* (red color), and *C frutescens* (green color). For each gene, samples with the same letter (a to h) refer to the difference between the samples not statistically significant; and samples with different letters refer to the significant difference ($p < 0.05$) estimated using Tukey's HSD test (Dubey et al. 2025).

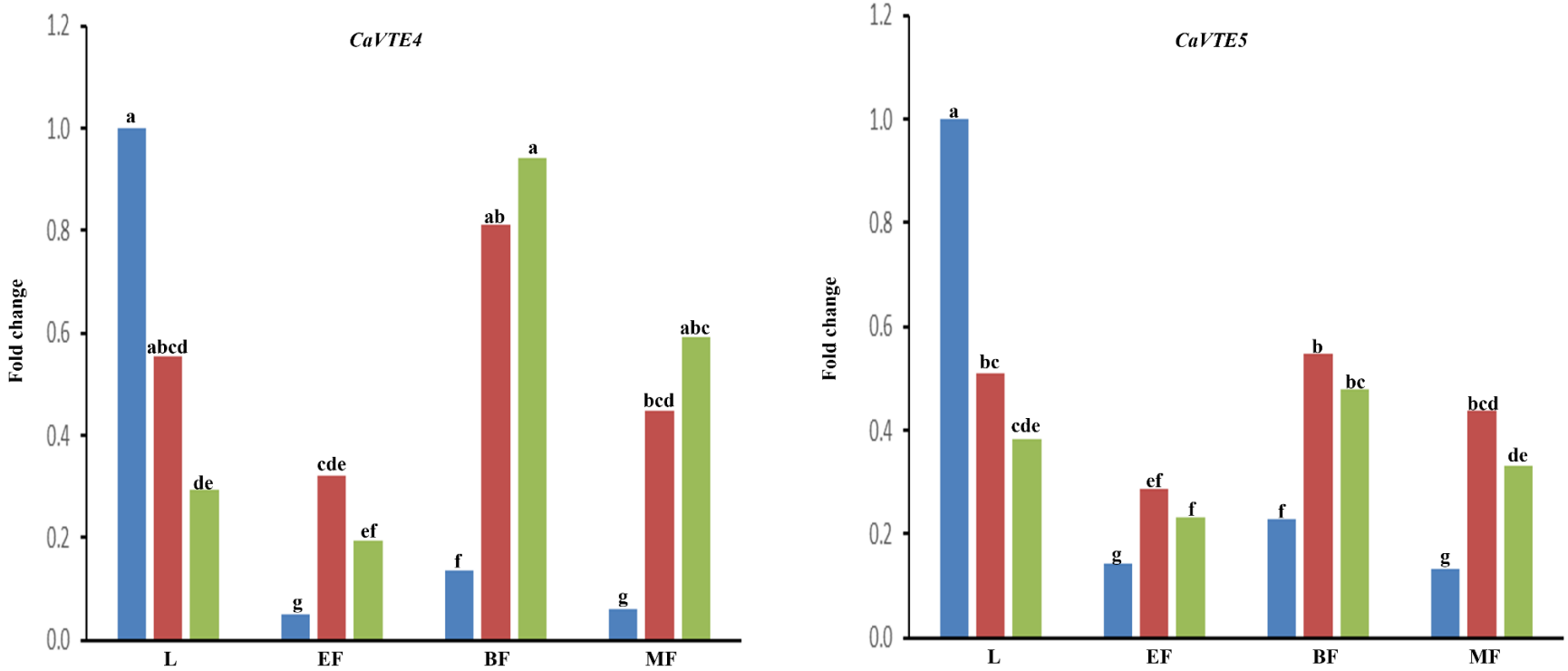
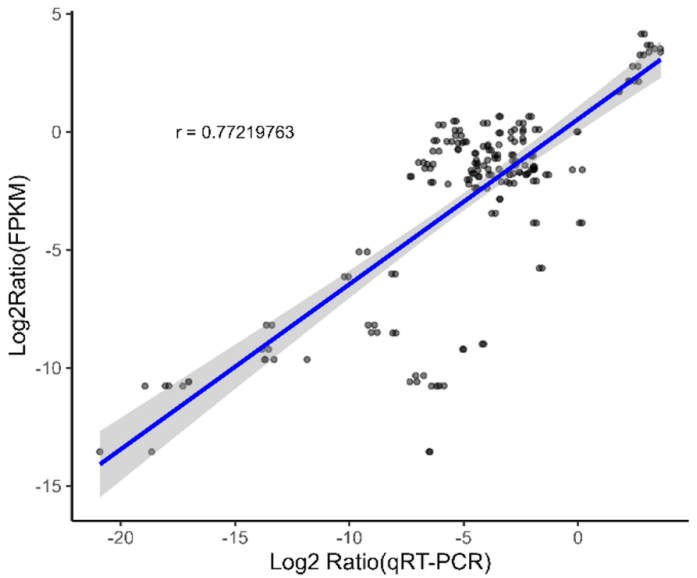


Fig. 3.14(E) Expression of genes involved (*CaVTE4* and *CaVTE5*) in Vitamin E biosynthetic pathway in four tissues - leaf (L), early fruit (EF), breaker fruit (BF), and mature fruit (MF) in *C. annuum* (blue color), *C. chinense* (red color), and *C frutescens* (green color). For each gene, samples with the same letter (a to h) refer to the difference between the samples not statistically significant; and samples with different letters refer to the significant difference ($p < 0.05$) estimated using Tukey's HSD test (Dubey et al. 2025).

CA	CC	CF	
0.71**	1.00**	0.72**	*CaGGDR*
0.25	-0.37	0.44	*CaHPPD*
0.32	0.85**	-0.66*	*CaIPI*
-0.95**	-0.81**	-0.51	*CaTAT*
0.31	0.40	0.21	*CaTYRA*
-0.18	0.24	0.45	*CaVTE1*
0.31	0.85**	0.33	*CaVTE2*
0.84**	0.60*	0.17	*CaVTE3*
0.45	-0.12	-0.30	*CaVTE4*
0.47	0.16	0.24	*CaVTE5*

(a)



(b)

Fig. 3.14(F) (a) Heat map in red-yellow-green color showing correlation of gene expression and Vitamin E content in *C. annuum* (CA), *C. chinense* (CC), and *C. frutescens* (CF) genotypes. * and ** showed significance at 0.05 and 0.01 levels, respectively. **(b)** The scatter plot shows the Pearson correlation between qRT-PCR data and transcriptome data for 10 Vitamin E genes (Dubey et al. 2025).

3.3.4 *In-silico* development of gene-based SSR markers for molecular breeding

For the utilization of the identified Vitamin C and E genes in the *Capsicum* breeding program, we identified a total of 185 SSR markers- 51 in Vitamin C genes and 134 in Vitamin E genes, respectively **[Supplementary Table 2(A) & 2(B) and Table 3.9]**. Among 51 SSRs, 34 were di-, 15 were tri- and two were pentanucleotide repeats, respectively, and the maximum number of repeats of the SSR motif was 25 **(Fig. 3.15)**. Out of 51 SSRs, 31 SSRs were present within the sequences of 22 Vitamin C genes. One gene (*CaGME*) contained three SSRs. The remaining 20 SSRs were present in the upstream sequences of the Vitamin C genes, respectively. Among 134 SSRs identified for Vitamin E biosynthesis genes, 92, 40 and one were di-, tri- and tetra nucleotide repeats, respectively. The maximum number of motif repeats was 31**(Fig. 3.15)**. Hundred and thirty-four SSRs were present within the gene and upstream sequences of the Vitamin E genes, respectively. Among the total SSRs identified, 24 SSRs were polymorphic among the three *Capsicum* spp.- *C. chinense*, *C. baccatum,* and *C. annuum* in the *in-silico* analysis. The primer sequences for amplification of the SSR motifs were designed using WebSat and the product sizes of the amplicons have been mentioned in **Table 3.**

Table 3.9 I*n silico* polymorphic SSR markers observed in Vitamin C and Vitamin E biosynthesis genes among *C. annuum, C. chinense,* and *C. baccatum* (Dubey et al. 2025).

Vitamin C									
S. No.	**Gene Name**	**Gene ID**			**Primer**	**SSR Repeats / Amplicon Size**			
		C. chinense	***C. baccatum***	***C. annuum***		***C. chinense***	***C. baccatum***	***C. annuum***	**Region**
1	*CYT*	*BC332_01951*		*LOC107842480*	FP-AAAAGGAATACGACGCTAACCA RP-ACTTGTGAAGAAGAGGAATCGG	(TC)6/211		(TC)15/239	Genic
2	*GR*	*BC332_07810*	*CQW23_10560*	*LOC107863849*	FP-CATGAGAACCCCTACAAGGAAG RP-TCACTGGAAGAAAAGAGCATGA	(TTC)7/382	(TTC)6/379	(TTC)6/379	Genic
3	*GPI*	*BC332_11605*	*CQW23_13083*	*LOC107867385*	FP-GGGTGAATAATGAGGCAACTTC RP-TAAGTGGAGGAGGGTGTTAGGA	(TC)6/367	(TC)6/367	(TC)7/369	Genic
4	*DHAR*	*BC332_14256*	*CQW23_01191*	*LOC107871328*	FP-CATGAGGTTGGTTTCGATTT RP-ACACACAGGGACTAAAAGAACAG	(TG)7/393	(TG)6/391	(TG)6/391	Genic
5	*GME*	*BC332_01951*	*CQW23_33222*		FP-AAAAGGAATACGACGCTAACCA RP-GAAGAAGAGGAATCGGGAAAA	(TC)9/221	(TC)11/225		upstream
6	*GMP*	*BC332_06972*	*CQW23_13083*		FP-ATGTTGGCCCGTGTGTACATGC RP-GGGAGTTTTGTGTTGTGTTGTG	(TA)8/119	(TA)6/115		upstream
7	*DHAR*		*CQW23_20242*	*LOC107871328*	FP-AGGCGGCTAACCCATAGACA RP-GGACGGAGGGAGTAGAACAGTA		(TA)10/385	(TA)16/397	upstream
8	*GME*		*CQW23_20243*	*LOC107879347*	FP-TTGCCGTGTAGCTTCTTGTCTA RP-GTGTACGGGTAAAAGGTTAGAATTG		(TA)7/387	(TA)9/391	upstream
Vitamin E									

S. No.	Gene Name	Gene ID			Primer	SSR Repeats/Amplicon Size			
		C. chinense	*C. baccatum*	*C. annuum*		*C. chinense*	*C. baccatum*	*C. annuum*	Region
1	*TIP41*	*BC332_22346*	*CQW23_20593*	*LOC107840209*	FP-GCTTTGGTTTCTTCTGTTTGGT RP-AACTTCACCATGTCAACACGTC	(TTG)6/274	(TTG)7/277	(TTG)8/280	Genic
2	*SDH/DHQ(1)*		*CQW23_01193*	*LOC107842508*	FP-CACCATCCTCCTTCTGTCTCTC RP-AAACCCCAGTAAATTCAACCCT		(TC)7/124	(TC)6/122	Genic
3	*SDH/DHQ(1)*		*CQW23_01193*	*LOC107842508*	FP-CTTCGGAGGTAGGTATGGACTG RP-GGAGTGGGAGAACAGACAAAAG		(TTG)7/276	(TTG)6/273	Genic
4	*TAT(1)*	*BC332_24974*	*CQW23_23221*	*LOC107845345*	FP-AGGTCAGCGCACACTCTATTCT RP-CCATATACTTCGTCAGCGATCA	(TGT)6/363	(TGT)10/374	(TGT)7/365	Genic
5	*VTE2*	*BC332_18733*	*CQW23_16910*	*LOC107850218*	FP-TCGGAAACAGCCTTTCTACTTC RP-CACAACTCCAACTCCAATTTCA	(TTG)6/369	(TTG)8/375	(TTG)6/369	Genic
6	*VTE2*	*BC332_18733*		*LOC107850218*	FP-ATGCTTGTCTTGGCACAGTTTA RP-TGGGTAGGAGTGAAGTAAGGGA	(AT)9/368		(AT)6/362	Genic
7	*CS(1)*	*BC332_12764*		*LOC107870324*	FP-TACGAGTGAGAGGGGATTCTGT RP-TCTGTTGTGACTTGGGAGTTTG	(CT)13/361		(CT)12/359	Genic
8	*HPPD(1)*	*BC332_13446*	*CQW23_12407*	*LOC107870882*	FP-CACCAAGTGGGGTATGAGG RP-GGGAAAGGGAGTGAGAAGAAG	(TA)6/368	(TA)7/370	(TA)6/368	Genic
9	*HPPD(1)*	*BC332_13446*		*LOC107870882*	FP-CACCAAGTGGGGTATGAGGTAT RP-GGAAAGGGAGTGAGAAGAAGATG	(TAT)6/379		(TAT)10/367	Genic
10	*HPPD(1)*	*BC332_13446*	*CQW23_12407*	*LOC107870882*	FP-ATGACGTAGGGCGTGTGTACTT RP-TCTTGAACTGGAGGTCTATCGG	(AT)7/328	(AT)6/326	(AT)7/328	Genic
11					FP-ATTATAGTTGCCGCTTGGTGAT		(TCT)11/329		Genic

		BC332_01232	*CQW23_01496*		RP-GGTCTCTCTCTCATCCTTGGAA	(TCT)7/338			
12	*HST*	*BC332_07849*		*LOC107863628*	FP-ACGAAAGACATGCCGTCAAT RP-CACTAATCAAAAGAGATCTAATTTTGGG	(AT)10/368		(AT)16/380	Upstream
13		*BC332_10449*	*CQW23_09740*		FP-TCTAATACATCCCACCAAACGA RP-TTCCTTCTTAGTCCGTTCCAAA	(TG)11/313	(TG)7/305		Upstream
14	*GGPS(2)*	*BC332_12065*	*CQW23_10916*	*LOC107867046*	FP-GCAAAGCACGTACCCTAAACTA RP-AGTACTGTCGTGGGTGGTG	(TA)12/405	(TA)6/391	(TA)7/393	Upstream
15	*HPPD(1)*		*CQW23_12407*	*LOC107870882*	FP-GATATTCGGGATTGACATTGGT RP-GCAGGAAATTAGCAAGGTGTTC		(AT)7/397	(AT)8/399	Upstream
16		*BC332_17116*	*CQW23_15958*		FP-TGTGTCCCTTATATTTGTCCCC RP-GCAAGAGGAAGCGTCAACA	(AT)10/400	(AT)6/317		Upstream

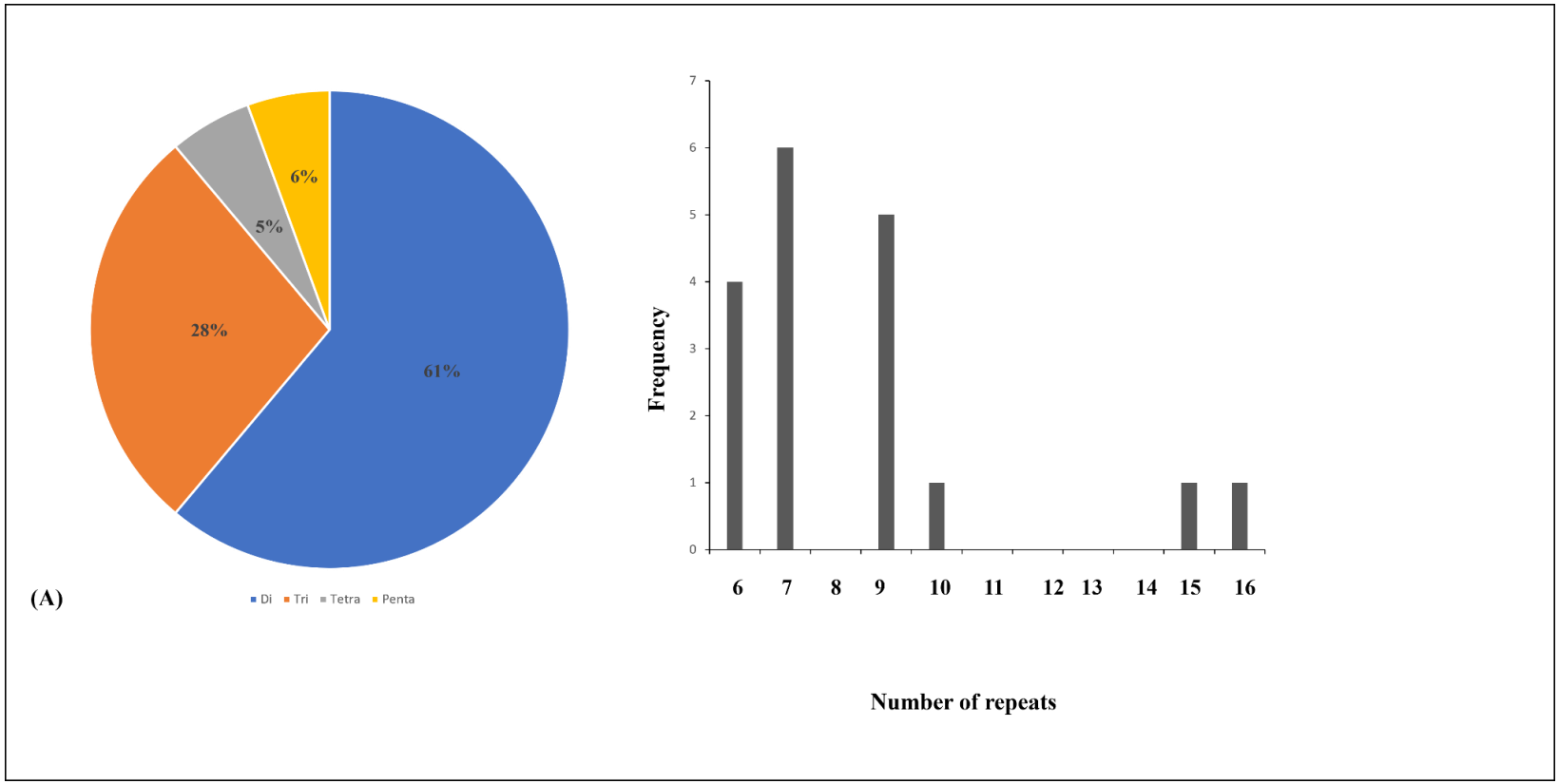


Fig. 3.15(A) Details Abundance, Classification, and Distribution of SSR motifs developed from genes involved in Vitamin C biosynthetic pathways in *C. annuum (*Dubey et al. 2025).

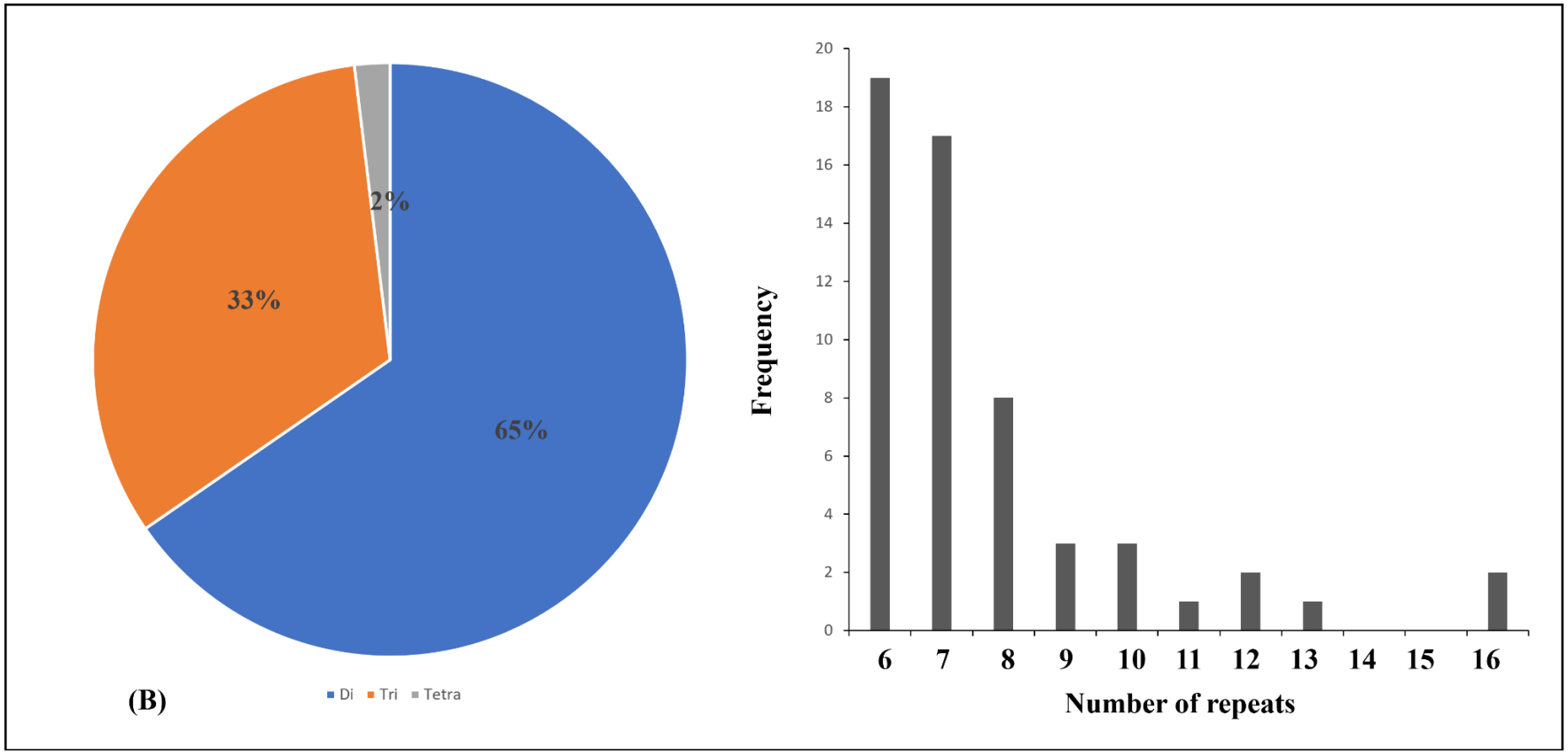


Fig. 3.15(B) Details Abundance, Classification, and Distribution of SSR motifs developed from genes involved in Vitamin E biosynthetic pathways in *C. annuum (*Dubey et al. 2025).

3.3.5 *In-silico* identification and development of Vitamin C and E genes based InDel markers

A total of 37 polymorphic InDel markers were identified within the Vitamin C and E biosynthetic genes and the upstream promoter regions, using the in-silico multiple sequence alignment. In the Vitamin C genes, 8 polymorphic InDel markers were identified (3 non-genic and 5 genic). The average length of InDel markers was around 45.12 bp, and the longest InDel was 75 bp as discovered in the *MIOX* gene. Based on the discovered polymorphism between *C. annuum*, *C. chinense*, and *C. baccatum* for Vitamin C biosynthesis genes in this study, these InDel markers can be validated via PCR using InDel specific primers **(Table 3.10)**. For Vitamin E genes, 29 InDel markers were identified (14 non-genic and 15 genic). The average length of InDel markers was around 59.82 bp, and the longest InDel was 156 bp as discovered in the *CM(2)* gene. The gene-specific InDel primers were designed manually, and the amplicon sizes were mentioned in three *Capsicum* species for each InDel marker in **Table 3.10.**

Table 3.10 Polymorphic Insertion-deletions (InDels) markers identified in the Vitamin C and E genes among the three *Capsicum* species-*C. annuum, C. chinense,* and *C. baccatum* (Dubey et al. 2025).

Vitamin C Genes-Upstream									
S. No.	**Gene Name**	***C. chinense* Gene ID**	***C. baccatum* Gene ID**	***C. annuum* Gene ID**	**InDel Size**	**Primer**	**Product Size**		
							C. chinense	***C. baccatum***	***C. annuum***
1	*GalDH*	*BC332_22176*	*CQW23_20548*	*LOC107840121*	43	FP-CATAGGTAACTCTCAATACAGCCCAAT RP-TGTGACACATTCTTTATAGGATGGACAA	245	202	245
2	*MIOX*	*BC332_24650*	*CQW23_06600*	*LOC107840943*	74	FP-ATAAAGTTGCTCCTCTAGTCGCAAC RP-TGCTTTCGACGTATTTTTGACCATCA	274	274	200
3	*GPI*	*BC332_11605*	*CQW23_10560*	*LOC107867385*	47	FP-GGAATCCTGAGGTAAAATTGGGGAAG RP-ACTCACATTTCGCTTTCATCTGATC	260	260	307
Vitamin C Genes									
1	*MIOX*	*BC332_24650*	*CQW23_06600*	*LOC107840943*	49	FP-GAGCAGCCTGAGTTTGGTATGTC RP-GATGCGCTCAATCATTCGTGTAAC	346	346	297
2	*AKR2*	*BC332_06244*	*CQW23_31568*	*LOC107864478*	49	FP-GCCCATTGAAATCACGGTGAG RP-CCAGCTTCTTAAGTTCTCCAACC	290	290	241
3	*MDHAR1*	*BC332_00155*	*CQW23_00039*	*LOC107864529*	34	FP-CGAGGAGTTCGATTCCTTCTAGAAG RP-TCGAAATCAAACACGAGAGCATC	304	338	338
4	*AKR1*	*BC332_15059*	*CQW23_22478*	*LOC107865540*	32	FP-GGAAGCATTCAACAAAGGCATCAC RP-CTTTCTTCACAGCATTGCCGC	399	399	431
5	*GPI*	*BC332_11605*	*CQW23_10560*	*LOC107867385*	33	FP-GGGCCATATTCTGTTCTTACTTCAAGAG RP-GATCCAACTCCATCAGTTGCCC	193	160	160

Vitamin E Genes-Upstream									
1	*CMS*	*BC332_22023*	*CQW23_33378*	*LOC107839798*	90	FP-TCTGATACCACTTGTTAGGAAAGACGG RP-GTGATTATAGTGGAGGTTTGGTCCC	272	272	182
2	*TIP41*	*BC332_22346*	*CQW23_20593*	*LOC107840209*	53	FP-GCGAACGTGATGGCAAGAAAG RP-TTTACAGAGAGGTACTTTGTGGCGG	253	253	200
3	*HDR*	*BC332_22432*	*CQW23_20958*	*LOC107840382*	21	FP-TTTACAGAGAGGTACTTTGTGGCGG RP-GTGCTACTGTTTCCTCCGACAGA	226	226	247
4	*TYRA*	*BC332_23589*	*CQW23_21512*	*LOC107842110*	60	FP-CCTTCATTGTGATTCACCTCTCAAGA RP-GGTAGTTGTAGTAATGATCCAACCCAG	324	384	384
5	*SDH/DHQ(1)*	*BC332_01953*	*CQW23_01193*	*LOC107842508*	25	FP-TGACCGAATTAAATGGTCTCAAATACTC RP-CCTAGCTTAAATTACTCCCTCCGC	362	387	387
6	*DXS(2)*	*BC332_29696*	*CQW23_27278*	*LOC107850768*	76	FP-TATATCCTTGTGGTGGAGTCTTTCC RP-ATTATGTTAACTTGACGCTTTTTCATGAC	465	389	389
7	*DXS(1)*	*BC332_01962*	*CQW23_01208*	*LOC107852315*	40	FP-AAGATTTTCTACAAAAACCTATGCATAGC RP-GTAAGTGGTGTCACTTTCGATTGACTC	362	362	402
8	*DHQS*	*BC332_05260*	*CQW23_04953*	*LOC107859855*	31	FP-ATTGACACTCATGCAAGGATAAAGTGA RP-GATCCAGCTAAATTTTTAAGTGAAGCTCC	122	122	91
9	*PDS*	*BC332_09881*	*CQW23_11317*	*LOC107861625*	90	FP-CAGGCATTTCCACCCAAATACCA RP-AATTAAATGTTGTTTAATACTACATCGCC	484	357	394
10	*GPPS*	*BC332_32257*	*CQW23_32033*	*LOC107866981*	37	FP-CAGAAATGAACCTCTGCTCACAC RP-CCTAAACGGACAAGGTAAACACTTATCC	261	335	298
11	*GGPS(1)*	*BC332_29497*	*CQW23_27298*	*LOC107870481*	82	FP-AACTATGGACTATTGTGAATTGTGTTG RP-TCGCTAGAAGTCATCGGTTCTAC	309	227	227
12	*HPPD(1)*	*BC332_13446*	*CQW23_12407*	*LOC107870882*	55	FP-CCCTAATCATGATAGTCGCTACTGGTAG RP-AGTGAGGGATTTTGGTAGTGGAATTAG	250	305	305
13	*SDH/DHQ(2)*	*BC332_17702*	*CQW23_15682*	*LOC107875208*	32	FP-AAATTTACGTGGAAATCCTTGCGG RP-CCGAATTTTACTCATCCGGCAAAG	282	282	314

14	*TAT(2)*	*BC332_19318*	*CQW23_17506*	*LOC107878342*	61	FP-CACCAATTCGAGGCATCACATT RP-ACGGGTGTATTAGGTTGCCTG	177	238	238
Vitamin E Genes									
1	*VTE1*	*BC332_21357*	*CQW23_19464*	*LOC107839135*	105	FP-GACGGAGGCAATTCACCTGGTA RP-GCCGTCCTCACCTGAGTT	313	208	208
2	*HDR*	*BC332_22432*	*CQW23_20958*	*LOC107840382*	91	FP-GTGACATCTGGTGCATCCACTC RP-CAGAGGGAGGAACGGAAAGAGG	275	275	184
3	*VTE2*	*BC332_18733*	*CQW23_16910*	*LOC107850218*	29	FP-TACCGTCGGAGGTAGCGG RP-CTGGACACTCTATGCGATTCAATCA	286	257	257
4	*PAT*	*BC332_30552*	*CQW23_28430*	*LOC107851006*	62	FP-TCTTAAAATCCCTCAACAACTCGACC RP-ACAATCTTCTCAGAGCTAGGCAC	301	301	363
5	*DXS(1)*	*BC332_01962*	*CQW23_01208*	*LOC107852315*	49	FP-AAGTTGTCTCCTCAAACGTGGAC RP-GATCTCTTCCAACAGCCATCCC	325	325	374
6	*CM(1)*	*BC332_05989*	*CQW23_30817*	*LOC107860495*	35	FP-GGTCTGGCCAGTATATGCGTAC RP-GGCTTGTAAGAGTGTTAACTGGACAC	181	181	216
7	*HST*	*BC332_07849*	*CQW23_07182*	*LOC107863628*	50	FP-CGTTGACACGAACACAAGGC RP-ACGACTTTATGTTTTCCTCCCAGTTAC	288	288	238
8	*CAC*	*BC332_12234*	*CQW23_11064*	*LOC107866788*	29	FP-TGGGAAATTGTTGGCCAAAGC RP-CAGTGAAGGGTTACTCCACCC	331	360	360
9	*GPPS*	*BC332_32257*	*CQW23_32033*	*LOC107866981*	35	FP-CACCTCAACCGCTGGTCTAAG RP-CTCGACGACAACTTCTCTCCTC	200	235	235
10	*CM(2)*	*BC332_11102*	*CQW23_10170*	*LOC107867808*	156	FP-GGACAAGTAAAAGATAGACGGAGGAAG RP-TCTGCTGCAGAATGCAGAACC	483	483	327
11	*HPPD(1)*	*BC332_13446*	*CQW23_12407*	*LOC107870882*	117	FP-GAAGGCGTGTGAAGGTGTAATG RP-AGAAATCACGAAATTTGGGGTTCAAC	608	608	491
12	*EPSPS*	*BC332_01232*	*CQW23_01496*	*LOC107873029*	50	FP-GGGAGGGTGGTGTGTACAC RP-GTGCGGAGATCAAGTTGAAGTAGTTG	271	221	221
13	*FPGS*	*BC332_17116*	*CQW23_15958*	*LOC107875563*	32	FP-CGTGGATAGTACGTGGAAGACGAC RP-CCCCGATTTTCTCCTCTAATGTCGTA	260	292	292

14	*EXPRESSED*	*BC332_18515*	*CQW23_17046*	*LOC107878010*	96	FP-GTGAACAGTAGCAAGTGGTGACG RP-TCAAGGAGCCTTGTTTGGTGG	208	208	112
15	*TAT(2)*	*BC332_19318*	*CQW23_17506*	*LOC107878342*	46	FP-GCGGAGCTACCCTTTGCC RP-ATAGGAAGTCTTGCAATCTTCGATCTTG	395	441	441

3.3.6 Synteny and Gene Duplication Analysis

The synteny and collinearity analysis among four *Solanaceae* genomes: *C. annuum, C. chinense, C. frutescens,* and *S. lycopersicum* **[Fig. 3.16 (A), (B), and (C)]**, identified a total of 370 CSSs with vitamin biosynthesis genes in them. The CSSs with vitamin genes ranged from 0.04 to 10.98 Mbp (million base pairs), and CSSs with Vitamin E genes ranged from 0.05 to 60.41 Mbp in size, respectively **(Table 3.11)**. A total of 124 and 246 CSSs harboring Vitamin C and Vitamin E genes, respectively, were identified among the four *Solanaceae* genomes (*C. annuum, C. chinense, C. baccatum,* and *S. lycopersicum*). In the syntenic blocks, around 16, 25, 29, and 15 Vitamin C genes, and 40, 50, 51, and 39 Vitamin E genes were present in the *C. annuum, C. baccatum, C. chinense,* and *S. lycopersicum* genomes, respectively **(Table 3.11)**.Of these syntenic blocks, a total of 69 and 172 syntenic blocks harbored collinear Vitamin C and Vitamin E genes. The highest number of CSSs for Vitamin C was 28, shared between *C. chinense* and *S. lycopersicum*. Out of the 124 CSSs with Vitamin C genes, 33 were present on different chromosomes, 7 on scaffolds, and 84 were present on the same chromosome among different genomes. For Vitamin E, the highest number of CSSs was found between the *C. chinense* and *S. lycopersicum* genomes, i.e., 51. Out of 246 CSSs with Vitamin E genes, 82 were present on different chromosomes, 20 on scaffolds, and 144 were on the same chromosomes among the different genomes. Overall, most of the CSSs were present on the same chromosomes in the different genomes **[Fig. 3.16 (A)].**

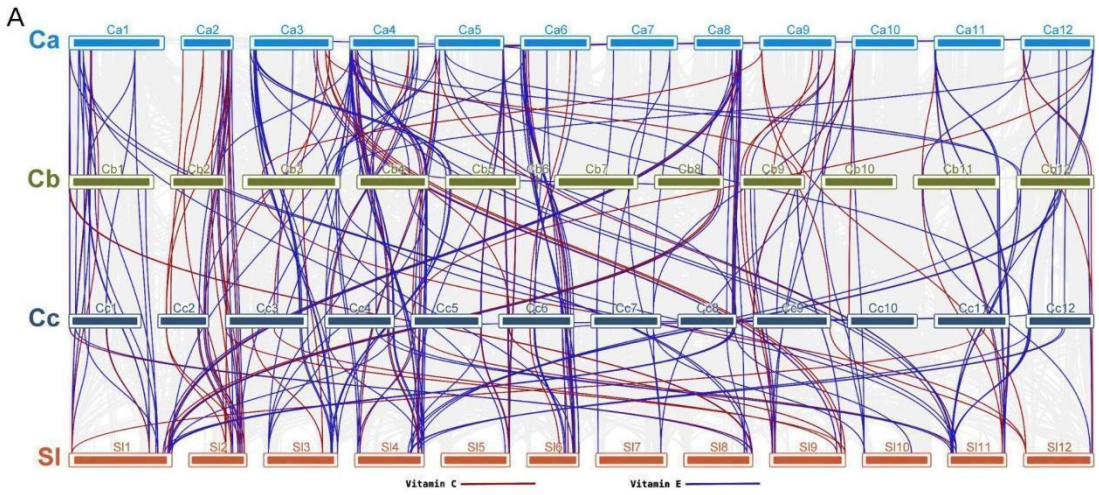


Fig.. 3.16(A) Conserved syntenic segments observed for Vitamin C and E biosynthesis genes among *C. annuum* (Ca), *C. chinense* (Cc), *C. baccatum* (Cb), and *S. lycopersicum* (Sl) genomes. Red and purple color indicates Vitamin C and Vitamin E biosynthesis genes, respectively (Dubey et al. 2025).

To study the effect of duplication events on the expansion of the Vitamin C and Vitamin E genes, homologous genes for vitamin biosynthesis were identified in *C. annuum*, *C. baccatum*, *C. chinense*, and *S. lycopersicum*. We found 700 pairs of vitamin gene duplicates or homologs across the genomes of four species: 240 homolog gene pairs for Vitamin C genes and 460 homolog gene pairs for Vitamin E genes **[Fig. 3.16 (B) and (C); Supplementary Table 3 and Supplementary Table 4]**.Among the 70 homolog gene pairs between *C. annuum* and *S. lycopersicum*, 21 were for Vitamin C, 9 out of 21 were present on the same chromosomes, and 49 were for Vitamin E, 28 out of 49 were present on the same chromosomes, respectively. A total of 129 homolog gene pairs were found between *C. annuum* and *C. baccatum*: 45 for Vitamin C, 15 out of 45 gene pairs were present on the same chromosomes, and 84 gene pairs for Vitamin E, 37 out of 84 were present on the same chromosomes, respectively, Among the *C. annuum* and *S.*

lycopersicum genomes, a total of 128 homolog gene pairs were found: 45 gene pairs for Vitamin C, 19 out of 45 were present on the same chromosomes, and 83 gene pairs for Vitamin E, 27 out of 83 were present on the same chromosomes, respectively. Similarly, a total of 161 homolog gene pairs were found between *C. baccatum* and *C. chinense* genomes: 57 gene pairs for Vitamin C, 30 out of 57 were present on the same chromosomes, and 104 gene pairs for Vitamin E, 55 out of 104 were present on the same chromosomes, respectively. For the *C. chinense* and *C. annuum* genomes, 106 homolog pairs were found: 36 pairs for Vitamin C, 20 out of 36 were present on the same chromosomes, and 70 gene pairs for Vitamin E, 53 out of 70 were present on the same chromosomes, respectively. Finally, a total of 106 homolog gene pairs were found between *C. chinense* and, *S. lycopersicum* genomes: 36 gene pairs for Vitamin C, 14 out of 36 were present on the same chromosomes, and 70 homolog gene pairs for Vitamin E, 39 out of 70 were present on the same chromosomes, respectively **[Fig. 3.16 (B) and (C); Supplementary Table 3 and Supplementary 4 Table 6]**

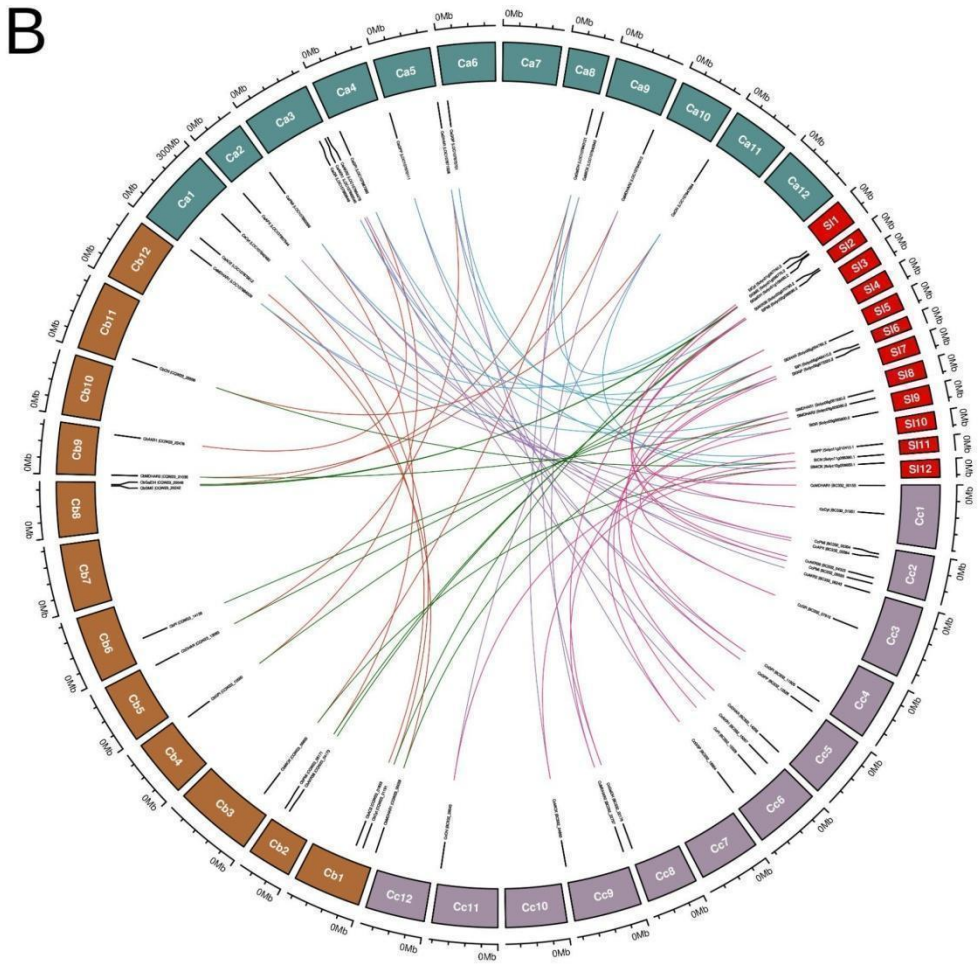


Fig.. 3.16 (B). The conservation of Vitamin C biosynthesis genes showing collinearity among *C. annuum* (Ca), *C. chinense* (Cc), *C. baccatum* (Cb), and *S. lycopersicum* (Sl) genomes (Dubey et al. 2025).

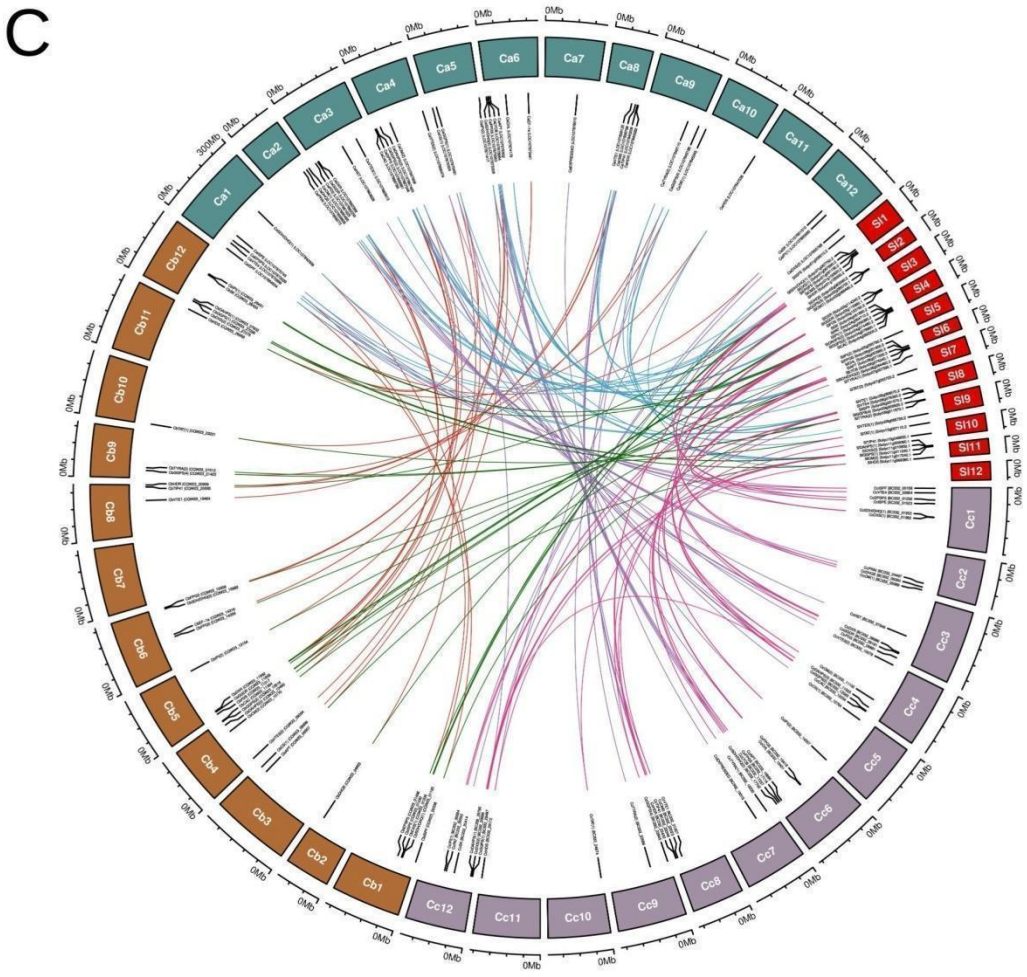


Fig. 3.16 (C). The conservation of Vitamin E biosynthesis genes showing collinearity among *C. annuum* (Ca), *C. chinense* (Cc), *C. baccatum* (Cb), and *S. lycopersicum* (Sl) genomes (Dubey et al. 2025).

Table 3.11. Number of conserved syntenic segments (CSSs) containing Vitamin C and E biosynthetic genes identified in *C. chinense, C. baccatum, C. annuum*, and, *S. lycopersicum* genomes (Dubey et al. 2025).

Number of CSSs for Vitamin C genes						Number of CSSs for Vitamin E genes					
S. No.	**Between Genome**	**No. of CSSs**	**Same Order**	**Reverse Order**	**Size Range**	**S. No.**	**Between Genome**	**No. of CSSs**	**Same Order**	**Reverse Order**	**Size Range**
1	*C. annuum* and *C. baccatum*	25	14	10	0.08 to 9.09 Mbp	1	*C. annuum* and *C. baccatum*	50	33	17	0.07 to 9.62 Mbp
2	*C. annuum* and *C. chinense*	25	11	14	0.10 to 9.55 Mbp	2	*C. annuum* and *C. chinense*	49	22	27	0.26 to 13.67 Mbp
3	*C. annuum* and *S. lycopersicum*	22	14	8	0.06 to 10.98 Mbp	3	*C. annuum* and *S. lycopersicum*	43	21	22	0.06 to 60.41Mbp
4	*C. baccatum* and *S. lycopersicum*	23	9	14	0.05 to 1.44 Mbp	4	*C. baccatum* and *S.lycopersicum*	49	26	24	0.05 to 1.16 Mbp
5	*C. chinense* and *S. lycopersicum*	28	17	11	0.04 to 1.66 Mbp	5	*C. chinense* and *S. lycopersicum*	51	19	22	0.07 to 1.77 Mbp
Total		123	65	57		Total		242	121	112	
Total Percentage of same order			52.80			Total Percentage of same order			46.50		

3.4 DISCUSSION

3.4.1 OBJECTIVE 1: Identification and expression analysis of genes involved in *Capsicum* fruit development and ripening:

In *Capsicum*, our understanding of the genetic mechanisms governing fruit development and ripening remains quite limited, with only a handful of genes having been characterized so far in earlier reports. In contrast, the tomato, a close relative within the Solanaceae family, has been the subject of extensive research. In tomatoes, the molecular and biosynthetic pathways and the specific genes involved in fruit development and ripening are well documented and thoroughly characterized.

Despite the main economic harvest of *Capsicum* being the fruit, the genes involved in fruit development/ripening have not been well characterized although available study reports are there in its close relative, tomato (Kim et al., 2014). Therefore, in the present study, we explored the *Capsicum* genome for identification of fruit development/ripening genes through comparative genomics. For this purpose, we used a total of 32 fruit development/ripening genes which are involved both in ethylene-dependent and independent fruit ripening pathways in tomato **(Fig. 3.1)** and could identify a total of 42 orthologous genes in the C. annuum reference genome. These genes are involved in different fruit development/ripening traits including coloring, cell wall formation, softening, aroma, and metabolite accumulation etc. **(details are given in Table 3.1)**. Significantly, one or more orthologs were identified for each of the 32 tomato genes in *Capsicum*, which suggest that the large size of the *Capsicum* genome may harbor more than one ortholog of tomato genes **(Table 3.2)**, Furthermore, of the 32 genes, 18 were

mapped onto the homologous *Capsicum* chromosomes suggesting these genes are still conserved in the syntenic chromosomal regions. Although *Capsicum* is the closest relative to tomato, the regulatory mechanisms (including gene expression) of these two crops differed, and are perhaps responsible for their diversified features (Kim et al., 2014). Our in-silico expression analysis revealed that out of 32 genes, only 12 genes were differentially expressed in *Capsicum* during fruit development/ripening: indicating different ripening behavior of *Capsicum* than that of tomato. Similar results have also been reported where expression pattern of *colorless non ripening gene (CNR)*, *Golden-2-like gene* (transcription factor that regulates plastid and chlorophyll levels thus responsible for coloration during fruit ripening), *ACO, ACS* and *HB-1* were very distinct in tomato and *Capsicum* (Kim et al., 2014). *Capsicum* is typically classified as non-climacteric fruit (unlike tomato which is a climacteric fruit), and this may be the reason that out of the 12 differentially expressed genes, 11 (excluding Ethylene receptor like protein, *ETR*) belonged to the ethylene independent pathway. In the present study, *ETR* (*Capsicum* gene ID; *LOC107873245*) found to be up regulated during fruit maturity in only *C. frutescens* supported the climacteric behavior of hot pepper as suggested in earlier studies (Hou et al., 2018; Paul et al., 2012), reinforcing the idea of species-specific ripening regulation within the *Capsicum* genus (Tan et al., 2012).

Although key regulators of the ethylene independent fruit ripening pathway are still not fully known, but *LeMADSRIN (RIN),* a MADS-box transcription factor, has been considered as one of the major regulators (Giovannoni, 2004; Lin et al., 2009). In the present study, interestingly *RIN* (*Capsicum* gene ID; *LOC107847473*) unexpectedly up-

regulated (more than 500 folds higher expression as compared to leaf) in breaker and mature fruits in all the three spp.; suggesting that *RIN* is a potential regulator of ethylene independent fruit ripening pathway in non-climacteric fruit, a role that is gaining support in other non-climacteric fruits (Wang et al., 2022). Furthermore, interacting partner genes of *RIN, like FUL1 (LOC107855404), FUL2 (LOC107845304), TAGL1 (LOC107878477)* showed maximum expression during fruit maturation as expected, although level of expression differed among three spp.; homologs of *Stay green 1 (SGR1, Capsicum* gene ID; *LOC107866321*), target of *RIN* and involved in chlorophyll degradation (Wang et al., 2016), also found to be up regulated during breaker and mature fruits stages suggested involvement of SGR1 gene in non-climacteric fruit ripening. Unlike, a few genes showed different expression pattern across spp., for example, *Golden 2 like gene (GLK2, Capsicum* gene ID; *LOC107845460*), which is known for regulating plastid and chlorophyll levels (coloration) during fruit ripening was up-regulated in *C. annuum* in contrast to *C. chinense* and *C. frutescens* where it was down-regulated during fruit maturation, highlighting the complex transcriptional control of ripening traits (Villa-Rivera et al., 2022). Based on the above results, the present study suggested that fruit development/ripening pathway may vary somewhat across three spp., although most of the pathway may be similar across spp.

Moreover, to utilize the fruit development and ripening genes identified in the present study in the breeding program, we also developed 49 SSR markers which are present either within or nearby to genes. As expected, a lesser number of SSRs (only 14, out of 49 SSRs) were present within the genes as compared to nearby regions, and only six were

polymorphic (out of 14) suggesting the high conservation of genic regions. From earlier studies it was suggested that the high conservation is maintained among the genes in different species as any mutation in this region may lead to drastic change (Kim et al., 2008, Zhang et al., 2014). In plants, di and trinucleotides repeat motifs have been found most abundant and vary from species to species (Kumpatla et al., 2005; Saha et al.,2017.; Varshney et al., 2002). In our study, di-nucleotide (AT/TA) SSRs were of maximum frequency followed by trinucleotide SSRs. To see the utilization of SSRs developed in population genetics study, we genotyped 47 *Capsicum* genotypes using 42 SSRs. Out of 42 SSRs, 23 (~55%) SSRs were found to be polymorphic which was comparable to earlier studies (Al-Qadumii et al., 2012; Meng et al., 2017; Rai et al., 2013); and out of 23 SSRs, six including SSR_CF-1 to SSR_CF-5 and SSR_CF-11 were present within the gene orthologs of *TCTR1, GLK, TOMLOX, PG2* and *FUL2* **(Table 3.2 and Table 3.5)**, respectively. Unexpectedly, we observed higher PIC for genic SSRs (0.31 to 0.51) than that of nearby SSRs (0.22 to 0.37). Highly polymorphic genic SSRs may serve as perfect markers in fruit breeding in *Capsicum* for the selection of associated genes; although non-genic SSRs also prove as potential flag marks for genes since they are tightly linked to genes (within 5 Kb). Wide range of other diversity parameters like heterozygosity (0.27 - 0.59), effective number of allele (1.36 – 2.4), gene diversity (0.26 – 0.58), and diversity index (0.51 – 0.97) also suggested that SSRs developed in our present study may prove useful for population genetics study and genomic selection approaches (Barchenger et al., 2018).

The genes identified in the *Capsicum* genome may be further functionally characterized

through overexpression and silencing/editing (e.g., CRISPR-Cas9) to study the details of their molecular mechanism so that these genes can be used for the manipulation of fruit development and ripening traits in *Ca*psicum crop (Yang et al., 2025). *Capsicum* L. displays rich diversity in fruit traits like shape, color, size, and ripening behavior (A. Kumar et al., 2021; Sarpras et al., 2016). Further, the SSR markers developed should be correlated with phenotypic traits in a large collection of natural populations or biparental mapping populations developed from contrasting fruit traits. Once validated, both SSRs and genes could be used for marker-assisted breeding and genomic selection to improve fruit traits in *Capsicum* (Adeyemo et al., 2023).

3.4.2 OBJECTIVE 2: Determination of Vitamin C and Vitamin E content at different fruit developmental stages in contrasting *Capsicum g*enotypes.

Vitamins are essential food components and play a significant role in the maintenance of human health. *Capsicum* is considered as a good source of Vitamin C and E (Daood et al., 1996; Padayatty et al., 2003; Wahyuni et al., 2011.; Zhuang et al., 2012.). Among the major *Solanaceae* crops (including tomato and eggplant), *Capsicum* has the highest content of Vitamin C (Rosa-Martínez et al., 2021). Several studies have also reported higher Vitamin C content in *Capsicum* fruit as compared to orange (Olatunji et al., 2019; Wahyuni et al., 2011.; Yahia et al., 2001.), tiger-nut, and turmeric (Toyer et al., 2021). Although Vitamin C and E biosynthesis genes have been well-characterized in species like tomato, which is closely related to *Capsicum* (Almeida et al., 2011; Ioannidi et al., 2009; Mellidou & Kanellis, 2017; Rodrõ Âguez-Concepcio et al., 2001), no comprehensive study has been conducted for the identification of the total Vitamin C and E biosynthesis genes in *Capsicum spp.* Furthermore, there is no report of estimation of Vitamins C and E content and their correlation with expression of genes at different fruit development stages in Northeast India *Capsicum* germplasm. Therefore, in this study, we conducted an experiment for the first time to estimate the Vitamin C and E content in three fruit developmental stages in 30 genotypes belonging to *C. annuum, C. chinense,* and *C. frutescens* from Northeast India. Our analysis showed highly variable vitamin content across the different fruit tissues which ranged from 15.21 to 41.77 mg/g (Vitamin C) and 3.32 to 112.4 mg/g (Vitamin E), respectively **(Table 3.7; Fig. 3.9, and Fig.3.10)**. We observed that in most of the *Capsicum* genotypes, Vitamin C content was higher as compared to that reported in previous studies (Gomes et al., n.d.; Kouassi et al., 2012;

Mennella et al., 2017). For instance, in a recent report, the content of Vitamin C in the sweet variety of *Capsicum*, i.e., Claudius (228.49 µg/g), was lower as compared to our study (Lidiková et al., 2021). The variations in the Vitamin C content in the *Capsicum* genotypes used in our study may be due to the differences in genetic background and the influence of the environment under which plants were grown.

In case of *Capsicum*, it is suggested that Vitamin C content increases as fruit matures, and MF (red) has higher Vitamin C content as compared to EF (Alós et al., 2013; Ghasemnezhad et al., 2011; Howard et al., 2000; Khadi et al., 1987; Lidiková et al., 2021; Martínez-Ispizua et al., 2021; Martínez et al., 2005; Olatunji et al., 2019; Yahia et al., 2001)**.** Another study reported the high to low trend for Vitamin C content in immature to mature fruit, with the highest accumulation in ripe fruit, respectively (Chiaiese et al., 2021). In our present study, a similar pattern of Vitamin C content (higher in MF) was observed in most of the *C. annuum* and *C. frutescens* genotypes except in the accessions Ca-3, Ca-5, Ca-9, and Cf-8 (higher in EF) **(Fig. 3.9)**. However, in *C. chinense* genotypes, the Vitamin C content was higher either in EF (in 6 genotypes, namely Cc-1, Cc-2, Cc-4, Cc-5, Cc-8, and Cc-9) or MF (in 4 genotypes, namely Cc-3, Cc-6, Cc-7, and Cc-10). These discrepancies may be explained by the different genetic compositions of the different genotypes of *Capsicum.* Among the different species of *Capsicum*, *C. frutescens* is reported to have maximum Vitamin C content (Żurawik et al., 2021). However, in our study, *C. chinense* showed the highest Vitamin C content, followed by *C. annuum* and *C. frutescens* in all or most of the fruit developmental stages.

In a 2019 study, Olatunji and his colleagues investigated the nutritional potential of

peppers by measuring Vitamin E levels in four different genotypes of *Capsicum* from *C. annuum* (three genotypes) and *C. frutescens* (one genotype). Their research, comparing four genotypes, revealed a clear hierarchy, the common bell pepper (*C. annuum var. grossum*) led by a significant margin, followed by *C. annuum var. acuminatum*, the bird's eye chili (*C. frutescens var. baccatum*), and finally *C. annuum var. abbreviatum*. In our study, the content of Vitamin E was higher than previously reported and was the maximum in some *C. annuum* and *C. frutescens* genotypes, followed by *C. chinense* genotypes. On the other hand, Vitamin E content increased or decreased with fruit development in all three *Capsicum* species depending upon the genotypes **(Table 3.7; Fig. 3.10)**. However, in *C. chinense*, four genotypes (namely Cc-1, Cc-5, Cc-6, and Cc-8) showed highest Vitamin E content at MF stage, four genotypes (Cc-2, Cc-3, Cc-7, and Cc-9) in BF, and two genotypes (Cc-4 and Cc-10) in EF, respectively. In *C. frutescens,* four genotypes (Cf-2, Cf-4, Cf-9, and Cf-10) were observed to exhibit higher content of Vitamin E at MF stage, four genotypes (Cf-2, Cf-5, Cf-7, and Cf-10) in BF, and only two genotypes (Cf-4 and Cf-7) were observed for higher content at EF stage, respectively. Similarly, in *C. annuum,* four genotypes (Ca-1, Ca-2, Ca-6, and Ca-8) were observed for higher content of Vitamin E at MF stage, three genotypes (Ca-5, Ca-7, and Ca-10) at BF stage, and three genotypes (Ca-5, Ca-6, and Ca-8) were found for higher content of Vitamin E at EF stage, respectively. Our findings provide a foundation for future research whether for breeding peppers with higher vitamin content or for understanding how these nutrients are made in plants. Since peppers are a major source of vitamins in many diets, improving their nutritional quality could have real health benefits

3.4.3 OBJECTIVE 3: Identification and expression analysis of genes responsible for regulation of Vitamins C and E in *Capsicum* spp.

This study investigates the molecular mechanisms of vitamin biosynthesis in *Capsicum* species to establish a genetic framework for nutritional enhancement. By characterizing the key enzymatic pathways and regulatory networks responsible for vitamin production in peppers. Since not all genes involved in vitamin biosynthesis have been identified in peppers, we used a comparative approach, looking at known genes from tomatoes and finding their counterparts in different *Capsicum* species. This method has been successful in studying fruit development in peppers before (Dubey et al., 2019), but until now, only a few Vitamin C (VitC) genes were known, and no Vitamin E (VitE) genes had been reported. Furthermore, this study expands our understanding of vitamin biosynthesis in *Capsicum* by providing the first comprehensive genomic and expression-based profiling of genes involved in the synthesis of both VitC and VitE. Previous research on *Capsicum* primarily focused on a limited number of VitC genes (Chiaiese et al., 2019), and no systematic study has found VitE genes. While earlier research in peppers was largely restricted to a handful of VitC-related genes (Chiaiese et al., 2019), with minimal emphasis on VitE pathways, our findings provide the first study of both VitC and VitE associated genes across three major *Capsicum* species. By leveraging comparative genomics with *Solanum lycopersicum* as the reference, we identified 110 genes in *C. annuum*, 129 in *C. baccatum*, and 106 in *C. chinense*, which represents the most extensive resource for vitamin biosynthesis in peppers to date.

Earlier to this study, only a limited number of VitC biosynthesis genes (e.g., *GME*, *VTC2*, *GLDH*) had been described in *Capsicum* (Chiaiese et al., 2019). Our study provides a substantial extension of vitamin biosynthesis. By leveraging a comparative genomics approach using *Solanum lycopersicum* as a reference, we successfully identified 110 genes (29 VitC, 81 VitE) in *C. annuum*, 129 (44 VitC, 85 VitE) in *C. baccatum*, and 106 (36 VitC, 70 VitE) in *C. chinense* biosynthesis-related genes. **(Supplementary Table 3)**. Importantly, we added two novel *GPI* and *PMI*, previously uncharacterized in pepper, along with expression-supported evidence for *GulLO*, which had been identified but function not studied in earlier studies (Ahmad et al., 2021). These additions not only strengthen the evidence for the L-galactose pathway but also point toward potential activity of alternative (*GulLO*-mediated) biosynthetic and recycling routes.

Our findings validate and extend studies in tomato, which identified around 12 VitC biosynthesis genes, in addition to multiple recycling and degradation-related components (Cruz-Rus et al., 2011; Ioannidi et al., 2009). Interestingly, the larger complement of VitC-related genes we document in *Capsicum* parallels reports of expanded ascorbate-related gene families in certain fruit crops with dynamic ripening systems (El Airaj et al., 2013; Mellidou & Kanellis, 2017). The high expression of certain candidates during ripening stages, particularly in *Capsicum* fruits, also reflects patterns seen in tomato (*SlVTC2*, *SlGME*) where activity peaks coincide with antioxidant demand and fruit maturation (Alhagdow et al., 2007; Ioannidi et al., 2009).

A similar trend is observed for Vitamin E biosynthesis, earlier studies found 41 genes in

tomatoes (Bhandari et al., 2013), 7 genes in rice, 11 genes in Populus (Daood et al., 1996) and 7 genes in Arabidopsis (Fritsche, Wang, & Jung, 2017). Our findings go further, identifying genes in each pepper species, giving us a much clearer picture of how VitE is made in these plants and identification of up to 85 VitE-related genes in *C. baccatum* **[Fig. 3.11 (A) and Supplementary Table 1]** suggests significant gene family expansion or diversification events that may be unique to the *Capsicum* lineage, possibly reflecting its ecological or physiological specialization.

A major strength of this study is the integration of in silico expression analysis with transcriptomic and qRT-PCR validation across fruit ripening stages **[Fig. 3.11, 3.12, 3.13, 3.17 and 3.18]**. The strong positive correlations ($r > 0.9$) between expression of several VitC and VitE genes and measured vitamin content suggest key roles for these candidates in controlling vitamin accumulation in developing fruits. Negative correlations, on the other hand, may reflect compensatory regulation, feedback inhibition, or flux toward alternative metabolic fates. Such complexity has previously been reported in tomato, where ascorbate levels are modulated not only by core biosynthetic genes but also by redox balancing processes and recycling enzymes (Alhagdow et al., 2007; Ioannidi et al., 2009). Stage-specific expression patterns further demonstrate temporal regulation throughout pepper ripening. In particular, an increase in VitC-related gene activity was observed during the breaker and red ripe stages, similar reported for *SlVTC2* in tomato (Cruz-Rus et al., 2011). Similarly, tocopherol biosynthesis genes in *Capsicum* showed increased expression during late ripening stages, consistent with tissue-specific accumulation patterns described previously (Osuna-García et al., 1998; Quadrana et al.,

2012).

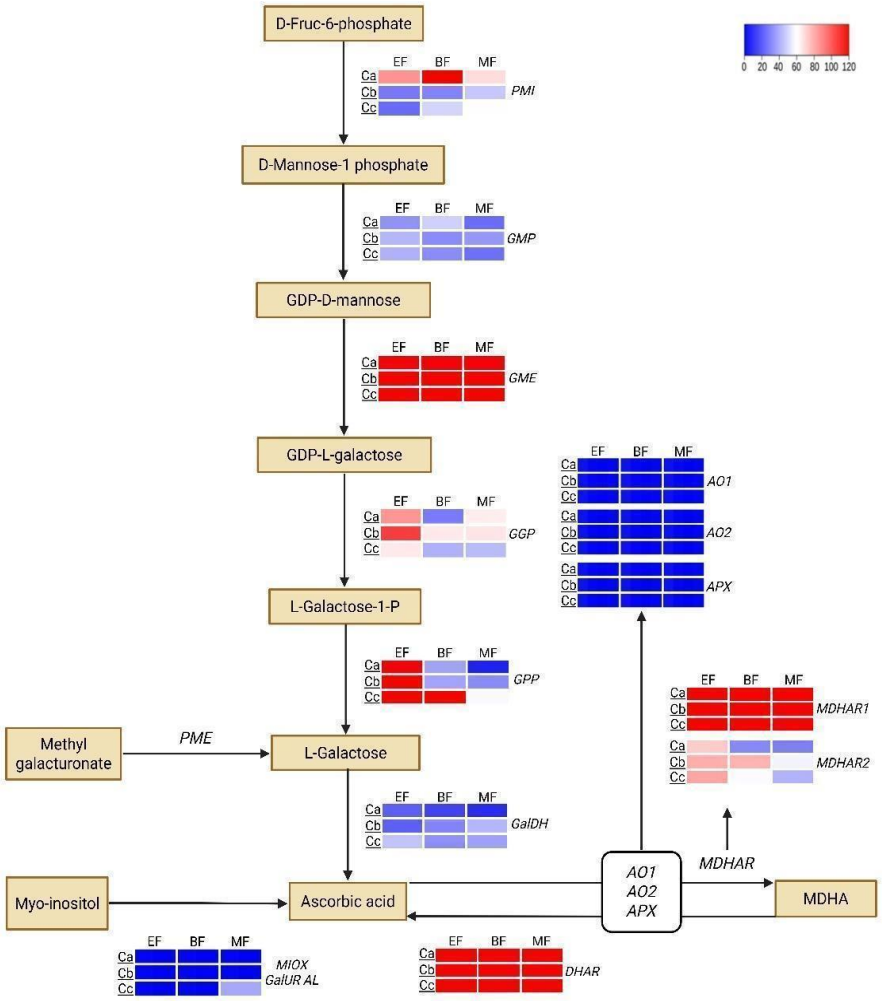


Fig.. 3.17 The Vitamin C biosynthesis pathway illustrating important enzymatic steps, genes [*PMI (Phosphomannose Isomerase), GMP* (*GDP-mannose pyrophosphorylase*), *GME* (*GDP-mannose-3', 5'-epimerase*), *GGP* (*GDP-L-galactose-1-phosphate phosphorylase*), *GPP* (*GDP-L-galactose-1-phosphate phosphorylase*) *GalDH* (*L-Galactose dehydrogenase*), *MDHR Monodehydroascorbate reductase*), *APX (Ascorbate Peroxidase), AO1(Ascorbate Oxidase1), AO2(Ascorbate Oxidase2), DHAR (Dehydroascorbate reductase) and MIOX (Myo-inositol oxygenase)*] and their transcriptome expression in three different fruit developmental stages (EF = Early Fruit, BF = Breaker Fruit, and MF = Mature Fruit) of three *Capsicum* species (Ca = *C. annuum,*

Cb = *C. baccatum,* and Cc = *C. chinense*). Scale given in upper right corner (blue—very low or negligible; red- high). The pathway has been modified from (Mellidou et al., 2021).

Collectively, our results provide a robust genetic framework for understanding vitamin biosynthesis in *Capsicum* and extending it to crop improvement strategies. The identification of novel pathway members such as *GPI* and *PMI* in peppers opens opportunities for targeted genome editing, while expression-supported candidates in the VitE pathway offer new routes for metabolic engineering to enhance tocopherol content. Given global micronutrient deficiencies, which affect billions worldwide and contribute significantly to health problems (Naidu, 2003; Walingo, 2005), strategies aimed at enhancing vitamin content in staple crops such as peppers are important for improving human nutrition and health.

This study demonstrates the power of comparative genomics in uncovering hidden metabolic potential across related plant lineages. While tomato has been the dominant model for studying fruit ripening and vitamin metabolism (Bhandari et al., 2013; Ioannidi et al., 2009), our results highlight that peppers harbor unique gene complements and regulatory dynamics that merit independent exploration.

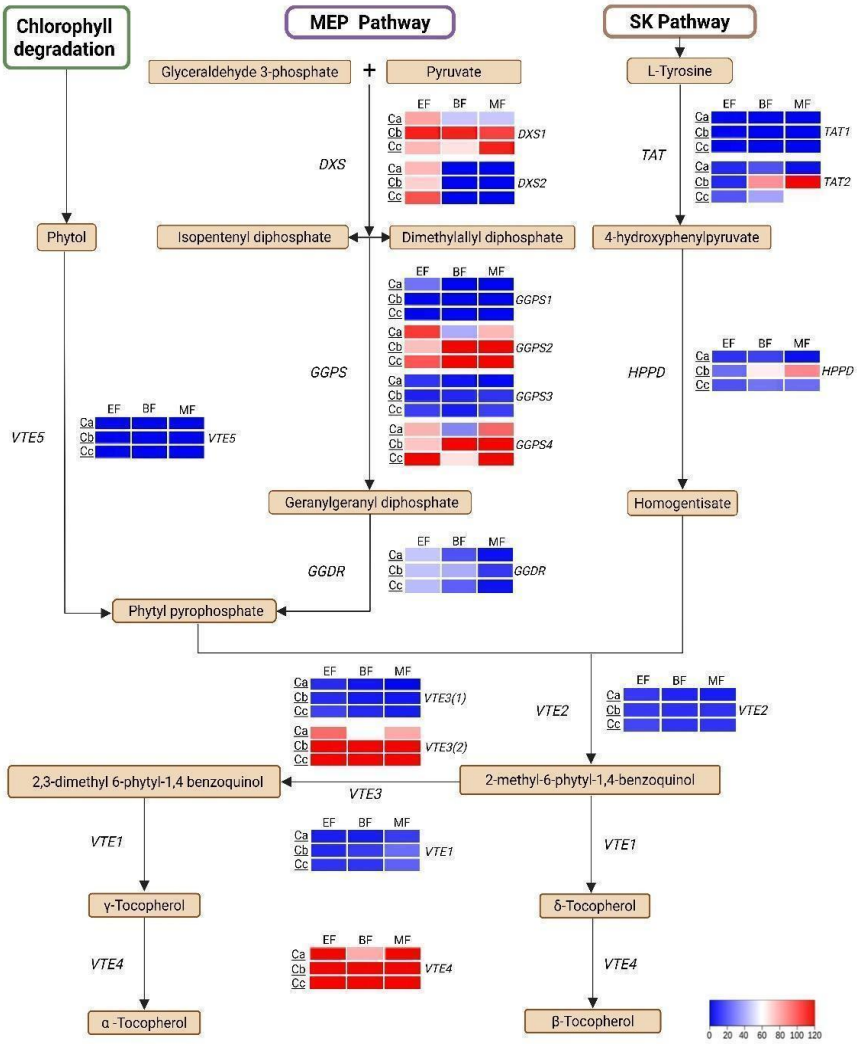


Fig. 3.18 The Vitamin E biosynthesis pathway illustrating important enzymatic steps, genes [*DXS (1- deoxy-Dxylulose- 5-phosphate synthase), GGPS(Geranylgeranyl pyrophosphate synthase), GGDR(Geranylgeranyl reductase), TAT(tyrosine aminotransferase), HPPD(4- hydroxyphenylpyruvatedioxygenase),VTE5(phytol kinase), VTE2(homogentisatephytyl transferase), VTE3(2-methyl- 6-phytyl- 1,4-benzoquinol methyltransferase), VTE1(tocopherol cyclise),* and *VTE4(γ-tocopherol C-methyl transferase)]* their transcriptome expression in three different fruit developmental stages (EF = Early Fruit, BF = Breaker Fruit, and MF = Mature Fruit) of three *Capsicum* species (Ca = *C. annuum,* Cb = *C. baccatum,* and Cc = *C. chinense*). Scale given in lower right corner (blue-very low or negligible; red-high). The pathway has been modified from (Meena et al., 2025).

Using the data generated in the present study, which laid the foundation for future *Capsicum* research, future work must proceed along two parallel paths: Functional characterization of identified genes, and association analysis of gene based molecular markers with phenotypes. First, candidate genes should be functionally characterized through approaches such as overexpression, silencing, or gene editing to confirm their mechanistic roles in fruit development and vitamin biosynthesis. Second, the developed SSR markers need validation through genotype–phenotype association studies in large natural populations or biparental mapping populations segregating for fruit traits and vitamins content. The integration of these validated genes and markers will provide a robust foundation for marker-assisted breeding to develop nutritionally enhanced *Capsicum* varieties.

CHAPTER 4

CONCLUSION, FUTURE SCOPE & SOCIAL IMPACT

Chapter 4: Conclusion, Future Scope & Social Impact

4.1. Conclusion- *Capsicum*s or chili peppers are important crop plants, widely used as vegetables, condiments, and in pharmaceutical industries, as they contain many health-beneficial metabolites. Although several cultivars and varieties have been developed through conventional breeding approaches, a detailed analysis of the genes and their expression profiles governing fruit traits and vitamin biosynthesis has not yet been conducted comprehensively in *Capsicum*. To fill this gap, we present the first comprehensive analysis of the *Capsicum* genome to identify genes involved in fruit development, ripening, and vitamin biosynthesis. Furthermore, their expression patterns and correlations with phenotypic traits were analyzed. In addition, molecular markers were developed based on genes associated with fruit development, ripening, and vitamin biosynthesis. The detailed findings and conclusions are presented below.

Objective 1: Identification and expression analysis of genes involved in *Capsicum* fruits development and ripening: -

The diversity within *Capsicum* harbors significant variations in fruit traits such as color, shape, size, and fruiting habits. However, genome-wide analyses of genes involved in fruit development and ripening are limited. Therefore, using known genes from tomato and a comparative genomics approach, we identified a total of 41 genes associated with fruit development and ripening in the *Capsicum annuum* genome, of which 38 were found on different chromosomes, and three were on unsigned scaffolds. Furthermore, 12 key genes were found to exhibit differential expression in contrasting fruit sizes among *C. annuum*

(medium-sized fruits), *C. chinense* (large fruits), and *C. frutescens* (small fruits) across three developmental stages, indicating a correlation between gene expression and fruit size. Notably, the non-climacteric regulator *MADS-RIN* (LOC107847473) showed striking upregulation (>500-fold) in mature *Capsicum* fruits, while *Ethylene receptor 2-like* (*LOC107873245*) displayed climacteric behavior unique to C. frutescens. These findings highlight the species-specific complexity of ripening regulation and bridge the gap between climacteric (ethylene-dependent) and non-climacteric models. In addition, we developed 49 SSR markers and validated in 47 contrasting germplasm belonging to ***C. annuum*** **(**comparatively medium fruit size except sweet peppers) , ***C. chinense*** with large fruit size and ***C. frutescens*** with small fruits After further validation through genetic mapping or functional characterization , the identified genes and markers could be used by breeders in *Capsicum* breeding programs to manipulate fruit size and ripening traits.

Objective 2: Determination of Vitamin C and E content at different fruit developmental stages in contrasting *Capsicum* genotypes: -

Apart from the wide diversity in fruit traits, *Capsicum* also exhibits substantial variation in fruit metabolites, including primary and secondary metabolites such as carotenoids, alkaloids (e.g., those contributing to pungency), and vitamins. *Capsicum* germplasm from India, particularly from Northeast India, has not been extensively explored for vitamin content. Therefore, in this study, we analyzed the levels of Vitamins C and E across three fruit developmental stages. Our results showed that genotypes of *C. annuum* and *C. frutescens* had higher concentrations of Vitamin C in the mature red fruit stage, whereas in *C. chinense,* elevated Vitamin C content was observed both at the early stage and in the

mature red stage. Overall, *C. chinense* emerged as a Vitamin C powerhouse, followed by *C. annuum* and *C. frutescens*. Conversely, higher Vitamin E accumulation was recorded in *C. annuum* and *C. frutescens*. The differential vitamin content across *Capsicum* species and genotypes suggests a genetic basis for this variation. Identifying genotypes with high vitamin content provides valuable genetic resources that can be incorporated into breeding programs to develop *Capsicum* varieties enriched in Vitamins C and E.

Objective 3: Identification and expression analysis of genes responsible for the regulation of Vitamin C and E in *Capsicum* species: -

As studies on the identification of genes involved in or related to Vitamin C and E biosynthesis are limited, we conducted a comprehensive analysis in three *Capsicum* species. Using reported tomato gene sequences, we identified 29 and 81 genes in *C. annuum*, 44 and 85 genes in C. baccatum, and 36 and 70 genes in *C. chinense* associated with Vitamin C and Vitamin E biosynthesis, respectively. Expression analysis of these genes using transcriptome data from contrasting *C. chinense, C. annuum*, and *C. frutescens* genotypes revealed differential expression of several genes. Notably, seven key genes- *CaGME, CaAKR38, CaGPI, CaGPP, CaGalDH, CaAKR1*, and *CaPMI*- showed significant correlations with Vitamin C content, while four genes- *CaGGDR, CaIPI, CaVTE2,* and *CaVTE3*- displayed strong correlations with Vitamin E content. These positive correlations suggest the potential involvement of these genes in the biosynthesis and accumulation of Vitamins C and E in *Capsicum* fruits. In addition, 24 SSR and 37

InDel markers were developed from vitamin pathway genes, providing actionable resources for biofortification efforts. Synteny analysis of chromosomal segments across Solanaceae genomes further highlighted both diversification and conservation patterns of Vitamin C and E gene-containing regions. By integrating molecular insights with practical tools, this research advances the development of *Capsicum* varieties with improved fruit quality, nutritional value, and adaptability, thereby addressing both agricultural and consumer needs.

In summary, this work lays the foundation for further investigations, such as:

1. Validation of markers in diverse germplasm to strengthen breeding pipelines.
2. Application of gene-editing tools (e.g., CRISPR) to manipulate desirable ripening and nutritional traits.
3. Cross-species synteny studies to explore evolutionary adaptations in Solanaceae crops.

Taken together, the identified genes and markers, after further validation, could be effectively deployed in breeding programs to manipulate fruit traits and enhance vitamin content in *Capsicum*.

4.2 Future Prospects

The research outcomes reported in our study have the following future applications and scope:

4.2.1 Application in *Capsicum* Improvement

(a) *Marker-Assisted Selection*: Validate and deploy the newly developed SSR and InDel markers to breed peppers with higher vitamin content, synchronized ripening, and extended shelf life. These tools will enable farmers to select superior plants more quickly and accurately. Once validated and incorporated into breeding programs, these markers can significantly enhance Vitamin C and E content.

(b) *Gene Editing*: Employ CRISPR/Cas9 to fine-tune key genes such as *MADS-RIN* (ripening), *CaGME* (Vitamin C), and *CaVTE4* (Vitamin E) to develop nutrient-rich varieties without compromising yield or flavor. This approach will also help to elucidate the molecular basis of fruit development, ripening, and vitamin biosynthesis in *Capsicum* species.

The integration of traditional breeding approaches, such as marker-assisted breeding using the developed molecular markers, with advanced biotechnological tools like genome editing to validate SSRs and identified genes, and their utilization in breeding programs, will accelerate the development of improved *Capsicum* varieties with desirable fruit traits and enhanced vitamin content.

4.2.3 Application in commercial product development as supplement to combat malnutrition

This work further lays the foundation for developing functional foods through marker-assisted breeding aimed at enhancing vitamin production after validation. High-vitamin and low-pungency *Capsicum* varieties, once developed, could be used to produce vitamin powders, natural colorants, or functional foods to combat malnutrition. Such varieties would also help farmers identify high-vitamin or stress-tolerant accessions suited to their local conditions.

4.3 Social Impact-

The research presented in this study on *Capsicum* fruit development, ripening, and vitamin biosynthesis has significant social implications, particularly in the areas of nutrition, agriculture, and economic development. The key social impacts of this work include:

4.3.1 Enhanced Nutritional Security

(i) The identification and characterization of genes involved in Vitamin C and E biosynthesis in *Capsicum* species provide a strong foundation for breeding programs aimed at enhancing the nutritional quality of peppers, which in turn will help combat malnutrition.

(ii) Since *Capsicum* spp. are staples in many diets worldwide, increasing their vitamin content could play a crucial role in addressing micronutrient deficiencies, particularly in regions where access to diverse foods is limited

4.3.2 Improved Agricultural Practices

(i) The development of gene-based SSR and InDel markers enables marker-assisted breeding, allowing farmers to cultivate *Capsicum* varieties with higher vitamin content, improved fruit quality, and extended shelf life.

(ii) These improvements can help reduce post-harvest losses and increase profitability for small-holder farmers, particularly in developing countries.

4.3.3 Economic Benefits for Farmers

The markers and genes identified in this study, once validated and applied in breeding programs, can facilitate the development of *Capsicum* varieties with desirable fruit traits and enhanced vitamin content. Such high-nutrient, improved-quality varieties could command premium prices in local and international markets, thereby improving farmers' livelihoods.

4.3.4 Health and Dietary Advancements

35

(i) *Capsicum* species are rich in antioxidants and vitamins, which are associated with reduced risks of chronic diseases such as cardiovascular disorders and cancer. The markers and genes identified in this research, once validated and used in breeding programs, will help enhance desirable fruit traits and nutritional content. Such improvements can contribute significantly to better public health outcomes.
(ii) The study's findings can also inform dietary recommendations and fortification strategies, particularly in populations with limited access to fresh produce.

4.3.5 Scientific and Educational Contributions

15 This research advances the understanding of genome-wide gene identification and expression analysis, laying a foundation for further investigations into fruit development,
15 ripening, and vitamin biosynthesis. These insights will benefit plant biologists and geneticists working on *Capsicum* as well as other related Solanaceae species.

RESEARCH SUMMARY

Identification and expression analysis of genes involved in vitamins biosynthesis and fruit ripening in *Capsicum* species

Meenakshi Dubey (2k18/PHD/BT/505)

Research Summary

Introduction

The crop plants under genus *Capsicum* belonging to the family Solanaceae, known for their diverse fruit traits such as size, shape, color, and, most notably, flavor are cultivated extensively. The center of origin for *Capsicum* is the Americas, with evidence pointing to Bolivia and surrounding regions as the primary center of diversification (Kraft et al., 2014). The five cultivated species, which are *C. annuum*, *C. baccatum*, *C. chinense*, *C. frutescens*, and *C. pubescens* have distinct genetic backgrounds and were domesticated in different regions, leading to a wide array of fruit morphologies and chemical compositions (Pickersgill, 2007; Sarpras et al., 2016). This genetic diversity is a valuable resource for breeding programs aimed at improving yield, disease resistance (Hill et al., 2017; Naegele et al., 2014), and nutritional quality. The characteristic hot taste of *Capsicum* fruits is attributed to capsaicinoids, a group of alkaloids synthesized in the placenta of fruits.

Beyond their culinary uses, *Capsicum* fruits are also widely used as vegetables (Bosland & Votava, 2012), and in traditional medicine, particularly in regions like Northeast India, where they are believed to offer a range of therapeutic benefits. *Capsicum* is valued for its high content of antioxidants, such as Vitamin C, carotenoids like β-carotene, phenolic compounds and mineral elements. These nutrients not only contribute to its health benefits such as having anti-carcinogenic and anti- inflammatory properties (Zimmer et al., 2012) but also enhance its role as a functional food, capable of providing essential vitamins and other beneficial compounds. The genetic diversity within the genus offers significant potential for improving crop yield, nutritional content, and disease resistance, making *Capsicum* a key focus of agricultural research.

Several important genes responsible for morphological and physiological trait variations in *Capsicum* have been reported (Sarpras et al., 2016; D. Wang & Bosland, 2006). Genetic and genomic resources such as ESTs, molecular markers, genetic linkage maps (Kong et al., 2012), transcriptome, and whole genome sequencing of different *Capsicum* species has led to identification genes and their expression patterns involved in various growth and development ,and stress responses (Hulse-Kemp et al., 2018; H.-J. Kim et al., 2008; S. Kim et al., 2017; Kim SeungIll et al., 2014; Qin et al., 2014; Zhong et al., 2013). Among the marker's development, several SSR markers have been developed in *Capsicum* species mostly in *C. annuum* (H. Chen et al., 2016; Chhapekar et al., 2020; Ramchiary et al., 2014). A few studies reported the molecular mechanism underlying the morphological and physiological changes during early and later fruit growth and development (Tiwari et al., 2013). The molecular mechanisms of vitamins biosynthesis as well as fruit

development and ripening are well established in tomato. Tomato has two *GOLDEN2-LIKE* transcription factors (*GLKs*) i.e., *GLK1* and *GLK2/UNIFORM*, *GLK1* is active in leaves, while *GLK2* is predominant in fruit. It is a transcription factor that regulates coloration during fruit ripening (Nguyen et al., 2014). A previous study reported that the AOX gene plays an important role in controlling tomato fruit ripening. *Ethylene insensitive 2 (EIN2)* gene was identified and characterized as a feedback loop and its important role in fruit ripening in tomatoes was explored. Subsequent silencing of *EIN2* resulted in changes in expression of several other genes related to ethylene production, photosynthesis, defense etc. and led to non- ripening phenotype (R. H. Wang et al., 2016).

Vitamins are very important beneficial health compounds. Only a few studies have reported analysis of Vitamin C and E content in *Capsicum*. *C. annuum* accessions from different places of South-Eastern Anatolia Region were used for Vitamin C estimation and Gaziantep red hot pepper was reported to contain the highest Vitamin C (Korkutata & Kavaz, 2015). A study conducted on five samples of *C. annuum* from Turkey suggested that microwave dried chili fruit samples have greater Vitamin C and E content than sundried samples (Karatas et al., 2021). Vitamins are essential nutrients of our food which are primarily obtained from plants (Macknight et al., 2017). Vitamin C, also known as ascorbic acid (AsA), is commonly found in fruits and vegetables, and it is one of the most remarkable vitamins for human well-being (Flier et al., 1986). Similarly, Vitamin E, also known as tocopherol, has powerful roles in eradicating various human health problems such as cardiovascular diseases, cancer, pulmonary tuberculosis, cataracts, and helps to improve immunity (Hussain et al., 2019; G. Y. Lee et al., 2018; Rizvi et al., 2014). In a

recent study, nine accessions of *Capsicum* spp. were examined for Vitamin C and other bioactive compounds and their correlation with antioxidant activity was analyzed (Lidiková et al., 2021). During fruit ripening of *C. annuum*, γ-tocopherol was found in seeds and α-tocopherol was present in the fruit pericarp (Osuna-García et al., 1998). Further, intra and interspecies variations of vitamins content were observed at different stages of fruit development (Wildman et al., 2016).

The Vitamin C biosynthetic pathway is a combination of the myo-inositol pathway, L-galactose pathway, L-glucose pathway and galacturonic pathway. Similarly, the Vitamin E biosynthetic pathway begins from the *PDS1* gene (Fritsche, Wang, & Jung, 2017). The genes *VTE1* (*Tocopherol cyclase*), *HPPD* (*4-hydroxyphenylpyruvate dioxygenase*), and *VTE5* (*phytol kinase*) are well characterized and found to be involved in Vitamin E biosynthesis in tomato fruits (Almeida et al., 2011, 2015; Quadrana et al., 2012).

Research gaps and Statement of the Problem

Despite several phenotypic, genetic, molecular, and omics-based studies conducted to understand the fruit traits and metabolite biosynthesis in *Capsicum* species (Ahmad et al., 2021, 2023; Jaiswal et al., 2020, 2022; Momo et al., 2022, 2025; Rawoof et al., 2022; Sarpras et al., 2016, 2019), the study of molecular mechanism underlying fruit development and ripening, and Vitamins biosynthesis in *Capsicum* remains limited. For breeding desired fruit traits and high vitamins content, identification of genes and their expression which correlates with phenotypes are important. Furthermore, the Northeast Indian *Capsicum* germplasm is not explored for analysis of vitamins content in different

development stages of fruits and no molecular markers to breed *Capsicum* varieties with desired fruit and vitamins content are reported. This gap needs to be addressed urgently; therefore, we have taken this study to: - 1. Identify and characterize genes involved in fruit development and ripening, 2. Quantify Vitamin C and E content across developmental stages and contrasting genotypes, 3. Understand the genetic regulation of vitamins biosynthesis, and 4. Develop molecular markers for marker-assisted breeding of *Capsicum* varieties for desired fruit traits and vitamins content.

Formulation of research objectives to solve the research gaps -To solve the research problem highlighted above, this study is divided into three objectives:

Objective 1: Identification and expression analysis of genes involved in *Capsicum* fruit development and ripening.

Objective 2: Determination of Vitamin C and E content at different fruit developmental stages in contrasting *Capsicum* genotypes.

Objective 3: Identification and expression analysis of genes responsible for regulation of Vitamin C and E in *Capsicum* spp.

Methods/ Approaches to achieve the above objectives

The present research employed a multi-disciplinary approach combining genomics, transcriptomics, biochemistry, and bioinformatics to carry out the work objectives mentioned above and are described below:

1. Identify orthologs of known tomato fruit development and ripening, and vitamins biosynthesis genes in *Capsicum* and analyze their expression patterns across different fruit

developmental stages in diverse *Capsicum* species. Identification of genes from the *Capsicum* genome was done using BLASTN searches against the *C. annuum* reference genome using tomato fruit development and ripening genes orthologs. Further expression analysis of genes was done in transcriptome data developed from three developmental stages of fruit belonging to *C. annuum* (medium size fruit), *C. chinense* (large fruit) and *C. frutescens* (small fruit). Transcriptome analysis was done following standard protocol using different analysis tools (FastQC, TrimGalore, TopHat, Cufflinks etc.,(Rawoof et al., 2022). Genes showing differential expressions in transcriptome were used for qRT-PCR validation using RNA extracted from three genotypes belonging to different fruit sizes at different development stages. Additionally, gene based simple sequence repeat (SSR) molecular markers were developed. SSRs were identified in and around targeted genes using WebSat tools, and primers were designed using g primer express 3.0.1 software. Genotyping of SSRs is done by PCR amplification and resolving on 5% polyacrylamide gels. For genotyping simple sequence repeat markers, 47 diverse *Capsicum* genotypes belonging to *C. annuum*, *C chinense* and *C. frutescens* were used. Allele sizes were recorded and used for diversity analysis.

2. Vitamins content analysis: Early, breaker, and mature fruits from 30 genotypes belonging to *C. annuum*, *C chinense,* and *C. frutescens* were collected by freezing into liquid nitrogen and stored in an ultra-deep freezer until used. Samples were extracted using standard methods (Daood et al., 2006; Saini & Keum, 2016). Vitamin contents were analyzed using HPLC (Dionex Ultimate 3000 model HPLC, manufactured by Thermo Fisher. A C18 column with dimensions of 150 × 4.6 mm and a particle size of 3 µm) and

quantified using validated protocols with three technical and three biological replicates (Dubey et al., 2019). For statistical analysis, data were analyzed to compare vitamins levels across species, genotypes, and developmental stages using standards of VitC and VitE (Sarpras et al., 2019).

3. Identification and expression analysis of genes involved in Vitamins C and E biosynthesis: A comprehensive identification of genes involved in Vitamins C and E Biosynthesis in *C. annuum, C. chinense* and *C. baccatum* genome was done using reference genome of these three *Capsicum* species as described for fruit development and ripening using tomato orthologs. We used the same protocol for analysis of transcriptome, qRT-PCR analysis, and development of SSRs based on genes involved in Vitamins C and E biosynthesis, as described for fruit development and ripening genes.

Results and discussion:

(1) This study explored the *Capsicum* genome and identified genes involved in fruit development and ripening through comparative genomics approach using 32 key tomato fruit development and ripening genes orthologs. A BLAST search against the *C. annuum* genome identified 41 orthologous genes in *Capsicum* genome with at least 80% similarity. 38 out of 41 were present in all 12. *C. annuum* chromosomes and 3 were on scaffold of genome, maximum 9 genes present in chromosome 10, followed by 8 genes on chromosome 1, 3 genes on each from chromosomes 2, 5, and 11, and 2 genes each on chromosomes 4, 6, 7,9, and 12 while Chromosomes 3 and 8 each have just one gene. For seven tomato genes, two or more orthologs were found. Synteny analysis revealed

that 18 of these genes showed conserved chromosomal positions between tomato and pepper, while the remaining 14 were scattered across the genome.

Following identification, the expression of these genes was analyzed using the available transcriptome data generated from flower, early fruit, breaker fruit, and mature fruit stages of *C. chinense, C. annuum*, and *C. frutescens*. Thirty-eight of the 41 genes analyzed were found expression in different tissues. Among these, 12 genes showed significant differential expression, suggesting their key roles in fruit development and ripening. Further, the expression patterns of these 12 differentially expressed genes were validated using qRT-PCR across five tissues belonging to contrasting C. *annuum* (medium fruit size), *C. chinense* (large fruit), and *C. frutescens* (small fruit). While similar expression patterns were observed, such as the extreme upregulation of *MADS-RIN* in ripening fruits, species-specific differences in expression intensity and tissue specificity were also identified for genes like *Agamous-like MADS-box protein AGL8 homolog (FUL2)* is involved in lycopene synthesis, and its silencing results in altered pigmentation. *Golden 2-like protein (GLK2)* is a transcription factor that regulates plastid and chlorophyll levels, making it responsible for coloration during fruit ripening. *Glycosyltransferase (NSGT1)* converts the cleavable di-glycosides of smoky-related phenylpropanoid volatiles into non-cleavable tri-glycosides, thereby preventing their de-glycosylation and release from tomato fruit upon tissue disruption. A strong correlation between the transcriptome and qRT-PCR data validated the overall findings, though some discrepancies suggested the necessity of further experimental validation. To leverage these genes for breeding, user-friendly PCR-based SSR markers were developed. A total of 49 SSRs were designed, with 14 derived from

within the gene sequences and 35 from close vicinity (within 5 kb upstream or downstream). The properties of these markers, including motif type and location, were characterized. Further, for genotyping and validation of these SSR's, 42 of the 49 SSRs were selected and tested. Twenty-three were found to be polymorphic, including six genic SSRs. Genetic diversity analysis using these polymorphic markers revealed a total of 50 alleles, with 2-3 alleles per locus. Key diversity parameters including Shannon's index (I), Nei's gene diversity, and Polymorphic Information Content (PIC) were calculated, confirming their utility for distinguishing between pepper varieties. Markers such as SSR_CF-5 and SSR_CF-3 were identified as being highly informative.

(2) This study provides the first report on analysis of Vitamin C and E content in Northeast *Capsicum* germplasm belonging to *C. annuum, C. chinense*, and *C. frutescens* at three fruit developmental stages: Early Fruit (EF), Breaker Fruit (BF), and Mature Fruit (MF). Vitamin C (ascorbic acid) content was highly variable across the 30 genotypes. Levels changed significantly across development stages, with the highest content found in genotype belonging to *C. chinense* i.e. Cc-9 (38.41 mg/g) at the EF stage and in genotypes of *C. annum* i.e. Ca-8 (41.76 mg/g) at the MF stage. On average across all stages, *C. chinense* possessed the highest Vitamin C content (25.33 mg/g), followed by *C. annuum* (24.25 mg/g) and *C. frutescens* (22.09 mg/g). In short, Vitamin C content depends greatly on genotype, species, and ripening stage, with *C. chinens*e having the highest average concentration.

Similarly, Vitamin E content varied significantly across genotypes and developmental stages. The highest content was found in genotype Ca-5 (111.43 mg/g) at the EF stage and in *C.*

frutescens genotype i.e. Cf-9 (112.4 mg/g) at the MF stage. On average, *C. annuum* had the highest Vitamin E content (85.63 mg/g), closely followed by *C. frutescens* (85.22 mg/g), and then *C. chinense* (83.41 mg/g). The study found that *C. chinense* had the highest average Vitamin C content, while *C. annu*um and *C. frutescens* had the highest average Vitamin E content. This indicates that the optimal *Capsicum* genotype/ species for breeding depends on the target vitamin. Our study recorded significantly higher Vitamin C content than previous reports (Gomes et al., 2019; Hamed et al., 2019; etc.), a finding consistent with the established pattern of accumulation during ripening where mature red fruit contains more than green fruit (Howard et al., 2000; Martínez et al., 2005; Olatunji et al., 2019). Similarly, the Vitamin E content was elevated, with the highest levels found in specific *C. annuum and C. frutescens* genotypes, aligning with known varietal differences (Olatunji et al., 2019) and the influence of maturity (Kim SeungIll et al., 2014; Nadeem et al., 2011). Therefore, the elevated vitamins levels are attributed to the advanced maturity of the fruit studied and the genetic superiority of the specific genotypes analyzed. Now we can say, this study found significantly higher Vitamin C and E contents than previous reports, a result attributed to advanced fruit maturity (red stage) and specific genotypic superiority, particularly.

(3) Genome wide analysis of Vitamins C and E was done as no report is available for Vitamin C and E biosynthesis in three *Capsicum* species (*C. annuum*, *C. baccatum,* and *C. chinense*) using 70 known tomato vitamins genes as a reference, 21 were from Vitamin C biosynthesis (Ioannidi et al., 2009) and 49 were from Vitamin E biosynthesis pathway (Quadrana et al., 2013). Our comprehensive analysis identified 29, 44, and 36 genes for

Vitamin C and 81, 85, and 70 genes for Vitamin E in *C. annuum*, *C. baccatum*, and *C. chinense* genomes, respectively. These genes were physically mapped onto the 12 pepper chromosomes. In *C. annuum*, 29 Vitamin C genes were physically mapped onto its 12 chromosomes, with most genes located on chromosome 3 (6 genes). In *C. baccatum*, 44 genes were mapped, with the highest number on chromosome 9 and the lowest on chromosomes 5 and 10 (1 gene each). In *C. chinense*, 36 genes were mapped, with most on chromosome 6 and the fewest on chromosomes 8 and 10 (1 gene each). Similar for Vitamin E, 81, 85, and 70 genes were mapped in the *C. annuum, C. baccatum*, and *C. chinense* genomes, respectively. In *C. annuum*, the maximum number of Vitamin E biosynthetic genes were present on chromosomes 2, 3, and 6 (10 genes on each chromosome). In *C. baccatum,* most genes were present on chromosomes 11, 9, and 10, and the minimum number of genes were on chromosomes 7, 9, and 10 (3 genes on each chromosome). In *C. chinense*, most genes were present on chromosome 6, and the minimum number was on chromosomes 9 and 10 (2 genes on each chromosome). The number of genes on each chromosome varied by species. This study provides a valuable genetic roadmap, pinpointing the specific locations of vitamin-related genes in the pepper genome, which is crucial for breeders to develop new pepper varieties with higher nutritional value. Earlier genomic studies on vitamin biosynthesis were primarily based on the tomato reference genome, identifying 21 genes for Vitamin C (Ioannidi et al., 2009) and 49 for Vitamin E (Quadrana et al., 2013). In contrast, this first-ever genome-wide analysis in three *Capsicum* species revealed a significantly greater number of genes for both vitamins, with the count and their physical chromosomal locations varying y among *C. annuum*, *C. baccatum*, and *C. chinense*. This comparison highlights a more complex genetic architecture

in peppers and provides a crucial genomic resource for targeted breeding of nutritionally enhanced varieties.

After gene identification, the expression of genes responsible for Vitamin C and E production was analyzed during different fruit developmental stages in three contrasting species of *Capsicum* (*C. annuum*, *C. chinense*, *C. frutescens*). The transcript abundance of vitamin biosynthesis genes changed significantly across fruit development stages (EF. BF, MF) and between species. For 18 out of 21 genes, their expression levels showed a strong positive correlation with the actual Vitamin C content measured in the fruit, meaning higher gene activity matched higher vitamin levels. Similarly for Vitamin E, genes like *CaVTE4, CaHDR*, and *CaEF-1α* were consistently highly expressed. This study moves beyond the foundational gene catalog of the vitamin E pathway established in tomato (Quadrana et al., 2013) by functionally characterizing expression in *Capsicum* species. It identifies 32 key genes whose expression strongly correlates with Vitamin E content, a significant advancement from merely identifying gene presence. Furthermore, it reveals species-specific expression patterns, such as high *CaVTE* activity in *C. chinense* and *C. frutescens* and *CaDXS* specificity to *C. chinense*, providing a genetic basis for the known varietal differences in Vitamin E accumulation reported in earlier phenotypic studies (Olatunji et al., 2019). Thirty-two out of 81 genes showed a strong positive correlation with the measured Vitamin E content. Unique expressions were found in certain species. For example, the *CaVTE* gene was highly active in *C. chinense* and *C. frutescens*, and the *CaDXS* gene was specific to *C. chinense*.

To confirm these genetic findings, the expression of Vitamin C and E biosynthesis genes was validated using qRT-PCR analysis. Eleven differentially expressed genes known to be crucial

for Vitamin C synthesis in tomatoes were chosen, with 7 from the main pathway and 4 from an alternate pathway. Gene expression varied significantly across different tissues and between the three *Capsicum* species. Some genes like *CaAKR38* and *CaAKR* were consistently less active in fruits than in leaves, while other genes like *CaMIOX* and *CaGME* were highly active in mature fruits, showing up to 40-fold higher expression. *C. chinense* and *C. frutescens* often showed similar gene expression patterns, which were frequently the opposite of those in *C. annuum,* InqRT-PCR validation provides crucial functional evidence in *Capsicum*, moving beyond the foundational gene catalogs established in tomato (Ioannidi et al., 2009; Quadrana et al., 2013). The discovery of strongly upregulated genes in mature fruit (*CaMIOX, CaGME*) aligns with ripening-associated expression patterns observed in tomato (Mellidou et al., 2012) but crucially reveals species-specific divergence, as *C. chinense* and *C. frutescens* exhibited distinct expression profiles from *C. annuum*. The expression levels of 7 out of the 11 genes showed a strong positive *correlation* with the actual Vitamin C content measured in the Capsicum, confirming their direct role in determining vitamin levels. The strong correlation between gene expression and vitamin content confirms the direct regulatory role of these biosynthetic genes, providing a mechanistic explanation for the known interspecies variation in *Capsicum* vitamin levels reported in earlier phenotypic studies (Olatunji et al., 2019; Wahyuni et al., 2013).

Similarly for Vitamin E, 10 differentially expressed genes involved in the final steps of tocopherol biosynthesis were selected for qRT-PCR validation. The expression of these genes varied across fruit development stages and *Capsicum* species. Gene *CaGGDR* showed

dramatically lower expression (down to 263-fold less) as the fruit matured. In contrast, the *CaTAT(1)* gene was highly active (up to 71-fold more) in later ripening stages, and the *CaHPPD* gene was 15-fold more expressed in breaker fruit of C. chinense compared to leaf. The expression of 6 out of the 10 genes showed a strong positive correlation with the actual Vitamin E content, confirming their direct role in vitamin accumulation. The gene expression patterns from the qRT-PCR experiment strongly agreed with the earlier transcriptome data, confirming the reliability of both analyses These findings align with and expand upon earlier reports in *Capsicum* and tomato, which also identified key tocopherol biosynthesis genes like *HPPD* and *TAT* as being highly expressed during fruit ripening and correlated with Vitamin E accumulation (Almeida et al., 2016; Quadrana et al., 2013). However, the dramatic downregulation of *CaGGDR* during maturation presents a contrasting pattern to some homologs in other species, suggesting a unique regulatory mechanism in *Capsicum* (Almeida et al., 2016). The strong correlation for six genes provides more precise genetic targets for breeding, building on previous studies that identified this pathway as crucial for nutritional quality.

.

To utilize the discovered vitamin biosynthesis genes in *Capsicum* breeding, two types of genetic markers (SSR's and InDels) were developed for tracking these valuable traits. A total of 185 SSR markers were developed, with 51 linked to Vitamin C genes and 134 to Vitamin E genes. These markers consisted primarily of di- and tri-nucleotide repeats, with the longest repeating motif extending to 31 repetitions. The majority of markers were located within the gene sequences themselves, while others were positioned in nearby upstream regulatory

regions. Importantly, 24 of these SSR markers demonstrated polymorphism between the three *Capsicum* species (*C. annuum*, *C. chinense, C. baccatum*) in computer-based analysis. Additionally, developed 37 polymorphic InDel markers, comprising 8 linked to Vitamin C genes and 29 to Vitamin E genes. These markers were distributed both within gene sequences (genic) and in adjacent promoter regions (non-genic), with sizes varying up to a 156 bp deletion in a Vitamin E gene. These natural variations were identified through comparative genomic analysis of the three pepper species. For both marker types, specific primers were designed, enabling breeders to employ simple PCR tests to track high-value vitamin genes and facilitate the development of more nutritious pepper varieties.

Synteny analysis was done to study the conservation and diversification of Vitamin C and E biosynthesis genes among *C. annuum*, *C. chinense*, *C. baccatum*, and tomato (*S. lycopersicum* genome). Synteny and collinearity analysis among four Solanaceae genomes (*C. annuum*, *C. chinense*, *C. frutescens*, and S. *lycopersicum*) identified 370 conserved chromosomal segments (CSSs) containing vitamin genes, showing these regions have been conserved through evolution. A significant number of vitamin biosynthesis genes are in collinear blocks, they are found in the same order on the chromosomes of these different species, indicating they were inherited from a common ancestor. The expansion of vitamin gene families was largely driven by duplication events, with 700 pairs of homologous genes (240 for Vitamin C, 460 for Vitamin E) found across the four species. The highest number of shared vitamin biosynthesis gene blocks was found between *C. chinense* and tomato, suggesting a particularly close evolutionary relationship in these metabolic pathways.

Future prospects

(1) This study presents the first comprehensive genome-wide identification of fruit development and ripening genes in *Capsicum*, including their expression analysis and SSR marker development and genotyping in northeastern *Capsicum* germplasm. The identified genes and developed SSR markers after further validation could be utilized for manipulating fruit traits in *Capsicum*.

(2) The identification of high Vitamin C and Vitamin E containing genotypes across different *Capsicum* species provides potential genetic stocks for breeding programs aimed at developing Vitamin C and Vitamin E enriched *Capsicum* varieties. This will facilitate the development of high commercial value biofortified *Capsicum* varieties and functional food for improved human health.

(3) This is the first comprehensive analysis of Vitamin C and Vitamin E biosynthesis genes in three *Capsicum* genomes, namely: *C. annuum, C. baccatum*, and *C. chinense*. Furthermore, the differentially expressed genes showing significant correlation with vitamins content could be validated for manipulating Vitamin C and Vitamin E levels in *Capsicum* fruits, enabling the development of high Vitamin C and Vitamin E content *Capsicum* varieties through biotechnological and molecular approaches. Additionally, the SSRs developed based on genes involved in Vitamin C and E biosynthesis could be associated with phenotypic traits through association mapping or QTL mapping in

biparental populations derived from contrasting Vitamin C and E containing genotypes. After validation they can be used for breeding.

Taken together, our findings laid a foundation for further *Capsicum* research on the manipulation of Vitamin C and Vitamin E and fruit traits, which will help to develop *Capsicum* varieties with high commercial value.

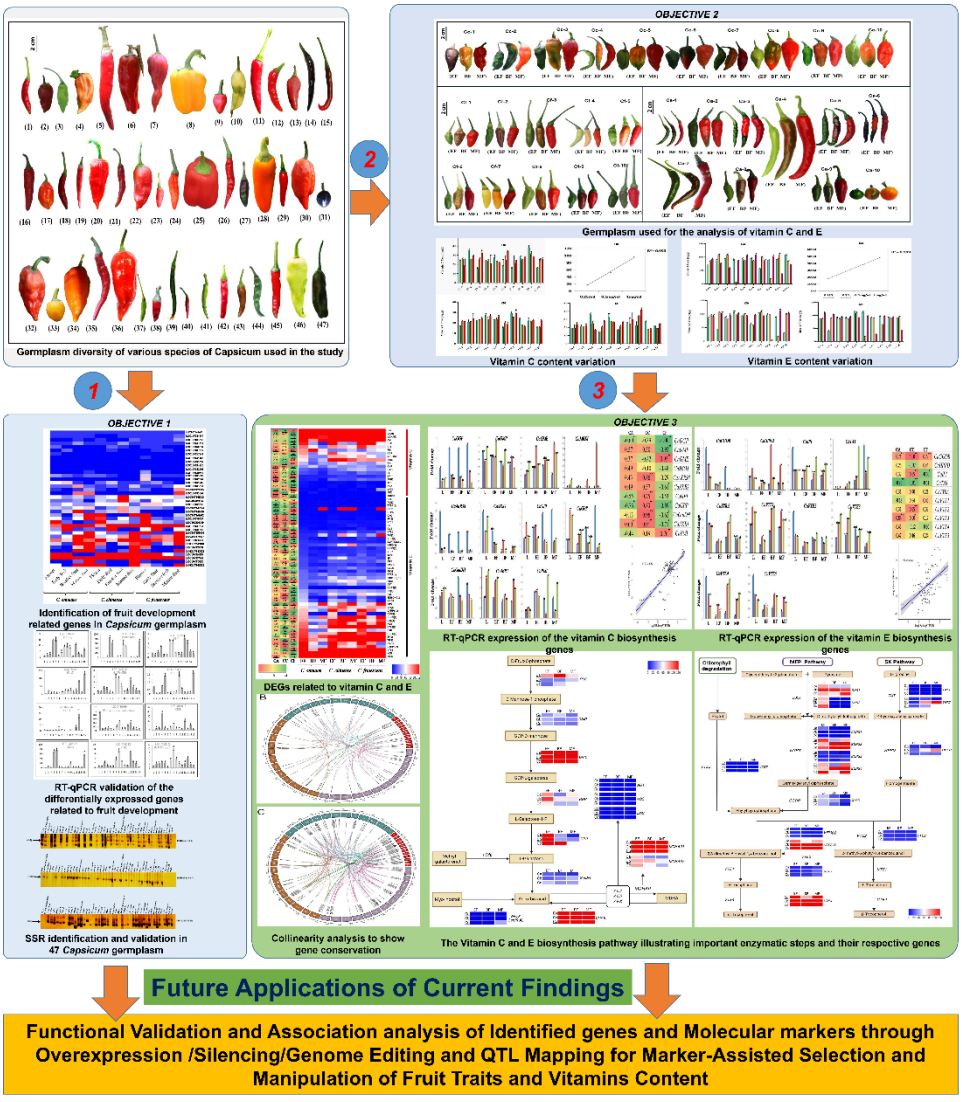


Fig.3.19. Graphical representation of overall summary of research findings under different objectives of study and their inter-relationships

CONFERENCES

1. Attending The Webinar on Plant MicroRNA Research held Department Of Botany, Aligarh Muslim University, Aligarh On 27th October 2020.

2. Participated In DTU-Sponsored One-Week International E-Workshop on Bioinformatics held At Department of Biotechnology Delhi Technological University on December 14-18, 2020.

3. Participated In The 26th International Conference of International Academy of Physical Sciences held at Mahatma Gandhi University, Kottayam, Kerala, India) during December 18-20, 2020.

4. Participated In DTU Sponsored One Day International E-Symposium on Women In Science held at Department ff Biotechnology Delhi Technological University on February 11, 2021.

5. Attended The International Webinar on Biotechnology for Crop Tolerance to Low and high Temperature Stresses, Organized by Department of Agriculture, Faculty of Agriculture, MMDU, Mullana-Ambala on January 15, 2022.

6. Participated In The 2nd International Conference on Recent Advances In Biotechnology and Nanobiotechnology (Int-BIONANO-2022) held at Amity Institute of Biotechnology, Amity University Madhya Pradesh, Gwalior During February 10-11, 2022.

7. Participated In The "International Conference on Applications of Natural Products Nanomaterials and Nano-Pharmaceuticals (ICAN3)" Organized By School of Life Sciences, B.S. Abdur Rahman Crescent Institute ff Science and Technology, Chennai, India In Association With Nano and Biomaterials Association (NBA), The Biotech Research Society (BRSI), India and Centre for Surface Technology and Applications (Cesta), Korea Aerospace University, Republic of Korea held on August 9 & 10, 2023.

8. Participated In The 3rd International Conference on Recent Advances In Biotechnology and Nanobiotechnology (Int-BIONANO-2024) held at Amity

Institute of Biotechnology, Amity University Madhya Pradesh, Gwalior during 5-6 March 2024.

9. Participated In The One-Day Online Training Programme on Soil Nutrient Management Organised by BUSINESS PLANNING AND DEVELOPMENT UNIT, JNKVV Jabalpur on 18th October 2024.

10. Participated In The GIAN Course Entitled "Medical Translational Research From Benchtop to Benchside" During February 24, 2025, To March 7, 2025, Organized by The Department Of Biotechnology Delhi Technological University.

11. Participated In The One-Month Online Internship on “Basics to Advanced Bioinformatics, Genomics And Ensembl & Biomart” Organised by Nextgenhelper, New Delhi during August 04 To August 30, 2025.

PUBLICATIONS

1. **Dubey, M.,** Islam, K., Rawoof, A., Tiwari, N., Jaiswal, V., Kumar, A., ... & Ramchiary, N. (2025). Identification and expression analysis of genes involved in biosynthesis of Vitamin C and E in *Capsicum* species. *The Nucleus*, 1-28.

2. Islam, K., Rawoof, A., Kumar, A., Momo, J., Ahmed, I., **Dubey, M.,** & Ramchiary, N. (2023). Genetic regulation, environmental cues, and extraction methods for higher yield of secondary metabolites in capsicum. *Journal of Agricultural and Food Chemistry*, *71*(24), 9213-9242.

3. Jaiswal, V., Rawoof, A., Gahlaut, V., Ahmad, I., Chhapekar, S. S., **Dubey, M.,** & Ramchiary, N. (2022). Integrated analysis of DNA methylation, transcriptome, and global metabolites in interspecific heterotic *Capsicum* F1 hybrid. *Iscience*, *25*(11).

4. Ahmad, I., Rawoof, A., **Dubey, M.,** & Ramchiary, N. (2021). ICP-MS based analysis of mineral elements composition during fruit development in *Capsicum* germplasm. *Journal of Food Composition and Analysis*, *101*, 103977.

5. Islam, K., Rawoof, A., Ahmad, I., **Dubey, M**., Momo, J., & Ramchiary, N. (2021). *Capsicum* chinense MYB transcription factor genes: Identification, expression analysis, and their conservation and diversification with other solanaceae genomes. *Frontiers in Plant Science*, *12*, 721265.

6. Jaiswal, V., Rawoof, A., **Dubey, M.,** Chhapekar, S. S., Sharma, V., & Ramchiary, N. (2020). Development and characterization of non-coding RNA based simple sequence repeat markers in *Capsicum* species. *Genomics*, *112*(2), 1554-1564.

7. **Dubey, M.**, Jaiswal, V., Rawoof, A., Kumar, A., Nitin, M., Chhapekar, S. S., ... & Ramchiary, N. (2019). Identification of genes involved in fruit development/ripening in *Capsicum* and development of functional markers. *Genomics*, *111*(6), 1913-1922.

Book Chapters

1. Ahmad, I., Nitin, M., Rawoof, A., **Dubey, M**., & Ramchiary, N. (2019). Noncoding RNAs in *Capsicum* genome. In *The Capsicum Genome* (pp. 173-186). Cham: Springer International Publishing.

2. Jaiswal, V., Gahlaut, V., **Dubey, M.,** & Ramchiary, N. (2019). Genes/quantitative trait loci and associated molecular mechanisms identified in *Capsicum* genome for tolerance to abiotic and biotic stresses. In *The Capsicum genome* (pp. 121-138). Cham: Springer International Publishing.

CERTIFICATES

PUBLICATIONS